Description Logics with Abstraction and Refinement: From $\mathcal{ALC}$ to $\mathcal{EL}$

Carsten Lutz[1,2] , **Lukas Schulze**[1]

[1]Department of Computer Science, Leipzig University, Germany

[2]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI)

{clu, lschulze}@informatik.uni-leipzig.de

Abstract

We study extensions of description logics from the widely used $\mathcal{EL}$ family with operators that make it possible to speak about different levels of abstraction. We analyze the computational complexity of reasoning and show that often, this complexity is significantly lower than in the corresponding extension of the more expressive description logic $\mathcal{ALC}$. By slightly varying the semantics, we also obtain a case that admits reasoning in polynomial time.

1 Introduction

Knowledge representation with ontologies often involves concepts that are situated at different levels of abstraction or, equivalently, at different levels of granularity. For example, the widely known medical ontology SNOMED CT contains the concepts Arm, Hand, Finger, Phalanx, Osteocyte, and Mitochondrion which may reasonably be viewed as all belonging to different, increasingly more fine-grained levels of abstraction. Existing ontology languages, however, do not provide any explicit support for representing and interrelating different abstraction levels.

Recently, this shortcoming has led to the proposal of a scheme for extending description logics (DLs) with operators that make it possible to explicitly speak about different abstraction levels and their interaction (Lutz and Schulze 2023). The main features of this scheme are as follows. Each of the (finitely many) abstraction levels is associated with a classical DL interpretation. A refinement function associates objects on more coarse-grained levels with an ensemble of objects on more fine-grained levels. Such an ensemble is simply a tuple of objects that the refined object decomposes into. This may for instance be in the sense of mereological parts, but the scheme is by no means restricted to mereology.

Operators based on conjunctive queries (CQs) make it possible to describe how objects relate to their refining ensembles. These operators come in two flavours. A *refinement operator* expresses that every object of a certain kind refines into a certain kind of ensemble. For example, the statement

$$L_2{:}q_A \ \underline{\mathsf{refines}}\ L_1{:}\mathsf{Aircraft},$$

cr	ca	rr	ra	Semantics	$\mathcal{EL}$ / $\mathcal{ELH}_r$	$\mathcal{ALC}$
X				standard	coNP	Exp
X	X			standard	PSpace	2Exp
		X		standard	2Exp	2Exp
X	X	X	X	standard	2Exp	2Exp
X		X		set ensemble	PTime	Exp
X	X			set ensemble	coNP-hard	?

Figure 1: The complexity of satisfiability in abstraction DLs.

where q denotes the conjunctive query

$$q_A = \mathsf{Fuselage}(x_1) \wedge \mathsf{Wings}(x_2) \wedge \mathsf{Stabilizer}(x_3) \wedge \mathsf{carries}(x_2, x_1) \wedge \mathsf{carries}(x_1, x_3),$$

expresses that every instance of Aircraft on the more coarse-grained abstraction level L_1 decomposes into an ensemble of three objects on the more fine-grained level L_2, as described by q_A. Conversely, an *abstraction operator* expresses that for every ensemble of a certain kind, there is an object that refines into it. Reusing the query q_A from above, for example, it would be reasonable to also state

$$L_1{:}\mathsf{Aircraft}\ \underline{\mathsf{abstracts}}\ L_2{:}q_A$$

expressing that every ensemble that consists of a fuselage, a set of wings, and a stabilizer, related as stated by q_A, forms an aircraft. While the operators illustrated above speak about concepts such as Aircraft that are refined or abstracted, there are analogous operators also for roles (that is, binary relations) such as carries.

The DLs with abstraction and refinement proposed in (Lutz and Schulze 2023) are based on the expressive description logics $\mathcal{ALC}$ and $\mathcal{ALCI}$. In this paper, we replace $\mathcal{ALC}$ with important members of the $\mathcal{EL}$ family of description logics, in particular with the eponymous $\mathcal{EL}$ and its extension $\mathcal{ELH}_r$ with role hierarchies and range restrictions. These DLs play an important role in practice for at least three reasons. First, they are among very few description logics that admit reasoning in polynomial time. Second, a mild extension of $\mathcal{ELH}_r$ was standardized by the W3C as the EL profile of the widely used OWL 2 ontology language (Motik *et al.* 2009). And third, many prominent large-scale ontologies such as SNOMED CT are formulated in $\mathcal{ELH}_r$ or mild extensions thereof.

Two guiding questions for our investigation are: (1) Are the resulting DLs with abstraction and refinement computationally more well-behaved than those based on $\mathcal{ALC}$? And (2) Can we even identify useful cases where reasoning is possible in polynomial time? We remark that polynomial time cannot be expected in the presence of abstraction operators because, whenever these operators are present, then there is an obvious polynomial time reduction from the homomorphism problem on directed graphs; this implies that reasoning (concept satisfiability, to be precise) is at least NP-hard. Refinement operators, however, do not preclude polynomial time reasoning up-front.

We first prove that the extension $\mathcal{ELH}_r^{\mathsf{abs}}$ of $\mathcal{ELH}_r$ with abstraction and refinement operators for both concepts and roles still enjoys the existence of universal models (defined in terms of homomorphisms). This is important because the existence of universal models makes a crucial difference when designing algorithms, and in fact universal models underlie all important polynomial time reasoning algorithms for description logics. To construct universal models, we give a non-trivial chase procedure tailored specifically to $\mathcal{ELH}_r^{\mathsf{abs}}$. The algorithms behind our upper complexity bounds then all rely on universal models.

Our findings on the complexity of satisfiability in $\mathcal{ELH}_r^{\mathsf{abs}}$ and various fragments thereof are summarized in Figure 1. There, 'cr' stands for concept refinement operators, 'ca' for concept abstraction, and likewise for 'rr' and 'ra' and roles in place of concepts. We remark that subsumption and (un)satisfiability can be reduced to one another in polynomial time in all considered logics. All stated results are completeness results with the lower bounds holding already for (the respective fragments of) $\mathcal{EL}^{\mathsf{abs}}$ and the upper bounds applying to $\mathcal{ELH}_r^{\mathsf{abs}}$. The results shown in gray are from (Lutz and Schulze 2023).

Full $\mathcal{ELH}_r^{\mathsf{abs}}$ and $\mathcal{EL}^{\mathsf{abs}}$ turn out to be computationally no more well-behaved than in the case where $\mathcal{ALC}$ is used as the base logic: satisfiability is 2EXPTIME-complete in both cases. This still holds when only role refinement is admitted. The picture changes, however, in the important case where only concept-based operators are used, but no role-based ones. With only concept refinement, the complexity reduces to CONP which we consider a significant improvement as it enables the use of SAT solvers to decide satisfiability. With both concept refinement and abstraction, satisfiability is PSPACE-complete which is still significantly lower than 2EXPTIME-completeness in the case where $\mathcal{ALC}$ is used as the base logic.

To attain polynomial time, we change the semantics: instead of tuples of objects, ensembles are now sets of objects. While this has a subtle impact on modeling (see Example 4 in the paper), it is still a very reasonable semantics. Under this semantics, we indeed achieve polynomial time reasoning when only concept and role refinement is admitted.

To comply with space restrictions, proof details are provided in the appendix.

Related Work. As already explained, we adopt the framework of (Lutz and Schulze 2023). It is loosely related to description logics of context (Klarman and Gutiérrez-Basulto 2016) and to other multi-dimensional DLs (Wolter and Zakharyaschev 1999). Granularity has also received attention in foundational ontologies, see e.g. (Bittner and Smith 2003). There are other approaches to combine description logic and abstraction/granularity, but from very different perspectives and in technically very different ways, see for example (Calegari and Ciucci 2010; Cima *et al.* 2022; Glimm *et al.* 2017; Lisi and Mencar 2018).

2 Preliminaries

Fix countably infinite sets $\mathbf{C}$ and $\mathbf{R}$ of *concept names* and *role names*. $\mathcal{EL}$-*concepts* C, D take the form $C, D ::= \top \mid A \mid C \sqcap D \mid \exists r.C$ where A ranges over concept names and r over role names. An $\mathcal{ELH}_r$-*ontology* is a finite set $\mathcal{O}$ of *concept inclusion (CIs)* $C \sqsubseteq D$ with C and D $\mathcal{EL}$-concepts, *role inclusions* $r \sqsubseteq s$ with $r, s \in \mathbf{R}$, and *range restrictions* $\top \sqsubseteq \forall r.C$ with $r \in \mathbf{R}$ and C an $\mathcal{EL}$-concept. We say $\mathcal{O}$ is an $\mathcal{EL}$-ontology if it contains no role inclusions and range restrictions.

An *interpretation* is a pair $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ with $\Delta^\mathcal{I}$ a non-empty set (the *domain*) and $\cdot^\mathcal{I}$ an *interpretation function* that maps every concept name $A \in \mathbf{C}$ to a set $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$ and every role name $r \in \mathbf{R}$ to a binary relation $r^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$. The interpretation function is extended to compound concepts by setting $\top^\mathcal{I} = \Delta^\mathcal{I}$, $(C_1 \sqcap C_2)^\mathcal{I} = C_1^\mathcal{I} \cap C_2^\mathcal{I}$, and $(\exists r.C)^\mathcal{I} = \{d \in \Delta^\mathcal{I} \mid \exists e \in C^\mathcal{I} : (d, e) \in r^\mathcal{I}\}$. An interpretation $\mathcal{I}$ *satisfies* a concept inclusion $C \sqsubseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$ and likewise for role inclusions; it satisfies a range restriction $\top \sqsubseteq \forall r.C$ if $(d, e) \in r^\mathcal{I}$ implies $e \in C^\mathcal{I}$. We say that $\mathcal{I}$ is a *model* of an ontology $\mathcal{O}$ if $\mathcal{I}$ satisfies all inclusions and range restrictions in $\mathcal{O}$. We write $\mathcal{O} \models r \sqsubseteq s$ if every model of $\mathcal{O}$ satisfies $r \sqsubseteq s$. One can decide whether $\mathcal{O} \models r \sqsubseteq s$ in polynomial time by computing the reflexive-transitive closure of the role inclusions in $\mathcal{O}$.

A *conjunctive query (CQ)* $q(\bar{x})$ takes the form $q(\bar{x}) \leftarrow \varphi(\bar{x})$ with $\bar{x}$ a tuple of variables and φ a conjunction of *concept atoms* $C(x)$ and *role atoms* $r(x, y)$ where C is a (possibly compound) $\mathcal{EL}$-concept, r is a role name, and x, y are variables from $\bar{x}$. We require that every variable from $\bar{x}$ occurs in some atom of q, but may omit this atom in writing in case it is $\top(x)$. We may write $\alpha \in q$ to indicate that α is an atom in φ. With $\mathsf{var}(q)$, we denote the variables in φ. The *arity* of q is the length of $\bar{x}$. We say that q is *connected* if the undirected graph with node set $\mathsf{var}(q)$ and edge set $\{\{v, v'\} \mid r(v, v') \in q \text{ for any } r \in \mathbf{R}\}$ is. Note that CQs as defined here do not admit quantified variables. The reason is that admitting such variables results in DLs with abstraction and refinement to become undecidable, even when based on $\mathcal{EL}$ (Lutz and Schulze 2023). In examples, we shall often write only $\varphi(\bar{x})$ in place of $q(\bar{x}) \leftarrow \varphi(\bar{x})$. We then choose a variable naming scheme such as x_1, x_2, x_3 that makes clear the order of the variables in $\bar{x}$ (and we then assume that there are no repeated variables in $\bar{x}$).

Let $q(\bar{x})$ be a CQ and $\mathcal{I}$ an interpretation. A mapping $h : \bar{x} \to \Delta^\mathcal{I}$ is a *homomorphism* from q to $\mathcal{I}$ if $C(x) \in q$ implies $h(x) \in C^\mathcal{I}$ and $r(x, y) \in q$ implies $(h(x), h(y)) \in r^\mathcal{I}$. A tuple $\bar{d} \in (\Delta^\mathcal{I})^{|\bar{x}|}$ is an *answer* to q on $\mathcal{I}$ if there is a homomorphism h from q to $\mathcal{I}$ with $h(\bar{x}) = \bar{d}$. We use $q(\mathcal{I})$ to denote the set of all answers to q on $\mathcal{I}$.

For any syntactic object O such as an ontology or a concept, we use $||O||$ to denote the *size* of O, that is, the number of symbols needed to write O using a suitable alphabet.

3 DLs with Abstraction and Refinement

We extend $\mathcal{ELH}_r$ to the DL $\mathcal{ELH}_r^{\mathsf{abs}}$ that supports abstraction and refinement, following (Lutz and Schulze 2023). Fix a countable set $\mathbf{A}$ of *abstraction levels*. An $\mathcal{ELH}_r^{\mathsf{abs}}$-ontology is a finite set of statements of the following form:

- *labeled concept inclusions* $C \sqsubseteq_L D$, role inclusions $r \sqsubseteq_L s$, and range restrictions $\top \sqsubseteq_L \forall r, C$,
- *concept refinements* $L{:}q(\bar{x})$ $\underline{\mathsf{refines}}$ $L'{:}C$,
- *concept abstractions* $L'{:}C$ $\underline{\mathsf{abstracts}}$ $L{:}q(\bar{x})$,
- *role refinements* $L{:}q(\bar{x}, \bar{y})$ $\underline{\mathsf{refines}}$ $L'{:}q_r(x, y)$,
- *role abstractions* $L'{:}r$ $\underline{\mathsf{abstracts}}$ $L{:}q(\bar{x}, \bar{y})$

where L, L' range over abstraction levels from $\mathbf{A}$, C, D over $\mathcal{EL}$-concepts, r over role names, q over conjunctive queries, and q_r over conjunctive queries of the form $C_1(x) \wedge r(x, y) \wedge C_2(y)$. In concept and role abstractions, we additionally require the CQ q to be connected.

We also consider various fragments of $\mathcal{ELH}_r^{\mathsf{abs}}$. With $\mathcal{ELH}_r^{\mathsf{abs}}$[cr,ca], for example, we mean the fragment of $\mathcal{ELH}_r^{\mathsf{abs}}$ that admits only concept refinement and concept abstraction, but neither role refinement nor role abstraction (which are identified by rr and ra). As in the base case, we drop $\mathcal{H}$ if no role inclusions are admitted and likewise for $\cdot_r$ and range restrictions.

We next define the semantics of $\mathcal{ELH}_r^{\mathsf{abs}}$, based on *A-interpretations* that include one traditional DL interpretation for each abstraction level. Formally, an A-interpretation takes the form $\mathcal{I} = (\mathbf{A}_\mathcal{I}, \prec, (\mathcal{I}_L)_{L \in \mathbf{A}_\mathcal{I}}, \rho)$, where

- $\mathbf{A}_\mathcal{I} \subseteq \mathbf{A}$ is the set of relevant abstraction levels;
- $\prec \subseteq \mathbf{A}_\mathcal{I} \times \mathbf{A}_\mathcal{I}$ is such that the directed graph $(\mathbf{A}_\mathcal{I}, \prec)$ is a tree;[1] intuitively, $L \prec L'$ means that L is less abstract than L' or, in other words, that the modeling granularity of L is finer than that of L';
- $(\mathcal{I}_L)_{L \in \mathbf{A}_\mathcal{I}}$ is a collection of interpretations $\mathcal{I}_L$, one for every $L \in \mathbf{A}_\mathcal{I}$, with pairwise disjoint domains; we use $L(d)$ to denote the unique $L \in \mathbf{A}_\mathcal{I}$ with $d \in \Delta^{\mathcal{I}_L}$;
- ρ is the *refinement function*, a partial function that associates pairs $(d, L) \in \Delta^\mathcal{I} \times \mathbf{A}_\mathcal{I}$ such that $L \prec L(d)$ with an *L-ensemble* $\rho(d, L)$, that is, with a non-empty tuple over $\Delta^{\mathcal{I}_L}$. We want every object to participate in only one ensemble and thus require that

$(*)$ for all $d \in \Delta^\mathcal{I}$ and $L \in \mathbf{A}_\mathcal{I}$, there is at most one $e \in \Delta^{\mathcal{I}_L}$ such that d occurs in $\rho(e, L(d))$.

For readability, we may write $\rho_L(d)$ in place of $\rho(d, L)$.

An A-interpretation $\mathcal{I} = (\mathbf{A}_\mathcal{I}, \prec, (\mathcal{I}_L)_{L \in \mathbf{A}_\mathcal{I}}, \rho)$ *satisfies*

- a labeled concept inclusion $C \sqsubseteq_L D$ if $\mathcal{I}_L$ satisfies $C \sqsubseteq D$, and likewise for role inclusions and range restrictions;
- $L{:}q(\bar{x})$ $\underline{\mathsf{refines}}$ $L'{:}C$ if $L \prec L'$ and for all $d \in C^{\mathcal{I}_{L'}}$, there is an $\bar{e} \in q(\mathcal{I}_L)$ such that $\rho_L(d) = \bar{e}$;
- $L'{:}C$ $\underline{\mathsf{abstracts}}$ $L{:}q(\bar{x})$ if $L \prec L'$ and for all $\bar{e} \in q(\mathcal{I}_L)$, there is a $d \in C^{\mathcal{I}_{L'}}$ such that $\rho_L(d) = \bar{e}$;
- $L{:}q(\bar{x}, \bar{y})$ $\underline{\mathsf{refines}}$ $L'{:}q_r(x, y)$ if $L \prec L'$ and for all $(d_1, d_2) \in q_r(\mathcal{I}_{L'})$, there is an $(\bar{e}_1, \bar{e}_2) \in q(\mathcal{I}_L)$ such that $\rho_L(d_1) = \bar{e}_1$ and $\rho_L(d_2) = \bar{e}_2$;
- $L'{:}r$ $\underline{\mathsf{abstracts}}$ $L{:}q(\bar{x}, \bar{y})$ if $L \prec L'$ and for all $(\bar{e}_1, \bar{e}_2) \in q(\mathcal{I}_L)$, there is a $(d_1, d_2) \in r^{\mathcal{I}_{L'}}$ such that $\rho_L(d_1) = \bar{e}_1$ and $\rho_L(d_2) = \bar{e}_2$.

An A-interpretation is a *model* of an $\mathcal{ELH}_r^{\mathsf{abs}}$-ontology if it satisfies all inclusions, refinements, etc in it.

Example 1. *We consider the domain of actions. Assume that there is a* $\mathsf{MealPrep}$ *action that refines into subactions:* $L_2{:}q_M$ $\underline{\mathsf{refines}}$ $L_1{:}\mathsf{MealPrep}$ *where*

$$q_M = \mathsf{Buying}(x_1) \wedge \mathsf{Cooking}(x_2) \wedge \mathsf{precedes}(x_1, x_2)$$

We might have budget-friendly meal preparation and buying actions:

$$\mathsf{BudgetMealPrep} \sqsubseteq_{L_1} \mathsf{MealPrep}$$
$$\mathsf{BudgetBuying} \sqsubseteq_{L_2} \mathsf{Buying}$$
$$\mathsf{BudgetBuying} \sqcap \exists \mathsf{bought}.\mathsf{Expensive} \sqsubseteq_{L_2} \bot$$

A budget-friendly meal preparation requires buying non-expensive ingredients: $L_2{:}q_B$ $\underline{\mathsf{refines}}$ $L_1{:}\mathsf{BudgetMealPrep}$ *where*

$$q_B(x_1, x_2) = \mathsf{BudgetBuying}(x_1).$$

We are interested in two reasoning problems: *concept satisfiability* and *concept subsumption*. Concept satisfiability means to decide, given an ontology $\mathcal{O}$, an $\mathcal{EL}$-concept C, and an abstraction level $L \in \mathbf{A}$, whether there is a model $\mathcal{I}$ of $\mathcal{O}$ such that $C^{\mathcal{I}_L} \neq \emptyset$. We then say that C is *L-satisfiable w.r.t.* $\mathcal{O}$ and call $\mathcal{I}$ an *L-model of* C *and* $\mathcal{O}$.

For concept subsumption, we are given an ontology $\mathcal{O}$, two concepts C and D, and an abstraction level $L \in \mathbf{A}$, and are asked to decide whether $C^{\mathcal{I}_L} \subseteq D^{\mathcal{I}_L}$ in every model $\mathcal{I}$ of $\mathcal{O}$. If this is the case we say that C *is L-subsumed by* D *w.r.t.* $\mathcal{O}$ and write $\mathcal{O} \models C \sqsubseteq_L D$.

We remark that the $\bot$-concept, interpreted as $\bot^\mathcal{I} = \emptyset$ in every interpretation $\mathcal{I}$, can be expressed in $\mathcal{EL}^{\mathsf{abs}}$[cr] at the expense of introducing fresh symbols: a CI $C \sqsubseteq_L \bot$ can be simulated by

$$L'{:}A(x) \ \underline{\mathsf{refines}} \ L{:}C \quad L'{:}r(x_1, x_2) \ \underline{\mathsf{refines}} \ L{:}C$$

where A, r, and L are a fresh concept name, role name, and abstraction level. This is because the two refinements require ensembles of different length. W.l.o.g., we thus use the $\bot$-concept whenever convenient.

Using $\bot$, (un)satisfiability and subsumption are easily interreducible in polynomial time. In fact, it is not hard to see that a concept C is L-unsatisfiable w.r.t. an ontology $\mathcal{O}$ iff C is L-subsumed by some fresh concept name A w.r.t. $\mathcal{O}$. Conversely, C is L-subsumed by D w.r.t. $\mathcal{O}$ iff C is L-unsatisfiable w.r.t. $\mathcal{O} \cup \{C \sqcap D \sqsubseteq \bot\}$. We thus state all our

[1] Dropping this restriction results in undecidability (Lutz and Schulze 2023).

results in terms of satisfiability and assume that it is understood that (up to complementation) they also apply to subsumption.

4 Upper Bounds

We prove upper complexity bounds for satisfiability in $\mathcal{ELH}_r^{\mathsf{abs}}$. A 2EXPTIME upper bound for full $\mathcal{ELH}_r^{\mathsf{abs}}$ follows from the results in (Lutz and Schulze 2023). We thus concentrate on the fragments $\mathcal{ELH}_r^{\mathsf{abs}}[\text{cr}]$ and $\mathcal{ELH}_r^{\mathsf{abs}}[\text{cr}, \text{ca}]$.

4.1 Simplifying Assumptions

We discuss some assumptions, all w.l.o.g., made throughout Section 4. First, we assume that the input ontology $\mathcal{O}$ is in *normal form*, meaning that:

1. all the CIs in $\mathcal{O}$ are of one of the following forms, where $A, A_1, \ldots, A_n, B$ are concept *names*:

$$\top \sqsubseteq_L A \qquad A_1 \sqcap \cdots \sqcap A_n \sqsubseteq_L B$$
$$A \sqsubseteq_L \exists r.B \qquad \exists r.A \sqsubseteq_L B$$

2. all range restrictions and (concept and role) refinements and abstractions contain only concept names, but no compound concepts, also inside of CQs.

By introducing new concept and role names, any $\mathcal{ELH}_r^{\mathsf{abs}}$ ontology $\mathcal{O}$ can be converted into an ontology $\mathcal{O}'$ in normal form that is a conservative extension of $\mathcal{O}$, i.e., every model of $\mathcal{O}'$ is also a model of $\mathcal{O}$, and every model of $\mathcal{O}$ can be extended to a model of $\mathcal{O}'$ by appropriately choosing the interpretations of the concept names that have been introduced during the conversion. The conversion takes only linear time, see for example (Baader *et al.* 2005). We further assume that the concept C_0 whose satisfiability is to be decided is a concept name, thus not compound. Finally, we assume that the abstraction level L_0 for which satisfiability is to be decided is the root of the tree $G_\mathcal{O}$ that is defined by the abstractions and refinements in $\mathcal{O}$. Let us make the latter more precise.

We use $\mathbf{A}_\mathcal{O}$ to denote the set of abstraction levels mentioned in $\mathcal{O}$ and $\prec_\mathcal{O}$ for the smallest relation on $\mathbf{A}_\mathcal{O}$ such that $L \prec_\mathcal{O} L'$ if $\mathcal{O}$ contains a concept refinement $L{:}q(\bar{x})$ $\underline{\mathsf{refines}}$ $L'{:}C$ or a concept abstraction $L'{:}C$ $\underline{\mathsf{abstracts}}$ $L{:}q(\bar{x})$. The *abstraction graph of an ontology* $\mathcal{O}$ is the directed graph

$$G_\mathcal{O} = (\mathbf{A}_\mathcal{O}, \prec_\mathcal{O}^{-1}).$$

Note that by the definition of the semantics, $\mathcal{O}$ being satisfiable implies that $G_\mathcal{O}$ is a tree.

Now assume that the abstraction level L_0 for which satisfiability is to be decided is not the root of $G_\mathcal{O}$, but L_R is. Then $G_\mathcal{O}$ contains a path $L_R = \widehat{L}_1, \ldots, \widehat{L}_k = L_0$ and we can extend $\mathcal{O}$ with concept refinements $\widehat{L}_{i+1}{:}\, A(x)$ $\underline{\mathsf{refines}}$ $\widehat{L}_i{:}\, A$ and $\widehat{L}_k{:}\, C_0(x)$ $\underline{\mathsf{refines}}$ $\widehat{L}_{k-1}{:}\, A$ for $1 \leq i < k$, with A a fresh concept name, and decide L_R-satisfiability of A w.r.t. the extended ontology.

4.2 Universal Models and The Chase

A crucial property of description logics of the $\mathcal{EL}$ family is the existence of universal models, defined in terms of homomorphisms. In particular, universal models are at the basis of all polynomial time algorithms for description logic reasoning that we are aware of. A fundamental observation that underlies the design of our algorithms is that universal models also exist for $\mathcal{ELH}_r^{\mathsf{abs}}$.

Let $\mathcal{I}_i = (\mathbf{A}_{\mathcal{I}_i}, \prec_i, (\mathcal{I}_{L,i})_{L \in \mathbf{A}_{\mathcal{I}_i}}, \rho_i)$ be an A-interpretation, for $i \in \{1, 2\}$. A function $h\colon \Delta^{\mathcal{I}_1} \to \Delta^{\mathcal{I}_2}$ is a *homomorphism* from $\mathcal{I}_1$ to $\mathcal{I}_2$ if the following conditions are satisfied, for all $d, e \in \Delta^{\mathcal{I}_1}$:

1. $L(d) = L(h(d))$;
2. $\prec_1 \subseteq \prec_2$;
3. $d \in A^{\mathcal{I}_1}$ implies $h(d) \in A^{\mathcal{I}_2}$ for all $A \in \mathbf{C}$;
4. $(d, e) \in r^{\mathcal{I}_1}$ implies $(h(d), h(e)) \in r^{\mathcal{I}_2}$ for all $r \in \mathbf{R}$;
5. $\rho_1(d, L) = \bar{e}$ implies $\rho_2(h(d), L) = h(\bar{e})$

where $h(\bar{e})$ is the tuple obtained from $\bar{e}$ by applying h component-wise. Note that this implies $\mathbf{A}_{\mathcal{I}_1} \subseteq \mathbf{A}_{\mathcal{I}_2}$.

Let C_0 be an $\mathcal{EL}$-concept, $\mathcal{O}$ an $\mathcal{ELH}_r^{\mathsf{abs}}$-ontology, and $L_0 \in \mathbf{A}$ an abstraction level. A model $\mathcal{I}$ of $\mathcal{O}$ with distinguished element $d \in C_0^\mathcal{I}$, where $L(d) = L_0$, is a *universal L_0-model of C_0 and $\mathcal{O}$* if the following holds: for every model $\mathcal{J}$ of $\mathcal{O}$ and every $e \in C_0^\mathcal{J}$ with $L(e) = L_0$, there exists a homomorphism h from $\mathcal{I}$ to $\mathcal{J}$ with $h(d) = e$. Our aim is to show the following.

Lemma 1. *Let C_0 be an $\mathcal{EL}$-concept, $\mathcal{O}$ an $\mathcal{ELH}_r^{\mathsf{abs}}$-ontology, and $L_0 \in \mathbf{A}$. If C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$, then there exists a universal L_0-model of C_0 and $\mathcal{O}$.*

Lemma 1 is proved by a somewhat intricate chase procedure. For technical reasons, this chase may construct structures that do not satisfy all the conditions required of A-interpretations. The chase does thus not run directly in A-interpretations, but rather on a weakening that we call interpretation candidates.

Let $\mathbf{K}$ be a countably infinite set of *constants*. A *fact* is an expression of the form $A(a)$ or $r(a, b)$ where A is a concept name, r a role name, and a, b are constants. Homomorphisms from conjunctive queries to sets of facts are defined in the expected way. An *interpretation candidate* is a triple $I = (F, \rho, \sim)$ where

- F is a *fact assignment*, that is, a function that maps each abstraction level $L \in \mathbf{A}_\mathcal{O}$ to a set of facts $F(L)$. We use $\mathsf{dom}(F(L))$ to denote the *domain* of $F(L)$, that is, $\mathsf{dom}(F(L)) = \{a \in \mathbf{K} \mid a \text{ is used in a fact in } F(L)\}$. We demand that the $F(L)$ have pairwise disjoint domains and may write $\mathsf{dom}(F)$ to denote $\biguplus_{L \in \mathbf{A}_\mathcal{O}} \mathsf{dom}(F_L)$. We further use $L(a)$, for any $a \in \mathsf{dom}(F)$, to denote the unique $L \in \mathbf{A}_\mathcal{O}$ such that $a \in \mathsf{dom}(F(L))$;
- ρ is a *refinement function*, that is, a partial function that associates pairs $(a, L) \in \mathsf{dom}(F) \times \mathbf{A}_\mathcal{O}$ such that $L \prec L(a)$ with an *L-ensemble* $\rho(a, L)$, that is, with a non-empty tuple over $\mathsf{dom}(F(L))$;

- $\sim$ is an *equivalence relation* on the set $\mathsf{dom}(F)$. If we *set* $a_1 \sim a_2$, we mean to set $\sim := \sim \cup \{(a_1, a_2)\}$ and add the smallest number of tuples such that $\sim$ is again an equivalence relation. We use $[a]$ to denote the *equivalence class* of $a \in \mathsf{dom}(F)$ w.r.t. $\sim$.

For readability we may write F_L instead of $F(L)$ and $\rho_L(a)$ instead of $\rho(a, L)$. Note that ρ, in contrast to the refinement function in A-interpretations, allows elements to be part of multiple ensembles.

Our chase procedure starts from the *initial interpretation candidate* $I_0 = (F^0, \rho^0, \sim^0)$ for C_0, L_0, and $\mathcal{O}$ where $F^0_L = \{A_\top(a_L)\}$ for all $L \in \mathbf{A}_\mathcal{O} \setminus \{L_0\}$, $F^0_{L_0} = \{C_0(a_0)\}$, ρ^0 is empty, and $\sim^0$ is the identity. It then applies a set of rules in a fair way, that is, every rule that is applicable will eventually be applied. The chase may also abort and report unsatisfiability of the input. We start by giving the rules that treat inclusions and range restrictions:

R1 if $A_1(a) \in F_L, \dots, A_n(a) \in F_L$, and $A_1 \sqcap \cdots \sqcap A_n \sqsubseteq_L B \in \mathcal{O}$, then add $B(a)$ to F_L;

R2 if $a \in \mathsf{dom}(F_L)$ and $\top \sqsubseteq_L A \in \mathcal{O}$, then add $A(a)$ to F_L;

R3 if $A(a) \in F_L$, $A \sqsubseteq_L \exists r.B \in \mathcal{O}$, then add $r(a, b)$ and $B(b)$ to F_L with b as a fresh constant;

R4 if $r(a, b) \in F_L$, $A(b) \in F_L$, and $\exists r.A \sqsubseteq_L B \in \mathcal{O}$, then add $B(a)$ to F_L.

R5 if $r(a, b) \in F_L$ and $r \sqsubseteq_L s \in \mathcal{O}$, then add $s(a, b)$ to F_L;

R6 if $r(a, b) \in F_L$ and $\top \sqsubseteq_L \forall r.C \in \mathcal{O}$, then add $C(a)$ to F_L;

Next up are the rules that pertain to concept refinements and abstractions in $\mathcal{O}$. We may use $x \in \bar{x}$ to express that variable x occurs in the tuple $\bar{x}$. For a CQ $q(\bar{x})$ and a tuple of constants $\bar{a}$ with $|\bar{x}| = |\bar{a}|$, we use $q(\bar{a})$ to denote the set of facts obtained from q by replacing in every atom the i-th variable in $\bar{x}$ by the i-th constant in $\bar{a}$, for $1 \leq i \leq |\bar{a}|$.

R7 if $A(a) \in F_L$, $L'{:}\, q(\bar{x})$ <u>refines</u> $L{:}\, A \in \mathcal{O}$, and $\rho_{L'}(a)$ is undefined, then set $\rho_{L'}(a) = \bar{a}$ for a tuple $\bar{a}$ of fresh constants with $|\bar{a}| = |\bar{x}|$;

R8 if $A(a) \in F_L$, $L'{:}\, q(\bar{x})$ <u>refines</u> $L{:}\, A \in \mathcal{O}$, $\rho_{L'}(a)$ is defined, and $|\bar{x}| = |\rho_{L'}(a)|$, then add $q(\rho_{L'}(a))$ to $F_{L'}$; if $|\bar{x}| \neq |\rho_{L'}(a)|$, then return 'unsatisfiable';

R9 if h is a homomorphism from q to F_L for any concept abstraction $L'{:}\, A$ <u>abstracts</u> $L{:}\, q(\bar{x}) \in \mathcal{O}$ and there is no $a \in \mathsf{dom}(F_{L'})$ with $\rho_L(a) = h(\bar{x})$, then introduce a fresh constant a and set $\rho_L(a) = h(\bar{x})$;

R10 if h is a homomorphism from q to F_L for any concept abstraction $L'{:}\, A$ <u>abstracts</u> $L{:}\, q(\bar{x}) \in \mathcal{O}$ and there is an $a \in \mathsf{dom}(F_{L'})$ with $\rho_L(a) = h(\bar{x})$, then add $A(a)$ to $F_{L'}$.

There are analogous rules for role refinement and role abstraction, given in the appendix. We also have rules that concern overlapping ensembles. Intuitively, overlapping ensembles require the identification of elements, but we do not want to do this in the chase itself to preserve monotonicity, that is, rule applications should always extend the interpretation candidate. We thus only record the necessary identifications in the '$\sim$' component of interpretation candidates.

R15 if $\rho_L(a_1) = \bar{e}_1$, $\rho_L(a_2) = \bar{e}_2$, there are $b_1 \in \bar{e}_1$ and $b_2 \in \bar{e}_2$ with $b_1 \sim b_2$ and $|\bar{e}_1| \neq |\bar{e}_2|$, then return 'unsatisfiable';

R16 if there are $b_1 \in \rho_L(a_1)$ and $b_2 \in \rho_L(a_2)$ with $b_1 \sim b_2$, then set $a_1 \sim a_2$;

R17 if $a_1 \sim a_2$, $\rho_L(a_1) = \bar{e}_1$, and $\rho_L(a_2) = \bar{e}_2$ with $|\bar{e}_1| = |\bar{e}_2|$, then set $\bar{e}_1[i] \sim \bar{e}_2[i]$ for $1 \leq i \leq |\bar{e}_1|$;

R18 if $a_1 \sim a_2$ and fact $f \in F_L$ contains constant a_1, then add to F_L the fact obtained from f by replacing some occurrence of a_1 with a_2.

R19 if $a_1 \sim a_2$, $\rho_L(a_1)$ is defined, and $\rho_L(a_2)$ is undefined, then add to F_L facts $A_\top(b_1), \dots, A_\top(b_n)$, with $b_1, \dots, b_n$ fresh constants and $n = |\rho_L(a_1)|$, and set $\rho_L(a_2) = (b_1, \dots, b_n)$ (where $A_\top$ is a fresh concept name).

A *chase sequence* is a sequence of interpretation candidates $I_0, I_1, \dots$ such that $I_0 = (F^0, \rho^0, \sim^0)$, I_{i+1} is obtained from I_i by applying one of the rules defined above. Every chase sequence $I_0, I_1, \dots$ gives rise to an interpretation candidate $I^* = (F^*, \rho^*, \sim^*)$ in the limit, with $F^* = \bigcup_i F_i$, $\rho^* = \bigcup_i \rho_i$, and $\sim^* = \bigcup_i \sim^i$. We also call I^* the *result* of chasing C_0 w.r.t. L_0 and $\mathcal{O}$. It can be shown that, up to isomorphism, all fair chase sequences deliver the same result.[2]

The chase is sound and complete in the following sense.

Lemma 2. *Let $\mathcal{O}$ be an $\mathcal{ELH}^{\mathsf{abs}}_r$-ontology in normal form whose abstraction graph $G_\mathcal{O}$ is a tree, C_0 a concept name, and L_0 an abstraction level. The L_0-chase on C_0 and $\mathcal{O}$ does not abort if and only if C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$.*

In the proof of the 'only if' direction of Lemma 2 (soundness), we start from a non-aborting chase sequence that delivers a result $I^* = (F^*, \rho^*, \sim^*)$, and then construct from I^* an L_0-model $\mathcal{I}$ of C_0 and $\mathcal{O}$. Intuitively, we apply filtration to make the equalities recorded in $\sim^*$ real equalities. This is achieved by setting $\mathcal{I} = (\mathbf{A}_\mathcal{I}, \prec, (\mathcal{I}_L)_{L \in \mathbf{A}_\mathcal{I}}, \rho)$ where

$$\begin{aligned}
\Delta^{\mathcal{I}_L} &= \{[a] \mid a \in \mathsf{dom}(F^*_L)\} \\
A^{\mathcal{I}_L} &= \{[a] \mid A(a') \in F^*_L \text{ and } a' \in [a]\} \\
r^{\mathcal{I}_L} &= \{([a], [b]) \mid r(a', b') \in F^*_L \text{ and } a' \in [a], b' \in [b]\} \\
\rho_L &= \{([a], ([b_1] \cdots [b_n])) \mid (a', (b'_1 \cdots b'_n)) \in \rho^*_L \text{ with} \\
&\qquad a' \in [a], b'_i \in [b_i] \text{ for } 1 \leq i \leq n\}.
\end{aligned}$$

The remaining components $\mathbf{A}_\mathcal{I}$ and $\prec$ are defined as $\mathbf{A}_\mathcal{O}$ and $\prec_\mathcal{O}$, respectively. We show in the appendix that $\mathcal{I}$ is not only an L_0-model of C_0 and $\mathcal{O}$, but even a universal such model, thus proving Lemma 1.

4.3 $\mathcal{ELH}^{\mathsf{abs}}_r[\mathrm{cr}]$ in CONP

Our aim is to prove the following.

Theorem 1. *Satisfiability in $\mathcal{ELH}^{\mathsf{abs}}_r[\mathrm{cr}]$ is in* CONP.

[2] Note that our rule R3 is oblivious in the sense that it may always add a fresh constant b even if there is already a b' with $r(a, b')$ and $B(b')$ in F_L.

It suffices to find an NP algorithm for unsatisfiability. Assume that the concept name C_0, the $\mathcal{ELH}_r^{\mathsf{abs}}$[cr]-ontology $\mathcal{O}$, and the abstraction level L_0 are given as an input, that is, we want to decide whether C_0 is L_0-unsatisfiable w.r.t. $\mathcal{O}$. If the abstraction graph $G_\mathcal{O}$ of $\mathcal{O}$ is not a tree, we directly return 'unsatisfiable'. Otherwise, the only remaining way in which unsatisfiability may arise is that there are two refinement statements that both apply to the same element of a model, but require ensembles of different length.

Example 2. *Consider the following ontology $\mathcal{O}$:*

$$C_0 \sqsubseteq \exists r.A_1$$
$$L_1\colon A_2(x) \ \underline{\mathsf{refines}}\ L_0\colon A_1$$
$$A_2 \sqsubseteq \exists s.A_3$$
$$L_2\colon B(x) \ \underline{\mathsf{refines}}\ L_1\colon A_3$$
$$L_2\colon r(x_1,x_2) \ \underline{\mathsf{refines}}\ L_1\colon A_3$$

The reader may try to construct an L_0-model of C_0 and $\mathcal{O}$, following the sequence of existential quantifications and refinements suggested by the order of the statements in $\mathcal{O}$.

As suggested by Example 2, our algorithm guesses a sequence of existential quantifications and refinements that lead to two 'incompatible' refinements. To make this precise, we need some preliminaries.

We use u to denote the *universal role*, that is, a fixed role name that is always interpreted as $\mathsf{u}^\mathcal{I} = \Delta^\mathcal{I} \times \Delta^\mathcal{I}$. An *ABox* is a finite set of facts as defined in the previous section. An interpretation $\mathcal{I}$ *satisfies* a concept assertion $C(a)$ if $a \in C^\mathcal{I}$, a role assertion $r(a,b)$ if $(a,b) \in r^\mathcal{I}$, and an ABox $\mathcal{A}$ if it satisfies all concept and role assertions in it. For an ABox $\mathcal{A}$, an $\mathcal{ELH}_r$-ontology $\mathcal{O}$, and an $\mathcal{EL}$-concept C, we write $\mathcal{A}, \mathcal{O} \models \exists \mathsf{u}.C$ if $C^\mathcal{I} \neq \emptyset$ in every model $\mathcal{I}$ of $\mathcal{O}$ that satisfies $\mathcal{A}$. It is known that given $\mathcal{A}$, $\mathcal{O}$, and C, it can be decided in polynomial time whether $\mathcal{A}, \mathcal{O} \models \exists \mathsf{u}.C$ (Krötzsch 2010). Note that a conjunctive query can be viewed as an ABox in an obvious way, by viewing variables as constants.

For a set of concept names S and $L, L' \in \mathbf{A}_\mathcal{O}$, we use $Q^{\mathsf{ref}}_{L,L'}(S)$ to denote the set of CQs $q(\bar{x})$ such that $\mathcal{O}$ contains a concept refinement $L'{:}q(\bar{x})$ $\underline{\mathsf{refines}}$ $L{:}C$ with $C \in S$. We assume that the conjunctive queries $q(\bar{x})$ in concept refinements in $\mathcal{O}$ use canonical variable names, that is, the variable with the left-most occurrence in $\bar{x}$ is x_1, the variable that occurs next is x_2, etc.

For an abstraction level $L \in \mathbf{A}_\mathcal{O}$, we use $\mathcal{O}_L$ to denote the $\mathcal{ELH}_r$-ontology that consists of all concept inclusions $C \sqsubseteq D$ such that $C \sqsubseteq_L D \in \mathcal{O}$, all role inclusions $r \sqsubseteq s$ such that $r \sqsubseteq_L s \in \mathcal{O}$, and all range restrictions $\top \sqsubseteq \forall r.C$ such that $\top \sqsubseteq_L \forall r.C \in \mathcal{O}$.

We are now ready to describe the algorithm. It guesses a sequence $S_1, L_1, \dots, S_n, L_n$ where $S_1, \dots, S_n$ are sets of concept names that occur in $\mathcal{O}$ and $L_1, \dots, L_n$ are abstraction levels, $n \leq |\mathbf{A}_\mathcal{O}|$. It accepts if the following conditions are satisfied, and rejects otherwise:

1. one of the following holds:
   - $L_1 = L_0$ and $\mathcal{O}_{L_1} \models C_0 \sqsubseteq_{L_1} \exists \mathsf{u}.(\bigsqcap S_1)$ or
   - $\mathcal{O}_{L_1} \models \top \sqsubseteq_{L_1} \exists \mathsf{u}.(\bigsqcap S_1)$;
2. for $1 \leq i < n$:
   $$\mathcal{A}_i, \mathcal{O}_{L_i} \models \exists \mathsf{u}.(\bigsqcap S_{i+1})$$
   where $\mathcal{A}_i$ is the union of all queries in $Q^{\mathsf{ref}}_{L_i,L_{i+1}}(S_i)$, viewed as ABoxes.[3]
3. There are concept refinements $L'\colon q(\bar{x})$ $\underline{\mathsf{refines}}$ $L\colon A \in \mathcal{O}$ and $L'\colon q'(\bar{x}')$ $\underline{\mathsf{refines}}$ $L\colon B \in \mathcal{O}$ such that $A, B \in S_n$ and $|\bar{x}| \neq |\bar{x}'|$.

Note that in Example 2, we have always interleaved a single existential restriction with each refinement statement. In general, however, there can be a more complex '$\mathcal{EL}$-derivation' between two subsequent refinements, and we abstract away from that by using the universal role.

Lemma 3. *The algorithm accepts iff C_0 is L_0-unsatisfiable w.r.t. $\mathcal{O}$.*

The proof of Lemma 3 crucially uses universal models as produced by the chase procedure from Section 4.2.

Note that, by what was said above, Conditions 1 to 3 can be checked in polynomial time. We have thus obtained an NP algorithm, as desired.

4.4 $\mathcal{ELH}_r^{\mathsf{abs}}$[cr, ca] in PSPACE

We now add concept abstraction, that is, we move from $\mathcal{ELH}_r^{\mathsf{abs}}$[cr] to $\mathcal{ELH}_r^{\mathsf{abs}}$[cr, ca]. This makes a significant difference because now we can also pass information upwards through the tree-shaped abstraction graph of the ontology, as illustrated by the following example.

Example 3. *Consider the following ontology $\mathcal{O}$:*

$$L_1{:}A_1(x) \ \underline{\mathsf{refines}}\ L_0{:}C_0 \qquad L_2{:}A_2(x) \ \underline{\mathsf{refines}}\ L_0{:}C_0$$
$$A_1 \sqsubseteq B_1 \qquad A_2 \sqsubseteq B_2$$
$$L_0{:}C_1 \ \underline{\mathsf{abstracts}}\ L_1{:}B_1(x) \qquad L_0{:}C_2 \ \underline{\mathsf{abstracts}}\ L_2{:}B_2(x)$$
$$C_1 \sqcap C_2 \sqsubseteq \bot$$

C_0 is L_0-unsatisfiable w.r.t. $\mathcal{O}$, but there is no (linear!) sequence of existential quantifications and refinements as in Example 2.

We want to prove the following, which is substantially more difficult than proving the coNP upper bound in the previous section. In fact, we view the following as a main result of this paper.

Theorem 2. *Satisfiability in $\mathcal{ELH}_r^{\mathsf{abs}}$[cr, ca] is in* PSPACE.

Let a concept name C_0, an $\mathcal{ELH}_r^{\mathsf{abs}}$[cr, ca]-ontology $\mathcal{O}$, and an abstraction level $L_0 \in \mathbf{A}_\mathcal{O}$ be given as an input. If the abstraction graph of $\mathcal{O}$ is not a tree, we immediately return 'unsatisfiable'.

Our algorithm has some resemblance with the standard non-deterministic PSPACE algorithm for the satisfiability of $\mathcal{ALC}$ concepts (without ontologies) that verifies the existence of a tree model of polynomial depth by traversing it in a depth-first manner, always keeping only a single path in memory (Baader *et al.* 2017). In our case, we want to verify the existence of an A-interpretation $\mathcal{I}$ that is an L_0-model of C_0 and $\mathcal{O}$. The tree that our algorithm traverses is

[3] Here we rely on canonical variable names.

$(\mathbf{A}_\mathcal{I}, \prec^{-1})$, which we can w.l.o.g. assume to be the abstraction graph of $\mathcal{O}$ (since universal models constructed by the chase have this property).

We are, however, confronted with two challenges. First, the 'upwards' nature of abstractions makes it difficult to traverse the tree in a depth-first manner. We address this by a suitable guessing strategy. And second, the interpretations $\mathcal{I}_L$ of a universal model, which correspond to the nodes of the traversed tree, are infinite and thus cannot be guessed. While infinite but regularly-shaped models can often be substituted by 'compact' finite models of polynomial size when designing algorithms for plain $\mathcal{EL}$ (Lutz *et al.* 2009), this is no longer true in the presence of CQs. To address this, we stick with infinite models $\mathcal{I}_L$, but *represent* them by compact (finite!) interpretations that we call pseudo-models. We then use a non-standard semantics for CQs on those compact representations.

Pseudo-Models For the following, one should imagine the interpretations $\mathcal{I}_L$ to take the shape of an infinite tree whose nodes are ensembles and domain elements that do not participate in an ensemble. In a pseudo-model $\mathcal{I}$, intuitively we identify ensembles / non-ensemble elements that are isomorphic, thus obtaining finiteness but losing the tree-shape.

Let $q(\bar{x})$ be a CQ and $\mathcal{I}$ an $\mathcal{EL}$-interpretation (representing the pseudo-model). Recall that in Section 2, we had associated an undirected graph G_q with q. We assume that $\mathcal{I}$ is equipped with a set of ensembles. Let h be a homomorphism from q to $\mathcal{I}$. We aim to identify a condition on h ('tameness') that allows us to obtain from h a homomorphism into the interpretation obtained by unraveling the pseudo-model $\mathcal{I}$ into an infinite tree-like interpretation.

We associate with h an equivalence relation $\sim_h$ on $\mathsf{var}(q)$ by setting $x \sim_h y$ if G_q contains a path $x = z_1, \ldots, z_n = y$ such that $h(z_1), \ldots, h(z_n)$ are all part of the same ensemble in $\mathcal{I}$. Let $G_{h,\mathcal{I}}$ be the directed graph whose nodes are the equivalence classes of $\sim_h$ and which has an edge (c_1, c_2) if there is an $r(x_1, x_2) \in q$ with $x_1 \in c_1$ and $x_2 \in c_2$. A node c of $G_{h,\mathcal{I}}$ is an *ensemble node* if there is an (equivalently: for all) $x \in c$ such that $h(x)$ is part of an ensemble in $\mathcal{I}$. We recommend to the reader to verify that all nodes that are not ensemble nodes are singleton classes. We say that h is *tame* if the following conditions are satisfied:

1. $G_{h,\mathcal{I}}$ is a tree, possibly with self-loops on ensemble nodes;
2. for all edges (c_1, c_2) in $G_{h,\mathcal{I}}$, there are $d_1, d_2 \in \Delta^\mathcal{I}$ such that for all $r(x_1, x_2) \in q$ with $x_1 \in c_1$ and $x_2 \in c_2$, we have $h(x_1) = d_1$ and $h(x_2) = d_2$.

Condition 2 reflects the fact that the tree-like interpretations $\mathcal{I}_L$ satisfy the following property: if d_i is an element in ensemble e_i, for $i \in \{1, 2\}$, and there is a role edge $(d_1, d_2) \in r^{\mathcal{I}_L}$, then d_1, d_2 are *unique* with this property.

An answer $\bar{d} \in q(\mathcal{I})$ is *tame* if there is a tame homomorphism h from $q(\bar{x})$ to $\mathcal{I}_L$ with $h(\bar{x}) = \bar{d}$. An A-interpretation $\mathcal{I}$ being a *pseudo-model* of $\mathcal{O}$ is defined in the same way as being a model of $\mathcal{O}$ except that in the semantics of concept abstractions, answers to a CQ q on an interpretation $\mathcal{I}_L$ are replaced with tame answers.

A central observation underlying the subsequent algorithm is that we can always find pseudo-models in which each maximal connected component has size polynomial in $||\mathcal{O}||$. Formally, a *maximal connected component (MCC)* of an A-interpretation $\mathcal{I}$ is an $\mathcal{EL}$-interpretation that can be obtained as follows: choose an abstraction level L, then choose a maximal subset $\Delta \subseteq \Delta^{\mathcal{I}_L}$ such that the following undirected graph is connected:

$$(\Delta, \{\{d, e\} \mid (d, e) \in r^{\mathcal{I}_L} \text{ for some role name } r \text{ or } d, e \text{ in } \bar{e} \text{ for some } L\text{-ensemble } \bar{e}\});$$

and finally take the restriction of $\Delta^{\mathcal{I}_L}$ to domain Δ. In the appendix, we prove the following.

Lemma 4. *If C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$, then there is an L_0-pseudo-model $\mathcal{I}$ of C_0 and $\mathcal{O}$ such that each MCC of $\mathcal{I}$ has at most $2 \cdot (||\mathcal{O}||^2 + ||\mathcal{O}||)$ elements.*

Our proof of Lemma 4 is rather laborious. The reason is that the structure of the universal models as constructed in Section 4.2 turns out to be surprisingly hard (and, to us, actually infeasible) to analyze. This is mainly due to the application of the filtration construction after chase termination. To avoid such an analysis, we first introduce another, more semantic construction of universal models. In this construction, we start from the universal models from Section 4.2 and 'combine small pieces of them' in a uniform, tree-like way. The structure of the resulting universal models, which we call *uniform*, is clear by definition. In particular, each $\mathcal{EL}$-interpretation $\mathcal{I}_L$ is a tree of ensembles / non-ensemble nodes, as described above. Starting from uniform universal models, we can then carefully craft pseudo-models by selecting ensembles and non-ensemble elements and 'rerouting' role edges.

The Algorithm The aim of our algorithm is to verify the existence of a pseudo-model of C_0 and $\mathcal{O}$, as per Lemma 4. To represent MCCs of that pseudo-model, we use mosaics. A *mosaic* is a tuple $M = (\mathcal{I}, L, E, \bar{e})$ that consists of

1. a model $\mathcal{I}$ of $\mathcal{O}_L$ such that $|\Delta^\mathcal{I}| \leq 2 \cdot (||\mathcal{O}||^2 + ||\mathcal{O}||)$,
2. an abstraction level $L \in \mathbf{A}_\mathcal{O}$,
3. a set E of non-overlapping ensembles, that is, non-empty tuples over $\Delta^\mathcal{I}$ that do not share elements, and
4. a tuple $\bar{e}$ over $\Delta^\mathcal{I}$ with $\bar{e} \in E$ or $\bar{e} = ()$.

We may write $\mathcal{I}^M$ for $\mathcal{I}$, and likewise for L^M, E^M, and $\bar{e}^M$. We further define a function

- $Q^{\mathsf{ref}}_{M,L'}$ that maps each $d \in \Delta^\mathcal{I}$ to the set of CQs $Q^{\mathsf{ref}}_{M,L'}(d) = \{q \mid L'{:}q(\bar{x}) \text{ \underline{refines} } L{:}A \in \mathcal{O} \text{ and } d \in A^\mathcal{I}\}$;
- $T^{\mathsf{abs}}_{M,L'}$ that maps each $\bar{d} \in E$ to the set of concept names $T^{\mathsf{abs}}_{M,L'}(\bar{d}) = \{A \mid L'{:}A \text{ \underline{abstracts} } L{:}q(\bar{x}) \in \mathcal{O} \text{ and } \bar{d} \in q(\mathcal{I}) \text{ is tame}\}$.

Note that, as mosaics are equipped with an explicit set E of ensembles, it is clear what we mean by a tame answer.

Our algorithm is now listed as Algorithm 1. In Line 3, we guess a set X_L of sets of concept names. This is related to the first challenge mentioned above and the idea is that for

Algorithm 1 Algorithm for satisfiability in $\mathcal{EL}^{\mathsf{abs}}$[cr,ca]

procedure $\mathcal{EL}$[cr,ca]-SAT(C_0, L_0)
  for all $L \in \mathbf{A}_\mathcal{O}$ **do**
    Guess a set $X_L \in 2^{2^{\mathbf{C}}}$ of sets concept names such that $|X_L| \leq ||\mathcal{O}||^3 + ||\mathcal{O}||^2$
  Guess a mosaic M such that $L^M = L_0$, $C_0^{\mathcal{I}^M} \neq \emptyset$, and $e^M = ()$
  $R \leftarrow$ RECURSE(M)
  for all $L \in \mathbf{A}_\mathcal{O}$ and $T \in X_L$ **do**
    Guess a mosaic M such that $L^M = L$, $(\bigsqcap T)^{\mathcal{I}^M} \neq \emptyset$, and $e^M = ()$
    $R \leftarrow R \wedge$ RECURSE(M)
  return R
procedure RECURSE$(M = (\mathcal{I}, L, E, \bar{e}))$
  for all $d \in \Delta^\mathcal{I}$ and $L' \in \mathbf{A}_\mathcal{O}$ s.t. $Q^{\mathsf{ref}}_{M,L'}(d) \neq \emptyset$ **do**
    Guess a mosaic $M' = (\mathcal{I}', L', E', \bar{e}')$
    if $\bar{e}' \notin q(\mathcal{I}')$ for some $q \in Q^{\mathsf{ref}}_{M,L'}(d)$ or $d \notin A^\mathcal{I}$ for some $A \in T^{\mathsf{abs}}_L(\bar{e}')$ **then return** false
    RECURSE(M')
  for all $L'{:}A$ abstracts $L{:}q(\bar{x}) \in \mathcal{O}$ and all tame answers $\bar{d} \in q(\mathcal{I})$ with $\bar{d} \neq \bar{e}$ **do**
    if $\bar{d} \notin E$ **then return** false
    Guess a set $T' \in X_{L'}$
    if $T^{\mathsf{abs}}_{M,L'}(\bar{d}) \not\subseteq T'$ or $\bar{d} \notin q(\mathcal{I})$ for some $L'{:}q(\bar{x})$ refines $L{:}A \in \mathcal{O}$ with $A \in T'$ **then return** false
  return true

every set $S \in X_L$, there must be an element on level L that satisfies all concept names in S, and that (copies of) these elements can be used to satisfy all abstractions that ever require a witness during the run of the algorithm. Intuitively, the algorithm repeatedly guesses mosaics and makes recursive calls to satisfy refinement statements from the ontology. More precisely, it is the tuple $\bar{e}^M$ in the fourth component of a mosaic M that, if not empty, is the ensemble which satisfies the refinement.

Lemma 5. *The algorithm accepts iff C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$.*

It is easy to see that the recursion depth of the algorithm is bounded by $|\mathbf{A}_\mathcal{O}|$ and that only a polynomial amount of space is consumed. Needless to say, we can eliminate nondeterminism by applying Savitch's theorem.

5 Lower Bounds

We prove lower complexity bounds that match the upper bounds presented in Section 4. We start with the following.

Theorem 3. *Satisfiability is*

1. CONP-*hard in* $\mathcal{EL}^{\mathsf{abs}}$[cr] *and*
2. PSPACE-*hard in* $\mathcal{EL}^{\mathsf{abs}}$[cr, ca].

The proofs of Points 1 and 2 of Theorem 3 are closely related. We start with Point 1, which is proved by reduction from unsatisfiability in propositional logic.

Let φ be a propositional formula that uses variables $p_1, \dots, p_n$ and only the junctors $\neg$ and $\wedge$. Let $\mathsf{sub}(\varphi)$ be the set of all subformulas of φ. We construct an $\mathcal{EL}^{\mathsf{abs}}$[cr]-ontology $\mathcal{O}$ that uses the following concept and role names:

- T_ψ and F_ψ, for each $\psi \in \mathsf{sub}(\varphi)$, to represent that ψ evaluates to true or false;
- P_i and $\overline{P}_i$, for $i \in \{1, \dots, n\}$, to represent assigning true or false to p_i.

The ontology $\mathcal{O}$ uses the abstraction levels $L_0 \succ \cdots \succ L_n$. When refining from L_i to L_{i+1}, we introduce two domain elements that represent the two possible truth assignments for variable p_{i+1}. This is achieved by including in $\mathcal{O}$ the following for $1 \leq i \leq n$:

$$L_i{:}\, P_i(x_1) \wedge \overline{P}_i(x_2) \ \underline{\mathsf{refines}}\ L_{i-1}{:}\, \top. \tag{1}$$

If desired, it is easy to make the query connected. To preserve the truth assignments to variables on finer levels, we add for $1 \leq i < n$ and $i < j < n$:

$$L_{j+1}{:}P_i(x_1) \wedge P_i(x_2) \ \underline{\mathsf{refines}}\ L_j{:}P_i \tag{2}$$

$$L_{j+1}{:}\overline{P}_i(x_1) \wedge \overline{P}_i(x_2) \ \underline{\mathsf{refines}}\ L_j{:}\overline{P}_i \tag{3}$$

This generates a binary tree of refinements of depth n, representing all possible truth assignments at the leaves, that is, by domain elements on level L_n. We evaluate φ on all these truth assignments and generate an inconsistency if φ ever evaluates to true:

$$P_i \sqsubseteq_{L_n} T_{p_i} \text{ and } \overline{P}_i \sqsubseteq_{L_n} F_{p_i} \text{ for } 1 \leq i \leq n \tag{4}$$

$$T_\psi \sqsubseteq_{L_n} F_{\neg\psi} \text{ and } F_\psi \sqsubseteq_{L_n} T_{\neg\psi} \text{ for all } \neg\psi \in \mathsf{sub}(\varphi) \tag{5}$$

and for all $\psi = \psi_1 \wedge \psi_2 \in \mathsf{sub}(\varphi)$:

$$T_{\psi_1} \sqcap T_{\psi_2} \sqsubseteq_{L_n} T_\psi \quad F_{\psi_1} \sqsubseteq_{L_n} F_\psi \quad F_{\psi_2} \sqsubseteq_{L_n} F_\psi \tag{6}$$

and finally:

$$T_\varphi \sqsubseteq_{L_n} \bot. \tag{7}$$

Lemma 6. *φ is unsatisfiable iff $\top$ is L_0-satisfiable w.r.t. $\mathcal{O}$.*

Point 2 of Theorem 3 is proved by reduction from the validity of quantified Boolean formulas (QBFs) of the form $\varphi_0 = Q_1 p_1 \cdots Q_n p_n \varphi$ with $Q_i \in \{\exists, \forall\}$ and φ a propositional formula that uses only the variables p_1 to p_n and the junctors $\neg$ and $\wedge$ (Arora and Barak 2009). We construct an $\mathcal{EL}^{\mathsf{abs}}$[cr, ca]-ontology $\mathcal{O}$ such that φ_0 is valid if and only if $\top$ is L_0-satisfiable w.r.t. $\mathcal{O}$.

The construction of $\mathcal{O}$ may be viewed as an extension of the construction from the previous reduction. In particular, we use the same concept and role names, plus a concept name F and a role name s. We next give details. To construct $\mathcal{O}$, we reuse statements (1) to (6) from the previous reduction, adding an $s(x_1, x_2)$-atom to the query in refinements (1) to (3). As in the previous reduction, this generates a full binary tree of refinements of depth n that represents all truth assignments as domain elements on level L_n. We next implement a bottom-up pass on this tree that evaluates the quantifiers in φ_0 using the concept name F. We first add to $\mathcal{O}$:

$$F_\varphi \sqsubseteq_{L_n} F. \tag{8}$$

For each $i \in \{1, \dots, n\}$ with $Q_i = \forall$ and $j \in \{1, 2\}$, we further add the following concept abstraction:

$$L_{i-1}\colon F \ \underline{\mathsf{abstracts}}\ L_i\colon q(x_1, x_2) \text{ where} \quad (9)$$
$$q(x_1, x_2) = F(x_j) \wedge s(x_1, x_2)$$

and for each $i \in \{1, \dots, n\}$ with $Q_i = \exists$, we add the concept abstraction

$$L_{i-1}\colon F \ \underline{\mathsf{abstracts}}\ L_i\colon q(x_1, x_2) \text{ where} \quad (10)$$
$$q(x_1, x_2) = F(x_1) \wedge F(x_2) \wedge s(x_1, x_2)$$

Note that, as required and due to our use of the role name s, the queries in these abstraction statements are connected. This is in fact the only reason why s was introduced. Finally, we add the following CI, representing our wish that φ_0 is valid:

$$F \sqsubseteq_{L_0} \bot. \quad (11)$$

Lemma 7. *φ_0 is valid iff $\top$ is L_0-satisfiable w.r.t. $\mathcal{O}$.*

Finally, we prove the following.

Theorem 4. *Satisfiability in $\mathcal{EL}^{\mathsf{abs}}[\mathrm{rr}]$ is 2EXPTIME-hard.*

This is achieved by reducing the word problem for exponentially space-bounded alternating Turing machines. More precisely, we adapt a reduction from (Lutz and Schulze 2023) used there to show that satisfiability in $\mathcal{ALC}^{\mathsf{abs}}[\mathrm{rr}]$ is 2EXPTIME-hard.

6 Getting To Polynomial Time

We now consider a semantic variation of $\mathcal{ELH}_r^{\mathsf{abs}}[\mathrm{cr}]$ that reduces the complexity of satisfiability from coNP to PTime. This variation is obtained by letting L-ensembles be *sets* rather than tuples, that is, dropping the order of elements in the ensemble. Moreover, refinements are now interpreted as a *partial* description of an ensemble, that is, the variables in the CQ used in the refinement describe elements of the ensemble that must exist, but other elements may exist as well.

In more detail, the refinement function ρ is now a partial function that associates every pair $(d, L) \in \Delta^{\mathcal{I}} \times \mathbf{A}_{\mathcal{I}}$ such that $L \prec L(d)$ with a non-empty subset of $\Delta^{\mathcal{I}_L}$ called an *L-ensemble*. We still require that every object participates in at most one ensemble, that is, Property $(*)$ from the original definition of the semantics is still required to be satisfied. The semantics of refinement statements is then as follows. An A-interpretation $\mathcal{I} = (\mathbf{A}_{\mathcal{I}}, \prec, (\mathcal{I}_L)_{L \in \mathbf{A}_{\mathcal{I}}}, \rho)$ *satisfies* a

- concept refinement $L{:}q(\bar{x})$ $\underline{\mathsf{refines}}$ $L'{:}C$ if $L, L' \in \mathbf{A}_{\mathcal{I}}$ and for all $d \in C^{\mathcal{I}_{L'}}$, there is an $\bar{e} \in q(\mathcal{I}_L)$ s.t. all elements of $\bar{e}$ are in $\rho_L(d)$;
- role refinement $L{:}q(\bar{x}, \bar{y})$ $\underline{\mathsf{refines}}$ $L'{:}q_r(x, y)$ if $L, L' \in \mathbf{A}_{\mathcal{I}}$ and for all $(d_1, d_2) \in q_r(\mathcal{I}_{L'})$, there is an $(\bar{e}_1, \bar{e}_2) \in q(\mathcal{I}_L)$ s.t. all elements of $\bar{e}_i$ are in $\rho_L(d_i)$, for $i \in \{1, 2\}$.

We call this semantics the *set ensemble semantics*. Note that under this semantics, we can no longer simulate $\bot$ by concept refinement. We instead assume that $\bot$ is explicitly available as a concept constructor (to ensure that satisfiability and subsumption are mutually reducible).

The following example illustrates the impact of switching to set ensemble semantics which, we believe, is fairly mild if the modeling discipline is adjusted in a suitable way.

Example 4. *Consider the following ontology $\mathcal{O}$:*

$$\mathsf{SportsCar} \sqsubseteq \mathsf{Car}$$
$$L_1{:}\mathsf{Engine}(x_1) \wedge \mathsf{Body}(x_2)\ \underline{\mathsf{refines}}\ L_2{:}\mathsf{Car}$$
$$L_1{:}\mathsf{TurboEngine}(x_1) \wedge \mathsf{Body}(x_2)\ \underline{\mathsf{refines}}\ L_2{:}\mathsf{SportsCar}$$

Under the standard semantics, every sports car refines into an ensemble of exactly two elements, the first one both an engine and a turbo engine, and the second one a body. Under set ensemble semantics, a sports car may refine into an ensemble of three elements: an engine, a turbo engine, and a body. If we add the natural concept inclusion

$$\mathsf{TurboEngine} \sqsubseteq \mathsf{Engine},$$

then the turbo engine is also an engine and, arguably, the difference between the two semantics becomes negligible.

We aim to prove the following.

Theorem 5. *Under the set ensemble semantics, satisfiability in $\mathcal{EL}^{\mathsf{abs}}[\mathrm{cr}, \mathrm{rr}]$ is in PTime.*

We prove Theorem 5 by providing a polynomial time reduction from L-satisfiability in $\mathcal{ELH}_r^{\mathsf{abs}}[\mathrm{cr}]$ to satisfiability in $\mathcal{ELHO}_{r,\bot}$, the extension of $\mathcal{ELH}_r$ with nominals and $\bot$. More precisely, we assume a countably infinite set $\mathbf{I}$ of individuals and admit expressions $\{a\}$, with $a \in \mathbf{I}$, as concepts. The semantics is given by $\{a\}^{\mathcal{I}} = a$ for all interpretations $\mathcal{I}$. It is known that satisfiability in $\mathcal{ELHO}_{r,\bot}$ is in PTime (Krötzsch 2010).

Let C_0 be an $\mathcal{EL}$-concept, $\mathcal{O}$ an $\mathcal{ELH}_r^{\mathsf{abs}}[\mathrm{cr}]$-ontology in normal form, and $L_0 \in \mathbf{A}_{\mathcal{O}}$, given as input. We assume w.l.o.g. that no two CQs in (refinements in) $\mathcal{O}$ share a variable. If $G_{\mathcal{O}}$ is not a tree, we may directly return 'unsatisfiable'. Otherwise, we construct in polynomial time an $\mathcal{ELHO}_{r,\bot}$-ontology $\mathcal{O}'$. Introduce a fresh role name r_L for each role name r in $\mathcal{O}$ and each abstraction level L in $\mathcal{O}$, and an additional fresh role name u. We include in $\mathcal{O}'$ the following concept inclusions:

1. $\top \sqsubseteq \exists u.L$ for all $L \in \mathbf{A}_{\mathcal{O}}$;
2. $L \sqcap A_1 \sqcap \cdots \sqcap A_n \sqsubseteq B$ for all $A_1 \sqcap \cdots \sqcap A_n \sqsubseteq_L B$ in $\mathcal{O}$ (with $A_1 \sqcap \cdots \sqcap A_n = \top$ if $n = 0$);
3. $\exists r_L.A \sqsubseteq B$ for all $\exists r.A \sqsubseteq_L B$ in $\mathcal{O}$;
4. $L \sqcap A \sqsubseteq \exists r_L.(L \sqcap B)$ for all $A \sqsubseteq_L \exists r.B$ in $\mathcal{O}$;
5. for all $L{:}q(\bar{x})$ $\underline{\mathsf{refines}}$ $L'{:}A$ in $\mathcal{O}$:
    - $L' \sqcap A \sqsubseteq \exists u.(L \sqcap B \sqcap \{a_x\})$ whenever $B(x)$ is an atom in q;
    - $L' \sqcap A \sqsubseteq \exists u.(L \sqcap \{a_x\} \sqcap \exists r_L.(L \sqcap \{a_y\}))$ whenever $r(x, y)$ is an atom in q;
6. for all $L{:}q(\bar{x}, \bar{y})$ $\underline{\mathsf{refines}}$ $L'{:}q_r(x, y)$ in $\mathcal{O}$ with $q_r = A_1(x) \wedge r(x, y) \wedge A_2(y)$:
    - $A_1 \sqcap \exists r_{L'}.A_2 \sqsubseteq \exists u.(L \sqcap B \sqcap \{a_x\})$ whenever $B(x)$ is an atom in q;
    - $A_1 \sqcap \exists r_{L'}.A_2 \sqsubseteq \exists u.(L \sqcap \{a_x\} \sqcap \exists s_L.(L \sqcap \{a_y\}))$ whenever $s(x, y)$ is an atom in q;

Moreover, $\mathcal{O}'$ contains the following:

7. the role inclusion $r_L \sqsubseteq s_L$ for every role inclusion $r \sqsubseteq_L s$ in $\mathcal{O}$;

8. the range restriction $\top \sqsubseteq \forall r_L.C$ for every range restriction $\top \sqsubseteq_L \forall r.C$ in $\mathcal{O}$.

Lemma 8. *C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$ under set ensemble semantics iff $C_0 \sqcap L_0$ is satisfiable w.r.t. $\mathcal{O}'$.*

This proves Theorem 5. An extension to the case that includes concept or role abstractions is not easily possible. In fact, it is straightforward to prove the following by a reduction from the homomorphism problems on directed graphs (the semantics of concept abstractions is defined as expected).

Theorem 6. *Under the set ensemble semantics, satisfiability is coNP-hard in $\mathcal{EL}^{\mathsf{abs}}[\mathrm{cr}, \mathrm{ca}]$.*

We remark that exactly the same reduction as given in this section also serves to reduce satisfiability in $\mathcal{ALC}^{\mathsf{abs}}[\mathrm{cr}, \mathrm{rr}]$ under set ensemble semantics to satisfiability in $\mathcal{ALCO}$, the extension of $\mathcal{ALC}$ with nominals. The latter problem is ExpTime-complete (Tobies 2001), which explains the entry for $\mathcal{ALC}^{\mathsf{abs}}$ under set ensemble semantics in Figure 1.

7 Conclusion

We have studied description logics of refinement and abstraction based on members of the $\mathcal{EL}$ family. While, compared to the $\mathcal{ALC}$ version, the computational complexity does not drop for the full logic, we have identified natural fragments where it does. We leave the complexity of other (less natural) fragments such as $\mathcal{ELH}_r^{\mathsf{abs}}[\mathrm{ca}]$ and $\mathcal{ELH}_r^{\mathsf{abs}}[\mathrm{ca}, \mathrm{ra}]$ as future work.

It would be interesting to consider DLs of abstraction and refinement based on the extension $\mathcal{ELI}$ of $\mathcal{EL}$ with inverse roles. Then already reasoning in the base logic is ExpTime-hard so we cannot expect any lower complexities. One might also define ontology languages with abstraction and refinement based on existential rules, see e.g. (Baget *et al.* 2011; Calì *et al.* 2010). It is then natural to extend the arity of all relations by one position that represents the abstraction level. Note, however, that since every object is required to refine only into a single ensemble, it does not seem possible to encode abstraction and refinement into existing (decidable) existential rule formalisms in a simple way.

Acknowledgments

The research reported in this paper has been supported by the German Research Foundation DFG, as part of Collaborative Research Center (Sonderforschungsbereich) 1320 Project-ID 329551904 "EASE - Everyday Activity Science and Engineering", University of Bremen (http://www.ease-crc.org/). The research was conducted in subproject "P02 - Ontologies with Abstraction".

This work is partly supported by BMBF (Federal Ministry of Education and Research) in DAAD project 57616814 (SECAI, School of Embedded Composite AI) as part of the program Konrad Zuse Schools of Excellence in Artificial Intelligence.

References

Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.

Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proc. of IJCAI*, pages 364–369. Professional Book Center, 2005.

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

Jean-François Baget, Marie-Laure Mugnier, Sebastian Rudolph, and Michaël Thomazo. Walking the complexity lines for generalized guarded existential rules. In *Proc. of IJCAI*, pages 712–717. IJCAI/AAAI, 2011.

Thomas Bittner and Barry Smith. A theory of granular partitions. In *Foundations of Geographic Information Science*, pages 117–149. Taylor & Francis, 2003.

Silvia Calegari and Davide Ciucci. Granular computing applied to ontologies. *Int. J. Approx. Reason.*, 51(4):391–409, 2010.

Andrea Calì, Georg Gottlob, Thomas Lukasiewicz, Bruno Marnette, and Andreas Pieris. Datalog+/-: A family of logical knowledge representation and query languages for new applications. In *Proc. of LICS*, pages 228–242. IEEE Computer Society, 2010.

Ashok K Chandra, Dexter C Kozen, and Larry J Stockmeyer. Alternation. *Journal of the ACM (JACM)*, 28(1):114–133, 1981.

Gianluca Cima, Marco Console, Maurizio Lenzerini, and Antonella Poggi. Monotone abstractions in ontology-based data management. In *Proc. of AAAI*, pages 5556–5563. AAAI Press, 2022.

Birte Glimm, Yevgeny Kazakov, and Trung-Kien Tran. Ontology materialization by abstraction refinement in horn SHOIF. In *Proc. of AAAI*, pages 1114–1120. AAAI Press, 2017.

Szymon Klarman and Víctor Gutiérrez-Basulto. Description logics of context. *J. Log. Comput.*, 26(3):817–854, 2016.

Markus Krötzsch. Efficient inferencing for OWL EL. In *Proc. of JELIA*, volume 6341 of *LNCS*, pages 234–246. Springer, 2010.

Francesca A. Lisi and Corrado Mencar. A granular computing method for OWL ontologies. *Fundam. Informaticae*, 159(1-2):147–174, 2018.

Carsten Lutz and Lukas Schulze. Description logics with abstraction and refinement. In *Proc. of KR*, pages 492–501, 2023.

Carsten Lutz, David Toman, and Frank Wolter. Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In *Proc. of IJCAI*. AAAI Press, 2009.

Boris Motik, Bernardo Cuenca Grau, Ian Horrocks, Zhe Wu, Achille Fokoue, and Carsten Lutz, editors. *OWL 2 Web Ontology Language Profiles (Second Edition)*. W3C Recommendation, 2009. Available at https://www.w3.org/TR/owl2-profiles/.

Christos H Papadimitriou. Computational complexity. In *Encyclopedia of computer science*, pages 260–265. 2003.

Stephan Tobies. *Complexity results and practical algorithms for logics in knowledge representation*. PhD thesis, RWTH Aachen University, Germany, 2001.

Frank Wolter and Michael Zakharyaschev. Multi-dimensional description logics. In *Proc. of IJCAI*, pages 104–109. Morgan Kaufmann, 1999.

A Proofs for Section 4.2

A.1 Missing Rules

We list the chase rules that treat role refinement and role abstraction, omitted in the main body of the paper:

R11 if h is a homomorphism from q_r to F_L for a role refinement $L'{:}q(\bar{x}, \bar{y})$ refines $L{:}q_r(x, y)$ in $\mathcal{O}$ then

- if $\rho_{L'}(h(x))$ is undefined, then set $\rho_{L'}(h(x)) = \bar{a}$ for a tuple $\bar{a}$ of fresh constants with $|\bar{a}| = |\bar{x}|$;
- if $\rho_{L'}(h(y))$ is undefined, then set $\rho_{L'}(h(y)) = \bar{b}$ for a tuple $\bar{b}$ of fresh constants with $|\bar{b}| = |\bar{y}|$;

R12 if h is a homomorphism from q_r to F_L for a role refinement $L'{:}q(\bar{x}, \bar{y})$ refines $L{:}q_r(x, y)$ in $\mathcal{O}$, $\rho_{L'}(h(x))$ and $\rho_{L'}(h(y))$ are defined, $|\bar{x}| = |\rho_{L'}(h(x))|$, and $|\bar{y}| = |\rho_{L'}(h(y))|$, then add $q(\rho_{L'}(h(x)), \rho_{L'}(h(y)))$ to $F_{L'}$; if $|\bar{x}| \neq |\rho_{L'}(h(x))|$ or $|\bar{y}| \neq |\rho_{L'}(h(y))|$, then return 'unsatisfiable';

R13 if h is a homomorphism from q to F_L for a role abstraction $L'{:}r$ abstracts $L{:}q(\bar{x}, \bar{y})$ in $\mathcal{O}$ then

- if there is no $a \in \mathsf{dom}(F_{L'})$ with $\rho_L(a) = h(\bar{x})$, then set $\rho_L(a) = h(\bar{x})$ for a fresh constant a;
- if there is no $b \in \mathsf{dom}(F_{L'})$ with $\rho_L(b) = h(\bar{y})$, then set $\rho_L(b) = h(\bar{y})$ for a fresh constant b;

R14 if h is a homomorphism from q to F_L for any role abstraction $L'{:}r$ abstracts $L{:}q(\bar{x}, \bar{y})$ in $\mathcal{O}$ and there are $a, b \in \mathsf{dom}(F_{L'})$ with $\rho_L(a) = h(\bar{x})$ and $\rho_L(b) = h(\bar{y})$, then add $r(a, b)$ to $F_{L'}$.

Soundness. We prove the two directions of Lemma 2 separately, starting with soundness.

Lemma 9. *Let $\mathcal{O}$ be an $\mathcal{ELH}_r^{\mathsf{abs}}$-ontology in normal form whose abstraction graph $G_\mathcal{O}$ is a tree, C_0 a concept name, and L_0 an abstraction level. If the L_0-chase on C_0 and $\mathcal{O}$ does not abort, then C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$.*

Let $I^* = (F^*, \rho^*, \sim^*)$ be the result of chasing C_0 w.r.t. L_0 and $\mathcal{O}$. To construct an L_0-model $\mathcal{I}$ of C_0 and $\mathcal{O}$, define $\mathcal{I} = (\mathbf{A}_\mathcal{I}, \prec, (\mathcal{I}_L)_{L \in \mathbf{A}_\mathcal{I}}, \rho)$ where

$$\begin{aligned}
\Delta^{\mathcal{I}_L} &= \{[a] \mid a \in \mathsf{dom}(F_L^*)\} \\
A^{\mathcal{I}_L} &= \{[a] \mid A(a') \in F_L^* \text{ and } a' \in [a]\} \\
r^{\mathcal{I}_L} &= \{([a], [b]) \mid r(a', b') \in F_L^* \text{ and } a' \in [a], b' \in [b]\} \\
\rho_L &= \{([a], ([b_1] \cdots [b_n])) \mid (a', (b_1' \cdots b_n') \in \rho_L^* \text{ with} \\
&\qquad a' \in [a], b_i' \in [b_i] \text{ for } 1 \leq i \leq n\}.
\end{aligned}$$

The remaining components $\mathbf{A}_\mathcal{I}$ and $\prec$ are defined as $\mathbf{A}_\mathcal{O}$ and $\prec_\mathcal{O}$, respectively. We remark that if $\mathcal{O}$ contains no abstraction statements, then the equivalence relation $\sim$ in the chase is just the identity function since rules R9, R10 and R13 to R19 are never applicable. It follows that the filtration step is not needed; in other words, the result $\mathcal{I}$ of filtrating I^* is simply I^* viewed as an interpretation.

We first verify that $\mathcal{I}$ is actually an A-interpretation.

Lemma 10. *$\mathcal{I}$ is an A-interpretation.*

Proof. We need to show the following:

1. The directed graph $(\mathbf{A}_\mathcal{I}, \prec)$ is a tree.
   Clear by definition of $\mathcal{I}$.
2. ρ is a partial function.
   Assume that $([a], ([b_1], \ldots, [b_n])), ([a], ([c_1], \ldots, [c_m]))$ are in ρ. Then there are $a_1, a_2 \in [a]$ such that $\rho_L^*(a_1) = (b_1', \ldots, b_n')$, $\rho_L^*(a_2) = (c_1', \ldots, c_m')$, $b_i' \sim b_i$ for $1 \leq i \leq n$, and $c_i' \sim c_i$ for $1 \leq i \leq m$. By R15 and since the chase does not abort, we have $n = m$. By R17, we have $b_i' \sim c_i'$ for $1 \leq i \leq n$. Consequently, $([b_1], \ldots, [b_n]) = ([c_1], \ldots, [c_m])$, as required.
3. No element in ρ is part of two distinct ensembles.
   Assume that $\rho_L([a_1]) = ([b_1], \ldots, [b_n])$ and $\rho_L([a_2]) = ([c_1], \ldots, [c_m])$ with $[b_\ell] = [c_k]$. Then there are $a_1' \in [a_1]$ and $a_2' \in [a_2]$ such that $\rho_L^*(a_1') = (b_1', \ldots, b_n')$, $\rho_L^*(a_2') = (c_1', \ldots, c_m')$, $b_i' \sim b_i$ for $1 \leq i \leq n$, and $c_i' \sim c_i$ for $1 \leq i \leq m$. From $[b_\ell] = [c_k]$, we obtain $b_\ell' \sim c_k'$. Consequently, R16 yields $a_1 \sim a_2$, thus $[a_1] = [a_2]$ as required. ❑

It remains to show the following.

Lemma 11. *$\mathcal{I}$ is an L_0-model of C_0 and $\mathcal{O}$.*

Proof. The first interpretation candidate in any chase sequence contains the fact $C_0(a_0)$ on level L_0, and thus $C_0(a_0) \in F_L^*$. By construction of $\mathcal{I}$ and since C_0 is a concept name, this implies $[a_0] \in C_0^{\mathcal{I}_{L_0}}$.

What remains to be shown is that all inclusions, range restrictions, abstractions and refinements of $\mathcal{O}$ are satisfied. We first observe that the following is implied by R18.

Claim 1. For any $L \in \mathbf{A}_\mathcal{O}$:

1. if $A(a) \in F_L^*$ and $a \sim a'$, then $A(a') \in F_L^*$;
2. if $r(a, b) \in F_L^*$, $a \sim a'$, and $b \sim b'$, then $r(a', b') \in F_L^*$.

We now consider inclusions and range restrictions:

- if $[a] \in \Delta^{\mathcal{I}_L}$ and $\top \sqsubseteq_L A \in \mathcal{O}$, then by definition of $\mathcal{I}$ we have $a \in \mathsf{dom}(F_L^*)$. R2 then implies $A(a) \in F_L^*$ and thus $[a] \in A^{\mathcal{I}_L}$ by definition of $\mathcal{I}$;
- if $[a] \in (A_1 \sqcap \cdots \sqcap A_n)^{\mathcal{I}_L}$ and $A_1 \sqcap \cdots \sqcap A_n \sqsubseteq_L B \in \mathcal{O}$, then by definition of $\mathcal{I}$ there are $a_1, \ldots, a_n \in [a]$ with $A_1(a_1), \ldots, A_n(a_n) \in F_L^*$. Point 1 of Claim 1 now implies $A_1(a), \ldots, A_n(a) \in F_L^*$. With R1 we get $B(a) \in F_L^*$ and the definition of $\mathcal{I}$ lets us obtain $[a] \in B^{\mathcal{I}_L}$;
- if $[a] \in A^{\mathcal{I}_L}$ and $A \sqsubseteq_L \exists r.B \in \mathcal{O}$, then by definition of $\mathcal{I}$ there is a $a' \in [a]$ with $A(a') \in F_L^*$. Point 1 of Claim 1 now implies $A(a) \in F_L^*$. With R3 we get $B(b), r(a, b) \in F_L^*$ for some constant $b \in \mathsf{dom}(F_L)$. The definition of $\mathcal{I}$ lets us obtain $([a], [b]) \in r^{\mathcal{I}_L}$ and $[b] \in B^{\mathcal{I}_L}$, as required;
- if $[a] \in (\exists r.A)^{\mathcal{I}_L}$ and $\exists r.A \sqsubseteq_L B \in \mathcal{O}$, then by the semantics there is a $[b] \in A^{\mathcal{I}_L}$ with $([a], [b]) \in r^{\mathcal{I}_L}$. By definition of $\mathcal{I}$ there are $a' \in [a]$ and $b' \in [b]$ such that $r(a', b') \in F_L^*$. Point 1 of Claim 1 yields $A(b') \in F_L^*$. From R4 we get $B(a') \in F_L^*$. From the definition of $\mathcal{I}$ we obtain $[a] \in B^{\mathcal{I}_L}$, as required.
- if $([a], [b]) \in r^{\mathcal{I}_L}$ and $r \sqsubseteq_L s \in \mathcal{O}$, then by Claim 1 $r(a, b) \in F_L$. R5 then implies $s(a, b) \in F_L$ and thus $([a], [b]) \in s^{\mathcal{I}_L}$ by definition of $\mathcal{I}$;

- if $([a],[b]) \in r^{\mathcal{I}_L}$ and $\top \sqsubseteq_L \forall r.C \in \mathcal{O}$, then by Claim 1 $r(a,b) \in F_L$. R6 then implies $C(b) \in F_L$ and thus $[b] \in C^{\mathcal{I}_L}$ by definition of $\mathcal{I}$;

Next up are the refinement and abstraction statements. Let F_L be a set of facts of level L and $q(\bar{x})$ a CQ. Similar to normal interpretations, we call a tuple $\bar{e}$ an *answer* to q on F_L, if there is a homomorphism h from q to F_L with $h(\bar{x}) = \bar{e}$. We also use $q(F_L)$ to denote the set of answers to q on F_L.

First an intermediary claim.

Claim 2. Let $([a_1],\dots,[a_n]) \in q(\mathcal{I}_L)$. Choose any $a'_1 \in [a_1],\dots,a'_n \in [a_n]$. Then $(a'_1,\dots,a'_n) \in q(F^*_L)$.

Proof of claim. Let $A(x_i)$ be a concept atom in q. Then $[a_i] \in A^{\mathcal{I}_L}$. By definition of $\mathcal{I}$, there is an $b_i \in [a_i]$ such that $A(b_i) \in F^*_L$. By Point 1 of Claim 1 and since $a'_i \sim b_i$, we have $A(a'_i) \in F^*_L$, as required. For role atoms $r(x_i,x_j) \in q$, we can argue similarly based on Point 2 of Claim 1. This finishes the proof the claim.

Now we are ready to prove that abstractions and refinements are satisfied.

- Assume that $[a] \in A^{\mathcal{I}_L}$ and there is a concept refinement $L' {:}\, q(\bar{x})$ refines $L {:}\, A$ in $\mathcal{O}$. Then by definition of $\mathcal{I}$, there is an $a' \in [a]$ with $A(a') \in F^*_L$. R7 then implies that $\rho^*_L(a')$ is defined and R8 that $\rho^*_L(a') \in q(F^*_{L'})$. Hence $\rho_L([a])$ is defined by definition of ρ_L and $\rho_L([a]) \in q(\mathcal{I}_L)$ by definition of $A^{\mathcal{I}_L}$ and $r^{\mathcal{I}_L}$.
- Assume that $([a_1],\dots,[a_n])$ is an answer to q on $\mathcal{I}_L$ and there is a concept abstraction $L' {:}\, A$ abstracts $L {:}\, q(\bar{x})$ in $\mathcal{O}$. By Claim 2, we have $(a_1,\dots,a_n) \in q(F^*_L)$. R9 guarantees that there is an $a \in \mathsf{dom}(F^*_{L'})$ with $\rho^*_L(a) = (a_1,\dots,a_n)$ and R10 that $A(a) \in F^*_{L'}$. Hence $[a] \in A^{\mathcal{I}_L}$ and $\rho_L([a]) = ([a_1],\dots,[a_n])$, by definition of $\mathcal{I}$.
- Assume that $([a_1],[a_2]) \in q_r(\mathcal{I}_{L'})$ and there is a role refinement $L'{:}q(\bar{x},\bar{y})$ refines $L{:}q_r(x,y)$ in $\mathcal{O}$. By Claim 2, we have $(a_1,a_2) \in q_r(F^*_{L'})$. R11 guarantees that $\rho^*_L(a_1) = \bar{a}_1$ and $\rho^*_L(a_2) = \bar{a}_2$ are defined and R12 that $(\bar{a}_1,\bar{a}_2) \in q(F^*_L)$. Hence $\rho_L([a_1])$ and $\rho_L([a_1])$ are defined and an answer to q on $\mathcal{I}_L$, by definition of $\mathcal{I}$.
- Assume that $([a_1],\dots,[a_n],[b_1,\dots,b_m]) \in q(\mathcal{I}_L)$ and there is a role abstraction $L'{:}r$ abstracts $L{:}q(\bar{x},\bar{y})$ in $\mathcal{O}$ with $|\bar{x}| = n$ and $|\bar{y}| = m$. By Claim 2, we have $(a_1,\dots,a_n,b_1,\dots,b_n) \in q(F^*_L)$. R13 guarantees that $\rho^*_L(a) = (a_1,\dots,a_n)$ and $\rho^*_L(b) = (b_1,\dots,b_m)$ are defined for some constants $a,b \in \mathsf{dom}(F^*_{L'})$ and R14 that $r(a,b) \in F^*_{L'}$. Hence $\rho_L([a])$ and $\rho_L([b])$ are defined and $([a],[b]) \in r^{\mathcal{I}_{L'}}$, by definition of $\mathcal{I}$.

❑

Completeness. We want to show the following.

Lemma 12. *Let $\mathcal{O}$ be an $\mathcal{ELH}^{\mathsf{abs}}_r$-ontology in normal form whose abstraction graph $G_\mathcal{O}$ is a tree, C_0 a concept name, and L_0 an abstraction level. If C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$, then the chase does not abort.*

We start with defining the notion of a homomorphism from an interpretation candidate to an A-interpretation. These homomorphisms are very similar to the ones from A-interpretation to A-interpretation defined in the main part of the paper.

Let $I = (F,\rho,\sim)$ be an interpretation candidate. We use $\mathbf{A}_I$ to denote all the abstraction levels for which F is defined and $\prec_I$ for the smallest relation such that $L \prec_I L'$ if $\rho_L(a) = b$ and $L(a) = L'$. Let $\mathcal{I} = (\mathbf{A}_\mathcal{I},\prec,(\mathcal{I}_L)_{L\in\mathbf{A}_\mathcal{I}},\rho')$ be an A-interpretation. A function $h{:}\ \mathsf{dom}(F) \to \Delta^\mathcal{I}$ is a *homomorphism* from I to $\mathcal{I}$ if the following conditions are satisfied for all $L \in \mathbf{A}_I$:

1. $\prec_I \subseteq \prec$;
2. $A(a) \in F_L$ implies $h(a) \in A^{\mathcal{I}_L}$ for all $A \in \mathbf{C}$;
3. $A_\top(a) \in F_L$ implies $h(a) \in \Delta^{\mathcal{I}_L}$;
4. $r(a,b) \in F_L$ implies $(h(a),h(b)) \in r^{\mathcal{I}_L}$ for all $r \in \mathbf{R}$;
5. $\rho_L(a) = \bar{b}$ implies $\rho'_L(h(a)) = h(\bar{b})$;
6. $a_1 \sim a_2$ implies $h(a_1) = h(a_2)$.

We now show an intermediary lemma that captures the most notable property of the chase, that is, the result of the chase can be found inside any model that satisfies the given concept.

Lemma 13. *Let $I_0, I_1, \dots$ be a (finite or infinite) non-aborting chase sequence. Then for every L_0-model $\mathcal{I}$ of C_0 and $\mathcal{O}$ and every $i \geq 0$, there is a homomorphism from I_i to $\mathcal{I}$.*

Proof. Let $I_0, I_1, \dots$ and $\mathcal{I}$ be as in the lemma. Further, let $I_i = (F^i,\rho^i,\sim^i)$ for all $i \geq 0$ and $\mathcal{I} = (\mathbf{A}_\mathcal{I},\prec,(\mathcal{I}_L)_{L\in\mathbf{A}_\mathcal{I}},\rho)$. We prove by induction on i that for every $i \geq 0$ there is a homomorphism h_i from I_i to $\mathcal{I}$.

For the induction start, we can define a homomorphism h_0 from I_0 to $\mathcal{I}$ by simply mapping the only constant a_0 from $\mathsf{dom}(F^0_{L_0})$ to an element $d \in C^{\mathcal{I}_{L_0}}$ and for all other $L \in \mathbf{A}_\mathcal{O} \setminus \{L_0\}$ mapping the only constant a_L from $\mathsf{dom}(F^i_L)$ to any $d \in \Delta^{\mathcal{I}_L}$.

For $i > 0$ assume that h_i is a homomorphism from I_i to $\mathcal{I}$. We have to show that there is a homomorphism h_{i+1} from I_{i+1} to $\mathcal{I}$. We do a case analysis according to the rule that was applied to obtain I_{i+1} from I_i:

- for R1, R2, R4 to R6, R8, R10, R12, R14, and R16 to R18 set $h_{i+1} = h_i$. Using the definition of homomorphisms it is straightforward to see that if any choice of elements satisfies the preconditions of any of these rules in I_i, then the h_i-image of these elements satisfies the same conditions in $\mathcal{I}$. Using the fact that $\mathcal{I}$ is a model of $\mathcal{O}$, we then get that h_{i+1} is a homomorphism from I_{i+1} to $\mathcal{I}$.
- R3. Assume that the rule is applied to fact $A(a) \in F^i_L$ and CI $A \sqsubseteq_L \exists r.B \in \mathcal{O}$. Let b be the constant introduced by the rule application. We have $h_i(a) \in A^\mathcal{I}$. Since $\mathcal{I}$ is a model of $\mathcal{O}$, there is thus a $(h_i(a),d) \in r^{\mathcal{I}_L}$ with $d \in B^{\mathcal{I}_L}$. Set $h_{i+1} = h_i \uplus \{(b,d)\}$.
- R7. Assume that the rule is applied to fact $A(a) \in F^i_L$ and concept refinement $L' {:}\, q(\bar{x})$ refines $L {:}\, A \in \mathcal{O}$. Furthermore, let $a_1,\dots,a_n$ be the fresh constants generated by the rule application and let $h_i(a) = d$. We have $d \in A^{\mathcal{I}_L}$. Since $\mathcal{I}$ is a model of $\mathcal{O}$, $\rho_{L'}(d)$ must be defined and an

answer to q on $\mathcal{I}_{L'}$. We set $h_{i+1} = h_i \uplus \{(a_i, d_i) \mid 1 \leq i \leq n\}$ where $\rho_L(d) = (d_1, \ldots, d_n)$.

- R9. Assume that the rule is applied to the concept abstraction $L'\colon A$ $\underline{\mathsf{abstracts}}$ $L\colon q(\bar{x}) \in \mathcal{O}$ and homomorphism h' from q to F_L^i. Let $a \in \mathsf{dom}(F_{L'}^i)$ be the fresh constant introduced by this application, implying $\rho_L^{i+1}(a) = h'(\bar{x})$. By the definition of homomorphisms, $h_i(h'(\bar{x}))$ is a homomorphism from q to $\mathcal{I}_L$. Thus $\bar{e} = h_i(h'(\bar{x}))$ is an answer to q on $\mathcal{I}_L$. Since $\mathcal{I}$ is a model of $\mathcal{O}$, there is thus an element $d \in \Delta^{\mathcal{I}_{L'}}$ with $\rho_L(d) = \bar{e}$. We set $h_{i+1} = h_i \uplus \{(a, d)\}$.
- R11. Assume that the rule is applied to a tuple $(a, b) \in q_r(F_L^i)$ that is an answer to q_r on F_L^i for a role refinement $L'{:}q(\bar{x}, \bar{y})$ $\underline{\mathsf{refines}}$ $L{:}q_r(x, y)$ in $\mathcal{O}$. Let $a_1, \ldots, a_n, b_1, \ldots, b_m$ be the fresh constants generated by the rule application, implying $\rho_{L'}^{i+1}(a) = (a_1, \ldots, a_n)$ and $\rho_{L'}^{i+1}(b) = (b_1, \ldots, b_n)$. Furthermore let $h_i(a) = d$ and $h_i(b) = e$. By definition of homomorphisms, we have $(d, e) \in q_r(\mathcal{I}_L)$. Since $\mathcal{I}$ is a model of $\mathcal{O}$, $\rho_{L'}(d)$ and $\rho_{L'}(e)$ have to be defined. We set $h_{i+1} = h_i \uplus \{(a_i, d_i) \mid 1 \leq i \leq n\} \uplus \{(b_i, e_i) \mid 1 \leq i \leq m\}$ where $\rho_L(d) = (d_1, \ldots, d_n)$ and $\rho_L(e) = (e_1, \ldots, e_m)$.
- R13. Assume that the rule is applied to the role abstraction $L'{:}r$ $\underline{\mathsf{abstracts}}$ $L{:}q(\bar{x}, \bar{y})$ in $\mathcal{O}$ and homomorphism h' from q to F_L^i. Let $a, b \in \mathsf{dom}(F_{L'}^i)$ be the fresh constants introduced by this application, implying $\rho_L^{i+1}(a) = h'(\bar{x})$ and $\rho_L^{i+1}(b) = h'(\bar{y})$. Furthermore let $\bar{a} = h'(\bar{x}, \bar{y})$, $|\bar{x}| = n$ and $|\bar{y}| = m$. By the definition of homomorphisms, $h_i(\bar{a}) \in q(\mathcal{I}_L)$. Since $\mathcal{I}$ is a model of $\mathcal{O}$, there are thus elements $d, e \in \Delta^{\mathcal{I}_{L'}}$ with $(d, e) \in r^{\mathcal{I}_{L'}}$, $\rho_L(d) = (d_1, \ldots, d_n)$, and $\rho_L(e) = (e_1, \ldots, e_m)$. We set $h_{i+1} = h_i \uplus \{(a, d)\} \uplus \{(b, e)\}$.
- R15. Cannot be used to produce I_{i+1} as it only aborts.
- R19. Assume that the rule is applied to constants a_1 and a_2. Then $\rho_L(a_1)$ is defined, say $\rho_L(a_1) = (c_1, \ldots, c_n)$. Let $b_1, \ldots, b_n$ be the fresh constants introduced by the rule application. We set $h_{i+1}(b_i) = h_i(c_i)$ for $1 \leq i \leq n$.

It is straightforward to verify that this definition of h_{i+1} satisfies all six conditions of homomorphisms. ❑

Now we return to proving Lemma 12 by proving its contrapositive which is captured by the following lemma.

Lemma 14. *If the chase aborts and returns 'unsatisfiable', then there is no L_0-model $\mathcal{I}$ of C_0 and $\mathcal{O}$.*

Proof. Let $I_i = (F^i, \rho^i, \sim^i)$ be the interpretation candidate such that in the construction of I_{i+1} the chase aborts and returns 'unsatisfiable'. Assume to the contrary of what we have to show that there exists an L_0-model $\mathcal{I} = (\mathbf{A}_\mathcal{I}, \prec, (\mathcal{I}_L)_{L \in \mathbf{A}_\mathcal{I}}, \rho)$ of C_0 and $\mathcal{O}$. Then by Lemma 13 there is a homomorphism h from I_i to $\mathcal{I}$. We make a case distinction according to the rule that makes the chase abort:

- R8. Then there are $A(a) \in F_L^i$ and concept refinement $L'\colon q(\bar{x})$ $\underline{\mathsf{refines}}$ $L\colon A \in \mathcal{O}$ such that $\rho_{L'}^i(a)$ is defined and $|\bar{x}| \neq |\rho_{L'}^i(a)|$. Definition of homomorphisms then implies that $h(a) \in A^{\mathcal{I}_L}$ and $|\rho_L(h(a))| = |\rho_{L'}^i(a)| \neq |x|$. This is a contradiction to $\mathcal{I}$ being a model of $\mathcal{O}$.
- R12. Then there are $(a, b) \in q_r(F_L^i)$ for a role refinement $L'{:}q(\bar{x}, \bar{y})$ $\underline{\mathsf{refines}}$ $L{:}q_r(x, y)$ in $\mathcal{O}$ such that $|\bar{x}| \neq |\rho_{L'}^i(a)|$ or $|\bar{y}| \neq |\rho_{L'}^i(b)|$. Let us assume that $|\bar{x}| \neq |\rho_{L'}^i(a)|$. By definition of homomorphisms, we have $(h(a), h(b)) \in q_r(\mathcal{I}_L)$ and $|\rho_{L'}(h(a))| = |\rho_{L'}^i(a)|$ and hence also $|\rho_L(h(a))| \neq |x|$. This is a contradiction to $\mathcal{I}$ being a model of $\mathcal{O}$. The case of $|\bar{y}| \neq |\rho_{L'}^i(b)|$ is analogous.
- R15. Then there are a_1, a_2 such that $\rho_L^i(a_1) = \bar{e}_1$ and $\rho_L^i(a_2) = \bar{e}_2$ with $|\bar{e}_1| \neq |\bar{e}_2|$ and there are $b_1 \in \bar{e}_1$, and $b_2 \in \bar{e}_2$ such that $b_1 \sim b_2$. By definition of homomorphisms, we then have $\rho_L(h(a_1)) = h(\bar{e}_1)$ and $\rho_L(h(a_1)) = h(\bar{e}_2)$, which in particular implies $|h(\bar{e}_1)| \neq |h(\bar{e}_2)|$ and thus $h(\bar{e}_1) \neq h(\bar{e}_2)$. The definition of homomorphisms also yields $h(b_1) = h(b_2)$. This contradicts the fact that $\mathcal{I}$ has no overlapping ensembles.

❑

A.2 Proof of Lemma 1

We prove the existence of universal models for $\mathcal{ELH}_r^{\mathsf{abs}}$ by using the chase defined in the previous section. More concretely we want to prove Lemma 1 which we repeat here for the reader's convenience.

Lemma 1. *Let C_0 be an $\mathcal{EL}$-concept, $\mathcal{O}$ an $\mathcal{ELH}_r^{\mathsf{abs}}$-ontology, and $L_0 \in \mathbf{A}$. If C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$, then there exists a universal L_0-model of C_0 and $\mathcal{O}$.*

Proof. Since C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$, by Lemma 14 the chase of C_0 w.r.t. L_0 and $\mathcal{O}$ does not abort. Consequently, the result $I^* = (F^*, \rho^*, \sim^*)$ of the chase exists. Let $\mathcal{I}'$ be an L_0-model of C_0 and $\mathcal{O}$. By Lemma 13, there are homomorphisms h_i from I_i to $\mathcal{I}'$, for all $i \geq 0$. Clearly, $\widehat{h} := \bigcup_{i \geq 0} h_i$ is a homomorphism from I^* to $\mathcal{I}'$.

As part of the soundness proof above, we had constructed from I^* an L_0-model $\mathcal{I} = (\mathbf{A}_\mathcal{I}, \prec, (\mathcal{I}_L)_{L \in \mathbf{A}_\mathcal{I}}, \rho)$ of C_0 and $\mathcal{O}$. We prove that $\mathcal{I}$ is in fact the desired universal model. For the reader's convenience, we recall that

$$
\begin{aligned}
\Delta^{\mathcal{I}_L} &= \{[a] \mid a \in \mathsf{dom}(F_L^*)\} \\
A^{\mathcal{I}_L} &= \{[a] \mid A(a') \in F_L^* \text{ and } a' \in [a]\} \\
r^{\mathcal{I}_L} &= \{([a], [b]) \mid r(a', b') \in F_L^* \text{ and } a' \in [a], b' \in [b]\} \\
\rho_L &= \{([a], [b_1] \cdots [b_n]) \mid (a', b'_1 \cdots b'_n) \in \rho_L^* \text{ and} \\
&\qquad a' \in [a], b'_i \in [b_i]\}
\end{aligned}
$$

Due to rule R18, the following properties are satisfied:

1. if $[a] \in A^{\mathcal{I}_L}$, then $A(a') \in F_L^*$ for all $a' \in [a]$;
2. if $([a], [b]) \in r^{\mathcal{I}_L}$, then $r(a', b') \in F_L^*$ for all $a' \in [a]$ and $b' \in [b]$.

Moreover, we observe the following.

Claim. If $\rho_L([a]) = ([b_1], \ldots, [b_k])$, then for all $a' \in [a]$ we have $\rho_L^*(a') = (b'_1, \ldots, b'_k)$ with $b'_i \in [b_i]$ for $1 \leq i \leq k$.

Proof of claim. Let $\rho_L([a]) = ([b_1], \ldots, [b_k])$ and $a' \in [a]$.

By construction of $\mathcal{I}$, there then exist $\widehat{a} \in [a]$ and $\widehat{b}_1 \in [b_1], \dots, \widehat{b}_k \in [b_k]$ such that $\rho^*_L(\widehat{a}) = (\widehat{b}_1, \dots, \widehat{b}_k)$. Due to rule R19, also $\rho^*_L(a')$ is defined. Due to R15 and since the chase is not aborting, the tuple $\rho^*_L(a')$ also has length k. Let $\rho^*_L(a') = (b'_1, \dots, b'_k)$. Rule R17 yields $b'_i \sim \widehat{b}_i$ for $1 \leq i \leq k$ and thus the claim is proved.

We define the desired homomorphism h from $\mathcal{I}$ to $\mathcal{I}'$ by traversing the tree $(\mathbf{A}_\mathcal{I}, \prec^{-1})$ in a top-down fashion. This boils down to choosing, for each $[a] \in \Delta^\mathcal{I}$, an $a' \in [a]$ and setting $h([a]) = \widehat{h}(a')$.

We start at the root L_0. For each $[a] \in \Delta^{\mathcal{I}_{L_0}}$, choose an arbitrary $a \in [a]$ and set $h([a]) = \widehat{h}(a)$.

Now assume that level L' of the tree was already treated and that $L \prec L'$. Consider any $[a] \in \Delta^{\mathcal{I}_{L'}}$ with $\rho_L([a])$ defined. Let $\rho_L([a]) = ([b_1], \dots, [b_k])$ and $h([a]) = \widehat{h}(a')$. By the claim, there are $b'_1 \in [b_1], \dots, b'_k \in [b_k]$ such that $\rho^*_L(a') = (b'_1, \dots, b'_k)$. Set $h([b_i]) = \widehat{h}(b'_i)$ for $1 \leq i \leq k$. Note that this is well-defined since every element of $\Delta^{\mathcal{I}_L}$ can occur in at most one ensemble. For every $[b] \in \Delta^{\mathcal{I}_L}$ that does not occur in any ensemble, choose an arbitrary $b \in [b]$ and set $h([b]) = \widehat{h}(b)$.

It is straightforward to see that h is a homomorphism from $\mathcal{I}$ to $\mathcal{I}'$. In fact, $\widehat{h}$ must satisfy the six conditions from the definition of homomorphisms from interpretation candidates to interpretations, and together with Point 1 and Point 2 from above, these conditions imply the five conditions that h must satisfy to be a homomorphism from $\mathcal{I}$ to $\mathcal{I}'$. ❑

B Proofs for Section 4.3

In this section we prove that the NP algorithm for checking unsatisfiability in $\mathcal{ELH}^{\mathsf{abs}}_r$[cr] as presented in Section 4.3 is correct.

We start with some observations about the chase. A *maximal connected component (MCC)* of a set of facts E is a maximal subset $E' \subseteq E$ such that the undirected graph

$$(\mathsf{dom}(E), \{\{d, e\} \mid r(d, e) \in E \text{ for some role name } r\})$$

is connected. An MCC of an interpretation candidate $I = (F, \rho, \sim)$ is an MCC of F. An *L-ensemble* in F is any tuple $\bar{e}$ such that for some $a \in \mathsf{dom}(F)$, we have $\rho_L(a) = \bar{e}$.

Let $I_0, I_1, \dots, I_k$ be a chase sequence w.r.t. an $\mathcal{ELH}^{\mathsf{abs}}_r$[cr]-ontology in normal form with $I^i = (F^i, \rho^i, \sim^i)$ for $1 \leq i \leq k$. An easy analysis of the chase rule reveals that there are three types of MCCs in F^k:

1. the MCC in $F^k_{L_0}$ that contains $C_0(a_0)$;
2. MCCs in F^k_L, with $L \in \mathbf{A}_\mathcal{O} \setminus \{L_0\}$, that contain $A_\top(a_L)$;
3. MCCs that are introduced by an R7 application.

Note in particular that since there are no abstractions and no role refinements in the ontology, Rules R9 to R19 of the chase are never applicable and thus the only rule that can introduce a new MCC is R7. MCCs of Type 1 and 2 do not contain any ensembles and MCCs of Type 3 contain a single ensemble, that is, there is a unique a such that $\rho^k(a) = (b_1, \dots, b_n)$ with all (equivalently: some) of $b_1, \dots, b_n$ in the MCC. We call a the *origin* of the MCC.

For a set of facts F and an $a \in \mathsf{dom}(F)$, we use $\mathsf{CN}_F(a)$ to denote the set $\mathsf{CN}_F(a) = \{A \mid A(a) \in F\}$. It is not difficult to show that Lemma 13 implies the following.

Lemma 15.

1. *If K is an MCC of Type 1 and $a \in \mathsf{dom}(K)$, then $\mathcal{O}_{L_0} \models C_0 \sqsubseteq_{L_0} \exists \mathsf{u}.(\bigsqcap \mathsf{CN}_K(a))$;*
2. *If K is an MCC of Type 2 on level L and $a \in \mathsf{dom}(K)$, then $\mathcal{O}_L \models \top \sqsubseteq_L \exists \mathsf{u}.(\bigsqcap \mathsf{CN}_K(a))$;*
3. *If K is an MCC of Type 3 on level L and with origin $a \in \mathsf{dom}(F^k_{L'})$ and $b \in \mathsf{dom}(K)$, then*

$$\mathcal{A}, \mathcal{O}_L \models \exists \mathsf{u}.(\bigsqcap \mathsf{CN}_K(b))$$

where $\mathcal{A}$ is the union of all queries in $Q^{\mathsf{ref}}_{L',L}(\mathsf{CN}_{F^k_{L'}}(a))$, viewed as ABoxes.

Now we are ready to prove Lemma 3 which we repeat here for the reader's convenience.

Lemma 3. *The algorithm accepts iff C_0 is L_0-unsatisfiable w.r.t. $\mathcal{O}$.*

Proof. "⇒". Assume that the algorithm accepts and that, to the contrary of what we need to show, there is an L_0-model $\mathcal{I}$ of C_0 and $\mathcal{O}$.

Condition 1 of the NP algorithm and $\mathcal{I}$ being a model of C_0 and $\mathcal{O}$ imply that $\mathcal{I}$ contains an element d_1 with $d_1 \in S_1^{\mathcal{I}_{L_1}}$. Condition 2 implies that for each $i \in \{1, \dots, n-1\}$, there is an element $d_i \in \Delta^{\mathcal{I}_{L_i}}$ with $d_i \in S_i^{\mathcal{I}_{L_i}}$. By Condition 3 and $\mathcal{I}$ being a model of $\mathcal{O}$, d_n would have to refine to two tuples of distinct arities which is impossible in A-interpretations. We have thus obtained a contradiction.

"⇐". Assume that C_0 is L_0-unsatisfiable w.r.t. $\mathcal{O}$. We can assume that the abstraction graph of $\mathcal{O}$ is a tree since otherwise our NP algorithm accepts right away. Since C_0 is unsatisfiable, Lemma 9 implies that the L_0-chase on C_0 and $\mathcal{O}$ aborts and returns 'unsatisfiable'. Assume that the chase sequence constructed until abortion is $I_0, I_1, \dots, I_k$, and that $I^i = (F^i, \rho^i, \sim^i)$ for $1 \leq i \leq k$.

We guide our algorithm into accepting by identifying a suitable sequence $S_1, L_1, \dots, S_n, L_n$ to be guessed. Since $\mathcal{O}$ does not contain any abstractions or role refinements, Rules R9 to R19 of the chase will never be applied and thus the only rule application that can make the chase abort is R8. Let a_n be the element for which R8 aborted (called 'a' in the formulation of R8). We set

$$S_n = \{A \mid A(a_n) \in F^k_{L(a_n)}\}$$

and $L_n = L(a_n)$. To define the remaining sequence, we work our way backwards through the chase.

If a_n is in an MCC of Type 1 or 2, then by Lemma 15, Condition 1 of our NP algorithm is satisfied and we are done. If a_n is in an MCC of Type 3, then let a_{n-1} be the origin of that MCC. We set $S_{n-1} = \{A \mid A(a_{n-1}) \in F^k_{L(a_{n-1})}\}$ and $L_{n-1} = L(a_{n-1})$. By Lemma 15, Condition 2 of our algorithm is satisfied. Now we repeat this procedure of determining S_i and L_i which results in a sequence of at most linearly many S_i and L_i, at most one per abstraction level in $\mathcal{O}$. This sequence will always reach an MCC of Type 1 or 2 and thus satisfying Condition 1, since on the root level of $\mathbf{A}_\mathcal{O}$ all MCCs are of Type 1 or 2.

It is easy to verify that all three conditions of our NP algorithm are satisfied by this sequence of S_i and L_I and thus it accepts, as required. ❑

C Proofs for Section 4.4

The main purpose of this section is to prove Lemma 5, that is, the correctness of our algorithm. We repeat the lemma here for the reader's convenience.

Lemma 5. *The algorithm accepts iff C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$.*

To prove the completeness ("if") part of the lemma, we first show that if C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$, then there is a pseudo-model of C_0 and $\mathcal{O}$ in which the maximally connected components are of size polynomial in $||\mathcal{O}||$.

C.1 Uniform Universal Models

It turns out that the universal models used in the proof of Lemma 1, being obtained from a filtration of the chase, are very hard to analyze. To avoid such an analysis, we resort to a different, 'semantic' definition of universal models. These are constructed by starting from the universal models whose existence is guaranteed by Lemma 1 and then 'piecing them together' in a very uniform, tree-like way.

Let C_0 be an $\mathcal{EL}$-concept, $\mathcal{O}$ an $\mathcal{ELH}^{\mathsf{abs}}_r$[cr, ca]-ontology, and $L_0 \in \mathbf{A}_\mathcal{O}$ such that C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$. We aim to construct a certain, highly uniformized universal L_0-model $\mathcal{I}$ of C_0 and $\mathcal{O}$.

Let $\mathcal{I}$ be an A-interpretation. With an *L-ensemble* in $\mathcal{I}$, we mean any tuple $\bar{e}$ over $\Delta^\mathcal{I}$ such that $\rho^\mathcal{I}_L(d) = \bar{e}$ for some $d \in \Delta^\mathcal{I}$. For a tuple $\bar{e}$ of elements over $\Delta^{\mathcal{I}_L}$, for some L, we use $\mathcal{I}|_{\bar{e}}$ to denote the $\mathcal{EL}$-interpretation that is obtained from $\mathcal{I}_L$ by restricting the domain to the elements of $\bar{e}$. In what follows, $\bar{e}$ shall either be an ensemble or a single element that is not part of any ensemble.

For every $\mathcal{EL}$-concept C and $L \in \mathbf{A}_\mathcal{O}$ such that C is L-satisfiable w.r.t. $\mathcal{O}$, Lemma 1 allows us to fix a universal L-model $\mathcal{U}_{C,L}$ of C and $\mathcal{O}$ with distinguished element $d_{C,L}$. With $\bar{e}_{C,L}$, we denote the unique ensemble in $\mathcal{U}_{C,L}$ that contains $d_{C,L}$, if existent, and the trivial tuple $(d_{C,L})$ otherwise. As a shortcut, we use $\mathcal{I}_{C,L}$ to denote the $\mathcal{EL}$-interpretation $\mathcal{U}_{C,L}|_{\bar{e}_{C,L}}$.

We construct a sequence of A-interpretations along with a list O_A of abstraction obligations. Start with the following A-interpretation $\mathcal{I}$:

- for all $L \in \mathbf{A}_\mathcal{O}$ and for $C = C_0$ if $L = L_0$ and $C = \top$ otherwise, $\mathcal{I}_L$ is $\mathcal{I}_{C,L}$
- $\rho^\mathcal{I}$ is empty;
- $\mathbf{A}^\mathcal{I} = \mathbf{A}_\mathcal{O}$ and $\prec = \prec_\mathcal{O}$.

Regarding O_A, we start with an empty list and then do the following, for all $L \in \mathbf{A}_\mathcal{O}$ and for $C = C_0$ if $L = L_0$ and $C = \top$ otherwise:

- if $d_{C,L}$ is part of an ensemble $\bar{e}$ in $\mathcal{U}_{C,L}$, append $(C, L, \bar{e}, \bar{e})$ to O_A.

We may use d_{C_0,L_0}, the distinguished element of $\mathcal{U}_{C_0,L_0}$, also as the distinguished element of $\mathcal{I}$. For easier reference, we denote it with $d_\mathcal{I}$.

For a role name $r \in \mathbf{R}$ and abstraction level $L \in \mathbf{A}_\mathcal{O}$, we use $R_{r,\mathcal{O}_L}$ to denote the set of role names $R_{r,\mathcal{O}_L} = \{s \mid \mathcal{O} \models r \sqsubseteq_L s\}$. Similarly, we use $C_{r,\mathcal{O}_L}$ to denote the concept $C_{r,\mathcal{O}_L} = \bigsqcap\{A \mid \top \sqsubseteq_L \forall s.A \in \mathcal{O} \text{ and } s \in R_{r,\mathcal{O}_L}\}$.

Now apply the following rules in a fair way:

1. if $d \in A^{\mathcal{I}_L}$, $A \sqsubseteq_L \exists r.B \in \mathcal{O}$, and there is no $e \in (B \sqcap C_{r,\mathcal{O}_L})^{\mathcal{I}_L}$ with $(d, e) \in s^{\mathcal{I}_L}$ for all $s \in R_{r,\mathcal{O}_L}$, then do the following:
   (a) let $D = B \sqcap C_{r,\mathcal{O}_L}$. Use an isomorphism ι to rename the elements of $\mathcal{I}_{D,L}$ to fresh elements, add the resulting interpretation to $\mathcal{I}_L$ and extend $s^{\mathcal{I}_L}$ with $(d, \iota(d_{D,L}))$ for all $s \in R_{r,\mathcal{O}_L}$;
   (b) if $d_{D,L}$ is part of an ensemble $\bar{e}$ in $\mathcal{U}_{D,L}$, append $(D, L, \bar{e}, \iota(\bar{e}))$ to O_A;
2. if $d \in \Delta^{\mathcal{I}_L}$, $C = \bigsqcap \mathsf{CN}_{\mathcal{I}_L}(d)$, $\rho^{\mathcal{U}_{C,L}}_{L'}(d_{C,L}) = \bar{e}$ is defined and $\rho^\mathcal{I}_{L'}(d)$ is undefined, then
   (a) use an isomorphism ι to rename the elements of $(\mathcal{U}_{C,L})|_{\bar{e}}$ to fresh elements and disjointly add the resulting interpretation to $\mathcal{I}_{L'}$;
   (b) set $\rho^\mathcal{I}_{L'}(d) = \iota(\bar{e})$;
3. if $(C, L, \bar{e}, \bar{e}')$ is the first tuple on the list O_A, then remove it; if $\rho^{\mathcal{U}_{C,L}}_L(f) = \bar{e}$, $L^{\mathcal{U}_{C,L}}(f) = L'$, and $D = \bigsqcap \mathsf{CN}_{\mathcal{U}_{C,L}}(f)$, then do the following:
   (a) use an isomorphism ι to rename the elements of $\mathcal{I}_{D,L'}$ to fresh elements and disjointly add the resulting interpretation to $\mathcal{I}_{L'}$;
   (b) set $\rho^\mathcal{I}_L(\iota(d_{D,L'})) = \bar{e}'$;
   (c) if $d_{D,L'}$ is part of an ensemble $\bar{f}$ in $\mathcal{U}_{D,L'}$, append $(D, L', \bar{f}, \iota(\bar{f}))$ to O_A.

Intuitively, Rule 1 deals with existential restrictions, Rule 2 with refinements, and Rule 3 with abstractions.

First, we prove that $\mathcal{I}$ is universal, and then that it is indeed an L_0-model of C_0 and $\mathcal{O}$.

Lemma 16. *Let $\mathcal{J}$ be a model of $\mathcal{O}$ with $d_0 \in C_0^{\mathcal{I}_{L_0}}$. Then there is a homomorphism h from $\mathcal{I}$ to $\mathcal{J}$ such that $h(d_\mathcal{I}) = d_0$.*

Proof. Let $\mathcal{I}^0, \mathcal{I}^1, \ldots$ be the sequence of A-interpretations encountered during the construction of $\mathcal{I}$. We show that for each $i \geq 0$, there is a homomorphism h_i from $\mathcal{I}^i$ to $\mathcal{J}$ with $h_i(d_{\mathcal{I}^i}) = d_0$. The desired homomorphism h is then $\bigcup_{i \geq 0} h_i$.

For $\mathcal{I}^0$, we start with an empty homomorphism h_0 and extend it as follows:

- since $\mathcal{I}^0_{L_0} = \mathcal{I}_{C_0,L_0}$, there must be a homomorphism h from $\mathcal{I}^0_{L_0}$ to $\mathcal{J}_{L_0}$ with $h(d_\mathcal{I}) = d_0$ by universality of $\mathcal{U}_{C_0,L_0}$. We add h to h_0;

- for the other levels $L \in \mathbf{A}_\mathcal{O} \setminus \{L_0\}$, we have $\mathcal{I}^0_L = \mathcal{I}_{\top,L}$ which again implies that there must be a homomorphism h from $\mathcal{I}^0_L$ to $\mathcal{J}_L$ by universality of $\mathcal{U}_{\top,L}$. We add h to h_0.

For $i > 0$ assume that h_i is a homomorphism from $\mathcal{I}^i$ to $\mathcal{J}$. We have to show that there is a homomorphism h_{i+1} from $\mathcal{I}^{i+1}$ to $\mathcal{J}$. We start with $h_{i+1} = h_i$ and do a case analysis on the rule that was applied to obtain $\mathcal{I}^{i+1}$ from $\mathcal{I}^i$:

- in case of Rule 1, let $d \in A^{\mathcal{I}_L}$ and $A \sqsubseteq_L \exists r.B \in \mathcal{O}$ be the element and CI it was applied on. By the IH, $h_i(d) \in A^{\mathcal{J}_L}$ and since $\mathcal{J}$ is a model of $\mathcal{O}$, there is an $e' \in D^{\mathcal{J}_L}$ with $D = (B \sqcap C_{r,\mathcal{O}_L})$ and $(h_i(d), e') \in s^{\mathcal{J}_L}$ for all $s \in R_{r,\mathcal{O}_L}$.

  Thus for the isomorphic (with isomorphism ι) copy of $\mathcal{I}_{D,L}$ added to $\mathcal{I}^{i+1}_L$ in Rule 1a, we set $h_{i+1}(\iota(d_{D,L})) = e'$ and are clearly able to extend h_{i+1} to the (possibly) remaining elements in the copy of $\mathcal{I}_{D,L}$ by the universality of $\mathcal{U}_{D,L}$;

- in case of Rule 2, let $d \in \Delta^{\mathcal{I}_L}$ and $C = \bigsqcap \mathsf{CN}_{\mathcal{I}_L}(d)$ be the element and concept it was applied on. By the IH, $h_i(d) \in C^{\mathcal{J}_L}$ and universality of $\mathcal{U}_{C,L}$ implies that there is a homomorphism h from $\mathcal{U}_{C,L}$ to $\mathcal{J}$ with $h(d_{C,L}) = h_i(d)$. Since Rule 2 was applied, $\rho^{\mathcal{U}_{C,L}}_{L'}(d_{C,L}) = \bar{e}$ must be defined which implies that $\rho^{\mathcal{J}}_{L'}(h_i(d)) = h(\bar{e})$ is defined as well, by definition of homomorphisms.

  Let ι be the isomorphism used in Part a of Rule 2. We can thus extend h_{i+1} by setting $h_{i+1}(\iota(\bar{e})) = h(\bar{e})$.

- in case of Rule 3, let $(C, L, \bar{e}, \bar{e}') \in O_A$ be the tuple that it was applied on. Further, let $\rho^{\mathcal{U}_{C,L}}_L(f) = \bar{e}$, $L^{\mathcal{U}_{C,L}}(f) = L'$, and $D = \bigsqcap \mathsf{CN}_{\mathcal{U}_{C,L}}(f)$.

  By the construction of O_A, we get that $d_{C,L}$ is part of $\bar{e}$ in $\mathcal{U}_{C,L}$. Remember that $\bar{e}'$ is an isomorphic copy of $\bar{e}$ by some isomorphism ι. Let $d = h_i(\iota(d_{C,L}))$. By the IH, we then have $d \in C^{\mathcal{J}_L}$ and by universality of $\mathcal{U}_{C,L}$, there is a homomorphism h from $\mathcal{U}_{C,L}$ to $\mathcal{J}$ with $h(d_{C,L}) = d$. The definition of homomorphisms then implies that d is part of the ensemble $h(\bar{e})$. We must also have $\rho^{\mathcal{J}}_L(h(f)) = h(\bar{e})$ and $h(f) \in D^{\mathcal{J}_{L'}}$.

  Let ι' be the isomorphism used in Rule 3 to create a copy of $\mathcal{I}_{D,L'}$. We extend h_{i+1} by setting $h_{i+1}(\iota'(d_{D,L'})) = h(f)$. It is straightforward to see that we can extend h_{i+1} to (possibly) remaining elements in $\mathcal{I}_{D,L'}$ by universality of $\mathcal{U}_{D,L'}$.

❏

In the following we often talk about isomorphic copies of elements or ensembles so let us define what isomorphism means in this context. Let $\mathcal{I}$ and $\mathcal{I}'$ be two A-interpretations, and $\bar{e}$ and $\bar{e}'$ tuples of n elements over $\Delta^{\mathcal{I}_L}$ and $\Delta^{\mathcal{I}'_L}$ respectively. We call $\bar{e}$ and $\bar{e}'$ *isomorphic* w.r.t. $\mathcal{I}$ and $\mathcal{I}'$, if $\bar{e}[i] \mapsto \bar{e}'[i]$ for $1 \leq i \leq n$ is an isomorphism from $\mathcal{I}_L|_{\bar{e}}$ to $\mathcal{I}'_L|_{\bar{e}'}$. We call two elements $d \in \Delta^{\mathcal{I}}$ and $e \in \Delta^{\mathcal{I}'}$ isomorphic w.r.t. to their interpretations, if (d) in $\mathcal{I}$ and (e) in $\mathcal{I}'$ are isomorphic. We use $\bar{e} \simeq_\mathcal{I} \bar{e}'$ to denote that $\bar{e}$ and $\bar{e}'$ are isomorphic in $\mathcal{I}$ and similar for $d \simeq_\mathcal{I} d'$.

Note that this definition of isomorphisms implies that if a tuple $\bar{e}$ in some interpretation $\mathcal{I}$ is an answer to some CQ q, then any tuple $\bar{e}'$ isomorphic to $\bar{e}$ in some interpretation $\mathcal{I}'$ is also an answer to q.

Lemma 17. *$\mathcal{I}$ is an L_0-model of C_0 and $\mathcal{O}$.*

Proof. For readability we will add o to the lower index of elements or tuples from universal models, intuitively denoting them as the 'original' elements, while $\mathcal{I}$ consists of copies of these elements. It is straightforward to see that $C_0^{\mathcal{I}_{L_0}} \neq \emptyset$ by the construction of $\mathcal{I}$.

To prove that $\mathcal{I}$ is a model of $\mathcal{O}$, let us first consider the concept inclusions in $\mathcal{O}$. Note that in the construction of $\mathcal{I}$, every element $d \in \Delta^{\mathcal{I}}$ is introduced as an isomorphic copy of some element $f \in \mathcal{U}_{C,L}$ where $\mathcal{U}_{C,L}$ is a universal L-model of C and $\mathcal{O}$. This clearly implies that all CIs of the form $\top \sqsubseteq_L A \in \mathcal{O}$ and $A_1 \sqcap \cdots \sqcap A_n \sqsubseteq_L B \in \mathcal{O}$ are satisfied for d. Similarly, CIs of the form $A \sqsubseteq_L \exists r.B \in \mathcal{O}$ are satisfied by the fair application of Rule 1.

For CIs of the form $\exists r.A \sqsubseteq_L B \in \mathcal{O}$, consider an element $d \in (\exists r.A)^{\mathcal{I}_L}$. By the semantics there is an $e \in A^{\mathcal{I}_L}$ with $(d, e) \in r^{\mathcal{I}_L}$. First, let us consider the case that d and e are part of the same ensemble in $\mathcal{I}$. The construction of $\mathcal{I}$ implies that every L-ensemble in $\mathcal{I}$ is an isomorphic copy of an L-ensemble in a model $\mathcal{I}'$ of $\mathcal{O}$ which immediately implies $d \in B^{\mathcal{I}_L}$, as required.

Now assume that d and e are not part of the same ensemble. Hence $(d, e) \in r^{\mathcal{I}_L}$ must have been introduced by an application of Rule 1. This implies $d \in A'$ for some $A' \sqsubseteq_L \exists r.B' \in \mathcal{O}$ such that e is an isomorphic copy of $d_{B',L}$ by some isomorphism ι. By construction of $\mathcal{I}$, this $d_{B',L}$ must then also satisfy $d_{B',L} \in A^{\mathcal{U}_{B',L}}$, since it is isomorphic to e.

Let $\mathcal{J} = \mathcal{U}_{C,L}$ and $d_o \in \Delta^{\mathcal{J}}$ be the universal model and element such that d was introduced as an isomorphic copy of d_o. Since d_o is part of a model of $\mathcal{O}$, there is an e_o with $(d_o, e_o) \in r^{\mathcal{J}_L}$ and $e_o \in B'^{\mathcal{J}_L}$. Universality of $d_{B',L}$ implies that $e_o \in A^{\mathcal{J}_L}$. Finally, $\mathcal{J}$ being a model of $\mathcal{O}$ implies $d_o \in B^{\mathcal{J}_L}$ and since d is an isomorphic copy of d_o, we get $d \in B^{\mathcal{I}_L}$, as required.

It is straightforward to see that all role inclusions and range restrictions of $\mathcal{O}$ are satisfied in any $(\mathcal{U}_{C,L})|_{\bar{e}_{C,L}}$, since it is part of a model of $\mathcal{O}$. We always add isomorphic copies of such interpretations in the rules, which implies that any roles introduced by such a copy still satisfy all role inclusions and range restrictions.

That roles introduced by a Rule 1 application also satisfy role inclusions and range restrictions immediately follows from the definition fo $C_{r,\mathcal{O}_L}$ and $R_{r,\mathcal{O}_L}$.

Now we are ready to prove that refinement and abstraction statements are satisfied.

- Assume that $d \in A^{\mathcal{I}_L}$ and there is a concept refinement $L'\colon q(\bar{x})$ <u>refines</u> $L\colon A \in \mathcal{O}$.

  If $\rho^{\mathcal{I}}_{L'}(d) = \bar{e}$ was set by a Rule 3 application, then there is a universal model $\mathcal{U}_{C,L}$ with $\rho^{\mathcal{U}_L}_{L'}(d_o) = \bar{e}_o$ and with d an isomorphic copy of d_o and $\bar{e}$ an isomorphic copy of $\bar{e}_o$. Since $\mathcal{U}_{C,L}$ is a model of $\mathcal{O}$, this implies that $\bar{e}_o$ and thus also $\bar{e}$ are an answer to q on their respective interpretation.

  Otherwise, let $C = \bigsqcap \mathsf{CN}_{\mathcal{I}_L}(d)$. Since $\mathcal{U}_{C,L}$ is a model of $\mathcal{O}$, it is straightforward to see that $\rho^{\mathcal{U}_{C,L}}_{L'}(d_{C,L}) = \bar{e}_o$

is defined and $\bar{e}_o$ is an answer to q on $\mathcal{U}_{C,L}$. Now the fair application of Rule 2 implies that $\rho^{\mathcal{I}}_{L'}(d) = \iota(\bar{e}_o)$ with ι some isomorphism and thus $\iota(\bar{e}_o)$ is an answer to q on $\mathcal{I}$, as required.

- Assume that $\bar{e}$ is an answer to a CQ q on $\mathcal{I}_L$ for some concept abstraction $L'\colon A$ abstracts $L\colon q(\bar{x}) \in \mathcal{O}$. We differentiate between three cases. The ensemble $\bar{e}$ might coincide with an ensemble in $\mathcal{I}$, it might partially overlap with an ensemble in $\mathcal{I}$, or it might not share any constants at all with an ensemble in $\mathcal{I}$. We will prove that only the first case can occur and the other two lead to a contradiction.

  Case 1: $\bar{e}$ (properly!) overlaps with an ensemble in $\mathcal{I}$.

  Then Condition 5 of homomorphisms and Lemma 16 would imply overlapping ensembles in every model of $\mathcal{O}$, contradicting the L_0-satisfiability of C_0 w.r.t. $\mathcal{O}$.

  Case 2: $\bar{e}$ coincides with an L-ensemble in $\mathcal{I}$.

  Then by the construction of $\mathcal{I}$, there are two cases. Either $\bar{e}$ was introduced by Rule 2 or $\bar{e}$ was part of a tuple in O_A in the construction of $\mathcal{I}$.

  If $\bar{e}$ was introduced by an application of Rule 2 for an element $d \in \Delta^{\mathcal{I}_{L'}}$ and $C = \bigsqcap \mathsf{CN}_{\mathcal{I}_{L'}}(d)$, then let $\mathcal{J} = \mathcal{U}_{D,L}$ be the universal model such that d was introduced as an isomorphic copy of some $d_o \in C^{\mathcal{J}_{L'}}$. By the rule application $\rho^{\mathcal{U}_{C,L}}_{L'}(d_{C,L}) = \bar{e}_o$ is defined. Universality of $\mathcal{U}_{C,L}$ then implies that $\rho^{\mathcal{J}}_{L}(d_o) = \bar{f}$ is defined and that $\bar{f}$ is an answer to q on $\mathcal{J}$.

  Now $\mathcal{J}$ being a model of $\mathcal{O}$ implies $d_o \in A^{\mathcal{J}_{L'}}$ and finally d being an isomorphic copy of d_o implies $d \in A^{\mathcal{I}_{L'}}$, as required.

  Otherwise, there must have been a Rule 3 application on a tuple $(C, L, \bar{e}_o, \bar{e}) \in O_A$. Let $\mathcal{J} = \mathcal{U}_{C,L}$ be the universal L-model of C and $\mathcal{O}$. The Rule 3 application on this tuple sets $\rho^{\mathcal{I}}_{L}(d) = \bar{e}$ with d an isomorphic copy of an $d_o \in \Delta^{\mathcal{J}_{L'}}$ such that $\rho^{\mathcal{J}}_{L}(d_o) = \bar{e}_o$. Recall that by the construction of $\mathcal{I}$, $\bar{e}$ is an isomorphic copy of $\bar{e}_o$, which implies that $\bar{e}_o$ is an answer to q on $\mathcal{J}$. Since $\mathcal{J}$ is a model of $\mathcal{O}$, we then obtain $d_o \in A^{\mathcal{J}_{L'}}$ and due to d being an isomorphic copy of d_o, it follows that $d \in A^{\mathcal{I}_{L'}}$, as required.

  Case 3: $\bar{e}$ does not share a constant with any L-ensemble in $\mathcal{I}$.

  We call an element of an A-interpretation *free* if it is not part of any ensemble. Note that Rules 1 and 3 may introduce free elements, and also the initial step may. An easy analysis of the construction of $\mathcal{I}$ shows that

  $(*)$ if some element $d \in \Delta^{\mathcal{I}}$ was introduced as an isomorphic copy of an element d_o from an interpretation $\mathcal{U}_{C,L}$, then d being free implies that d_o is free.

  We next observe that any tuple $\bar{f}$ of free and connected elements in $\Delta^{\mathcal{I}}$ forms a tree $T = (V, E)$ with

  $$V = \{d \mid d \in \Delta^{\mathcal{I}} \text{ is part of } \bar{f}\}$$
  $$E = \{(d, e) \in V \times V \mid (d, e) \in r^{\mathcal{I}_L} \text{ for some } r \in \mathbf{R}\}.$$

  This is the case since roles outside of ensembles are only introduced by Rule 1 which always introduces a fresh element as an endpoint.

  Since CQs in abstraction statements are connected and the elements in $\bar{e}$ are all free, these elements must form a tree $T = (V, E)$.

  Claim. Let $d_R \in V$ be the root of T and $\mathcal{J} = \mathcal{U}_{C,L}$ the universal model such that d_R was introduced as an isomorphic copy of $d_{C,L}$. Then there is a homomorphism h from $\mathcal{I}_L|_V$ to $\mathcal{J}_L$ with $h(d_R) = d_{C,L}$.

  Proof of the claim. We define h step by step, traversing T from the root to the leaves. We start with setting $h(d_R) = d_{C,L}$ which is a (partial) homomorphism since d_R is an isomorphic copy of $d_{C,L}$.

  Now assume that $h(d)$ is already defined and $(d, e) \in E$. Since e was introduced by a Rule 1 application, there is an $A \sqsubseteq_L \exists r.B \in \mathcal{O}$ with $d \in A^{\mathcal{I}_L}$ and $(d, e) \in s^{\mathcal{I}_L}$ for all $s \in R_{r,\mathcal{O}_L}$ such that e is an isomorphic copy of $d_{C,L}$ with $C = (B \sqcap C_{r,\mathcal{O}_L})$. We have $h(d) \in A^{\mathcal{J}_L}$ and $\mathcal{J}$ being a model of $\mathcal{O}$ implies that there is an $f \in C^{\mathcal{J}_L}$ with $(h_i(d), f) \in s^{\mathcal{J}_L}$ for all $s \in R_{r,\mathcal{O}_L}$. We set $h(e) = f$. This finishes the proof of the claim.

  By the claim, there is a homomorphism h from $\mathcal{I}_L|_V$ to $\mathcal{J}_L$ with $h(d_R) = d_{C,L}$. Hence $h(\bar{e})$ is an answer to q on $\mathcal{U}_{C,L}$ with $d_{C,L}$ part of that answer. By $(*)$, this contradicts our assumption that d_R is free.

❑

This concludes the proof that $\mathcal{I}$ is a uniformized universal model. What follows is some further analysis of the structure of $\mathcal{I}$ which allows us to confirm the concrete upper bound used in Line 3 of the algorithm (size of the X_L).

We call an ensemble $\bar{e}$ in $\mathcal{I}$ an *abstraction ensemble* if there was a Rule 3 application that set $\rho^{\mathcal{I}}(f, L) = \bar{d}$ for some f. Otherwise, we call $\bar{e}$ a *refinement ensemble*. Intuitively, we will show that there are only polynomially many abstraction ensembles that are pairwise non-isomorphic. Additionally, if two abstraction ensembles $\bar{e}$ and $\bar{f}$ are isomorphic in $\mathcal{I}$, then they abstract to two elements that satisfy the same concept names in $\mathcal{I}$.

A difficulty that arises when trying to prove the second point is the following. Let $\bar{e}$ be an ensemble in $\mathcal{I}$ that is a copy of $\bar{e}_{C,L}$ and $\bar{f}$ an ensemble in $\mathcal{I}$ that is a copy of $\bar{f}_{D,L}$ with $\bar{e} \simeq_{\mathcal{I}} \bar{f}$ and $C \neq D$. What we would need to show now is that in $\mathcal{U}_{C,L}$ and $\mathcal{U}_{D,L}$, the ensembles $\bar{e}_{C,L}$ and $\bar{f}_{D,L}$ abstract to elements that satisfy the same concept names. This implies analyzing how the chase constructs these models, including the filtration step. As we already pointed out in the main part of the paper, this is rather complex. Thus instead we will inspect the construction of $\mathcal{I}$, in particular the set of abstraction obligations.

Recall that $\mathcal{I}$ was constructed by the union of a sequence of interpretations $\mathcal{I}^0, \mathcal{I}^1, \ldots$ and with each of them a list of abstraction obligations $O^0_A, O^1_A, \ldots$.[4] Let $O^*_A = \bigcup_{i \geq 0} O^i_A$

[4]In the algorithm we use only O_A but we can easily define O^i_A to be the O_A used for I^i.

be the union of all abstraction obligations encountered during the construction of $\mathcal{I}$. Recall that a tuple $\bar{o} \in O_A$ with $\bar{o} = (C, L, \bar{e}, \bar{f})$ consists of a concept C, abstraction level L, tuple $\bar{e}$ that the distinguished element of $\mathcal{U}_{C,L}$ is part of and isomorphic copy $\bar{f}$ of that tuple in $\mathcal{I}$. We may use $C^{\bar{o}}$ to denote the *concept C of $\bar{o}$* and $L^{\bar{o}}$ to denote *the level L of $\bar{o}$*.

Note that a simple analysis of $\mathcal{I}$ implies that for every tuple $(C, L, \bar{e}, \bar{f}) \in O_A^*$, Rule 3 sets $\rho_L^{\mathcal{I}}(d) = \bar{f}$ for some $d \in \Delta^{\mathcal{I}}$. Hence the set of all abstraction ensembles in $\mathcal{I}$ is exactly the set of all ensembles $\bar{f}$ with $(C, L, \bar{e}, \bar{f}) \in O_A^*$. We say a tuple $\bar{o} \in O_A^*$ was *introduced* by some Rule i application, if $\bar{o} \in O_A^j$, $\bar{o} \notin O_A^{j-1}$, and $\mathcal{I}^j$ was obtained from $\mathcal{I}^{j-1}$ by a Rule i application. Note that there can be no duplicate tuples introduced to O_A^* since the last component in $\bar{o}$ is always constructed using fresh elements.

An easy analysis of the construction of $\mathcal{I}$ reveals the following.

Lemma 18. *For every tuple $(C, L, \bar{e}, \bar{e}')$ ever added to O_A^*, $\bar{e}$ is the ensemble in $\mathcal{U}_{C,L}$ that contains $d_{C,L}$ and $\bar{e}'$ is an isomorphic copy of $\bar{e}$ in $\mathcal{I}_L$.*

Intuitively, this means that C and L uniquely determine the ensemble $\bar{e}$ and up to isomorphism also $\bar{e}'$. We can thus classify these tuples by just the concept and level, which helps to give an upper bound on the number of such concepts per level.

Lemma 19. *Let $L \in \mathbf{A}_{\mathcal{O}}$ and $S_L = \{C \mid (C, L, \bar{e}, \bar{e}') \in O_A^*\}$. Then $|S_L| \leq ||\mathcal{O}||^2 + ||\mathcal{O}||$.*

Proof. For every level $L \in \mathbf{A}_{\mathcal{O}}$, let $\mathsf{d}(L)$ denote the number of descendants of L in $G_{\mathcal{O}}$ (not including L). The following is easy to see.

Claim 1. Let $L \in \mathbf{A}_{\mathcal{O}}$ be a level with $\mathsf{d}(L) = n$ such that L has the children $L_1, \dots, L_m$ in $G_{\mathcal{O}}$. Then

$$\mathsf{d}(L) = m + \sum_{i=1}^{m} \mathsf{d}(L_i).$$

We next prove the following central claim.

Claim 2 For any level $L \in \mathbf{A}_{\mathcal{O}}$, if $\mathsf{d}(L) \leq n$, then $|S_L| \leq (n+1) \cdot (||\mathcal{O}|| + 1)$.

Proof of claim. We do an induction on n.

Base case: $n = 0$. Assume that there is a level $L \in \mathbf{A}_{\mathcal{O}}$ with $\mathsf{d}(L) = 0$. Thus L has no child in $G_{\mathcal{O}}$. By construction of $\mathcal{I}$, it is clear that the abstraction graph $(\mathcal{I}, \prec^{-1})$ of $\mathcal{I}$ is a subgraph of $G_{\mathcal{O}}$. Thus L is also a leaf abstraction level in $(\mathcal{I}, \prec^{-1})$. Then all tuples $\bar{o} \in O_A^*$ with $L^{\bar{o}} = L$ must be introduced by Rule 1 applications plus one entry for the initial $\mathcal{I}_{\top,L}$ or $\mathcal{I}_{C_0,L}$. It cannot be that Rule 3 introduces such a $\bar{o}$ since there is no abstraction level that elements of L refine to. For every tuple $\bar{o} \in O_A^*$ with $L^{\bar{o}} = L$ that is introduced by Rule 1, it is clear by definition of Rule 1, that $C^{\bar{o}}$ depends only on some CI of the form $A \sqsubseteq \exists r.B$ in $\mathcal{O}$. Since there are of course at most $||\mathcal{O}||$ such CIs in $\mathcal{O}$, we obtain an upper bound of $|S_L| \leq ||\mathcal{O}|| + 1$ which proves the IH for $n = 0$.

Induction step: $n = i$. Assume that the IH holds for $n = i - 1$. We want to prove that it is then also satisfied for $n = i$. Let $L \in \mathbf{A}_{\mathcal{O}}$ be such that $\mathsf{d}(L) \leq i$. If $\mathsf{d}(L) < i$, then we can apply the IH to obtain $|S_L| \leq i \cdot (||\mathcal{O}|| + 1)$ and are done. Hence let us assume that $\mathsf{d}(L) = i$. Let us also assume that L has children $L_1, \dots, L_m$ in $G_{\mathcal{O}}$.

We want to find an upper bound on the size of S_L so let us again consider how tuples $\bar{o} \in O_A^*$ with $L^{\bar{o}} = L$ are introduced to O_A^*. For the same reasons as in the base case, there can be at most $||\mathcal{O}|| + 1$ concepts in S_L that come from tuples in O_A^* introduced by Rule 1 and the initial tuple in O_A^* for level L.

In contrast to the base case, there can now also be tuples $\bar{o} \in O_A^*$ with $L^{\bar{o}} = L$ that have been introduced by a Rule 3 application on some tuple $\bar{o}' \in O_A^*$ with $L^{\bar{o}'} = L_i$ for some $i \in \{1, \dots, m\}$. Let k be the number of concepts $C \in S_L$ such that there is such a tuple $\bar{o}$ with $C^{\bar{o}} = C$. It is clear that $|S_L| \leq ||\mathcal{O}|| + 1 + k$ and thus it remains to analyze k.

If Rule 3 is applied on a tuple $(C, L_i, \bar{e}, \bar{e}') \in O_A^*$ and introduces a tuple $(D, L, \bar{f}, \bar{f}')$ to O_A^*, then D depends only on C and L. Recall that S_{L_i} is the set of concepts C such that there is a tuple $\bar{o}' \in O_A^*$ with $C^{\bar{o}'} = C$ and $L^{\bar{o}'} = L_i$. In the worst case (upper bound) each of these tuples (for each of the children of L) induces a Rule 3 application that generates a tuple with a fresh concept for level L. Thus $k \leq \sum_{i=1}^{m} S_{L_i}$. This is the starting point of the following chain of inequalities.

$$\begin{aligned}
k &\leq \sum_{i=1}^{m} S_{L_i} && (1)\\
&\leq \sum_{i=1}^{m} \big((\mathsf{d}(L_i) + 1) \cdot (||\mathcal{O}|| + 1)\big) && (2)\\
&= (||\mathcal{O}|| + 1) \cdot \sum_{i=1}^{m} (\mathsf{d}(L_i) + 1) && (3)\\
&= (||\mathcal{O}|| + 1) \cdot \big(m + \sum_{i=1}^{m} (\mathsf{d}(L_i))\big) && (4)\\
&= (||\mathcal{O}|| + 1) \cdot \big(m + d(L) - m\big) && (5)\\
&\leq (||\mathcal{O}|| + 1) \cdot i && (6)
\end{aligned}$$

From 1 to 2 we use the IH on S_{L_i}. From 4 to 5 we apply Claim 1 on the sum. From 5 to 6, we apply the assumption that $d(L) \leq i$ since we are in the case that $n = i$. We obtain:

$$\begin{aligned}
|S_L| &\leq ||\mathcal{O}|| + 1 + k\\
&\leq ||\mathcal{O}|| + 1 + (||\mathcal{O}|| + 1) \cdot i\\
&= (||\mathcal{O}|| + 1) \cdot (i + 1)
\end{aligned}$$

which finishes the induction step for $n = i$ and concludes the proof of the claim.

We have $\mathsf{d}(L) \leq |\mathbf{A}_{\mathcal{O}}| \leq ||\mathcal{O}||$ for all $L \in \mathbf{A}_{\mathcal{O}}$. Thus Claim 2 implies $|S_L| \leq ||\mathcal{O}|| \cdot (||\mathcal{O}|| + 1)$ for all $L \in \mathbf{A}_{\mathcal{O}}$, as required. ❑

We call an element $d \in \Delta^{\mathcal{I}}$ an *abstraction element* if $\rho_L^{\mathcal{I}}(d) = \bar{e}$ and $\bar{e}$ is an abstraction ensemble in $\mathcal{I}$. Recall that in the PSPACE algorithm we guess (in Line 3) a set of sets of concept names that represent these abstraction elements.

Now we are ready to prove the upper bound for the size of those sets.

Lemma 20. *For all $L \in \mathbf{A}_{\mathcal{O}}$, the set*

$$Y_L = \{\mathsf{CN}_{\mathcal{I}}(d) \mid d \in \Delta^{\mathcal{I}_L} \text{ and } d \text{ is abstraction element}\}$$

satisfies $|Y_L| \leq ||\mathcal{O}||^3 + ||\mathcal{O}||^2$.

Proof. Let $d, d' \in \Delta^{\mathcal{I}_L}$ be abstraction elements. Then there are tuples $\bar{o}_1 = (C, L', \bar{e}, \bar{e}')$ and $\bar{o}_2 = (C, L'', \bar{f}, \bar{f}')$ in O^*_A, such that $\rho^{\mathcal{I}}_{L'}(d) = \bar{e}'$ and $\rho^{\mathcal{I}}_{L''}(d') = \bar{f}'$. We show that if $L' = L''$, then $\mathsf{CN}_{\mathcal{I}}(d) = \mathsf{CN}_{\mathcal{I}}(d')$.

We have $\rho^{\mathcal{I}}_{L'}(d) = \bar{e}'$ and $\rho^{\mathcal{I}}_{L'}(d') = \bar{f}'$, and this must have been set by applications of Rule 3. An analysis of this rule shows that then there must be f, f' such that $\rho^{\mathcal{U}_{C,L}}_{L'}(f) = \bar{e}$ and $\rho^{\mathcal{U}_{C,L}}_{L'}(f') = \bar{f}$. By Lemma 18, we have $\bar{e} = \bar{f}$ and thus also $f = f'$. Again considering the definition of Rule 3, we must have $\mathsf{CN}_{\mathcal{I}}(d) = \mathsf{CN}_{\mathcal{I}}(d')$ as desired.

To $|Y_L|$, it thus suffices to count the number of distinct pairs (C, L) in tuples $(C, L, \bar{e}, \bar{e}')$ ever added to O^*_A, and this is bounded from above by $||\mathcal{O}|| \cdot (||\mathcal{O}||^2 + ||\mathcal{O}||)$ due to Lemma 19. ❑

For an abstraction ensemble $\bar{e}$, we call an element $d \in \bar{e}$ the *origin* of $\bar{e}$, if there is a role edge $(f, d) \in r^{\mathcal{I}_{L(d)}}$ such that f is not part of $\bar{e}$. The construction of $\mathcal{I}$ implies that there is at most one origin per abstraction ensemble.

As outlined in the main part of the paper, we want to construct pseudo-models with MCCs of polynomial size by using a uniformized universal model. In the following section, we have to prove that tame matches of abstraction CQs to such a pseudo-model imply a match to a standard model. We next prove an intermediary lemma that helps with this.

Intuitively, if an abstraction ensemble has an origin o, then the concept names that o satisfies determine the whole ensemble that o is part of.

Lemma 21. *Let $\bar{d}$ be an abstraction ensemble in $\mathcal{I}_L$ for some $L \in \mathbf{A}_{\mathcal{O}}$, and d the origin of $\bar{d}$. If $\mathsf{CN}_{\mathcal{I}}(d) = \mathsf{CN}_{\mathcal{I}}(e)$ for some $e \in \Delta^{\mathcal{I}_L}$, then e is part of an ensemble $\bar{e}$ such that there is a homomorphism h from $\mathcal{I}|_{\bar{d}}$ to $\mathcal{I}|_{\bar{e}}$ with $h(d) = e$.*

Proof. Since $\bar{d}$ is an abstraction ensemble, there was an application of Rule 3 that sets $\rho^{\mathcal{I}}(f, L) = \bar{d}$ for some f. Say that this application has processed the tuple $(C, L, \bar{g}, \bar{d})$. By Lemma 18, $\bar{g}$ is the ensemble of $\mathcal{U}_{C,L}$ that contains $d_{C,L}$ and $\bar{d}$ is an isomorphic copy thereof. Moreover, the construction of $\mathcal{I}$ ensures that the origin d of $\bar{d}$ is a copy of $d_{C,L}$, and $C = \bigsqcap S$ for some $S \subseteq \mathsf{CN}_{\mathcal{I}}(d)$. To see this, note that if $(C, L, \bar{g}, \bar{d})$ was added to O^*_A initially or by Rule 3, then no element of $\bar{d}$ has a role predecessor outside of $\bar{d}$, and thus $\bar{d}$ has no origin. This is because roles outside of ensembles are only introduced by Rule 1 applications, and Rule 1 always introduces fresh elements as role successors. Thus $(C, L, \bar{g}, \bar{d})$ was added to O^*_A by Rule 1 and an easy analysis of this rule shows that d is a copy of $d_{C,L}$ and $C = \bigsqcap S$ for some $S \subseteq \mathsf{CN}_{\mathcal{I}}(d)$.

By the universality of $\mathcal{U}_{C,L}$, there is a homomorphism h from $\mathcal{U}_{C,L}$ to $\mathcal{I}$ with $h(d_{C,L}) = e$. Since $\bar{g}$ is an ensemble in $\mathcal{U}_{C,L}$, there is a g such that $\rho^{\mathcal{U}_{C,L}}(g, L) = \bar{g}$ for some g. By Point 5 of the definition of homomorphisms, we have $\rho^{\mathcal{I}}(h(g), L) = h(\bar{g})$. From $h(d_{C,L}) = e$ and $d_{C,L}$ being part of $\bar{g}$, we know that e is part of the ensemble $\bar{e} = h(\bar{g})$. Points 3 and 4 of the definition of homomorphisms (between A-interpretations) yields the desired homomorphism (between $\mathcal{EL}$-interpretations) from $\mathcal{I}|_{\bar{d}}$ to $\mathcal{I}|_{\bar{e}}$. ❑

C.2 Pseudo-Models with Small Component Size

The aim of this section is to prove:

Lemma 4. *If C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$, then there is an L_0-pseudo-model $\mathcal{I}$ of C_0 and $\mathcal{O}$ such that each MCC of $\mathcal{I}$ has at most $2 \cdot (||\mathcal{O}||^2 + ||\mathcal{O}||)$ elements.*

Let $\mathcal{U}$ be the uniform universal L_0-model of C_0 and $\mathcal{O}$ with $d_{\mathcal{U}} \in C_0^{\mathcal{U}_{L_0}}$. We construct a sequence of A-interpretations $\mathcal{I}^0, \mathcal{I}^1, \ldots$. The A-interpretation that we are interested in is then obtained as the limit of the sequence $\mathcal{I}^0, \mathcal{I}^1, \ldots$.

The construction of $\mathcal{I}^0, \mathcal{I}^1, \ldots$ is such that for all $i \geq 0$ and all $L \in \mathbf{A}_{\mathcal{O}}$, we have $\Delta^{\mathcal{I}^i_L} \subseteq \Delta^{\mathcal{U}_L}$.

We start with some preliminaries. For every tuple of elements $\bar{e}$ over $\Delta^{\mathcal{U}_L}$, for some $L \in \mathbf{A}_{\mathcal{O}}$, we use $S(\bar{e})$ to denote a smallest subset of $\Delta^{\mathcal{U}_L}$ that satisfies the following properties:

1. all elements of $\bar{e}$ are in $S(\bar{e})$;
2. if $e \in S(\bar{e})$ and there is a CI $A \sqsubseteq_L \exists r.B \in \mathcal{O}$ such that the MCC of $\Delta^{\mathcal{U}_L}$ that contains e has elements $d, e_{r,B}$ such that $(d, e_{r,B}) \in r^{\mathcal{U}_L}$, $e_{r,B} \in B^{\mathcal{U}_L}$, there is no ensemble in $\mathcal{U}$ that contains both d and $e_{r,B}$, and $e_{r,B}$ is the element satisfying the fewest concept names in addition to the other conditions, then choose such an $e_{r,B}$ and include it in $S(\bar{e})$;
3. if $d \in S(\bar{e})$ is part of the ensemble $\bar{e}'$ in $\mathcal{U}$, then include in $S(\bar{e})$ all elements of $\bar{e}'$.

For every element $e \in S(\bar{e})$, introduce a fresh element $e^\dagger$ and set $S^\dagger(\bar{e}) := \{e^\dagger \mid e \in S(\bar{e})\}$.

For every $L \in \mathbf{A}_{\mathcal{O}}$ and tuple of elements $\bar{e}$ over $\Delta^{\mathcal{U}_L}$, we define an $\mathcal{EL}$-interpretation $\mathcal{J}_{\bar{e}}$. This proceeds in two steps. In the first step, we set

$$\begin{aligned}
\Delta^{\mathcal{J}_{\bar{e}}} &= S(\bar{e}) \cup S^\dagger(\bar{e}) \\
A^{\mathcal{J}_{\bar{e}}} &= \{e, e^\dagger \mid e \in A^{\mathcal{U}_L} \cap S(\bar{e})\} \\
r^{\mathcal{J}_{\bar{e}}} &= \{(d, e), (d^\dagger, e^\dagger) \mid (d, e) \in r^{\mathcal{U}_L} \cap (S(\bar{e}) \times S(\bar{e}))\}.
\end{aligned}$$

In the second step, we further extend $r^{\mathcal{J}_{\bar{e}}}$ as follows, for each role name r. Consider every element $d \in S(\bar{e})$ and CI $A \sqsubseteq_L \exists r.B \in \mathcal{O}$ such that $d \in A^{\mathcal{J}_{\bar{e}}} \setminus (\exists r.B)^{\mathcal{J}_{\bar{e}}}$. Then $d \in A^{\mathcal{U}_L}$ and since $\mathcal{U}_L$ is a model of $\mathcal{O}_L$, there is a $(d, e) \in r^{\mathcal{U}_L}$ with $e \in B^{\mathcal{U}_L}$. There is no ensemble in $\mathcal{U}$ that contains both d and e, as then we would have $e \in S(\bar{e})$, implying $(d, e) \in r^{\mathcal{J}_{\bar{e}}}$ and $e \in B^{\mathcal{J}_{\bar{e}}}$, contradicting $d \notin (\exists r.B)^{\mathcal{J}_{\bar{e}}}$. It follows that an element $e_{r,B}$ was chosen in Step 2 above and we can set

$$s^{\mathcal{J}_{\bar{e}}} = s^{\mathcal{J}_{\bar{e}}} \cup \{(d, e^\dagger_{r,B}), (d^\dagger, e_{r,B})\} \text{ for all } s \in R_{r,\mathcal{O}_L}.$$

It is easy to verify that every $\mathcal{EL}$-interpretation $\mathcal{J}_{\bar{e}}$ is a model of $\mathcal{O}_L$. Note that we use the copies $d^\dagger$ of elements d to avoid adding edges to ensembles in the second step that do not exist in $\mathcal{U}_L$.

We now construct the interpretation $\mathcal{I}^0$. Set $\rho^{\mathcal{I}_0}$ to be empty, $\mathbf{A}^{\mathcal{I}_0} = \mathbf{A}_\mathcal{O}$ and $\prec^{\mathcal{I}_0} = \prec_\mathcal{O}$. It remains to populate the $\mathcal{EL}$-interpretations $\mathcal{I}_L^0$, for all $L \in \mathbf{A}_\mathcal{O}$. Let $L \in \mathbf{A}_\mathcal{O}$. We choose an initial element $d_L \in \Delta^{\mathcal{U}_L}$ and then set $\mathcal{I}_L^0 = \mathcal{J}_{S(d_L)}$. To make precise the choice of d_L, recall that we have built $\mathcal{U}$ by starting with $\mathcal{U}_L = \mathcal{I}_{C,L}$ where $C = C_0$ if $L = L_0$ and $C = \top$ otherwise. Further, recall that $\mathcal{I}_{C,L}$ is a restriction of the universal model $\mathcal{U}_{C,L}$ with distinguished element $d_{C,L}$. We choose d_L to be $d_{C,L}$ (and consequently $d_L = d_\mathcal{U}$ if $L = L_0$).

The rest of the sequence of A-interpretations $\mathcal{I}^0, \mathcal{I}^1, \dots$ is obtained by repeatedly satisfying abstraction and refinement statements in $\mathcal{O}$, guided by $\mathcal{U}$, each time adding an MCC in essentially the same way as described above. This is done in a fair way, that is, we do not defer the satisfaction of any abstraction and refinement statement forever so that in the A-interpretation $\mathcal{I}$ obtained in the limit of $\mathcal{I}^0, \mathcal{I}^1, \dots,$ all abstractions and refinements are satisfied. We now give details.

There are two rules for obtaining $\mathcal{I}^{i+1}$ from $\mathcal{I}^i$ and we assume that they are applied in a fair way to obtain the sequence $\mathcal{I}^0, \mathcal{I}^1, \dots$. For a tuple of elements $\bar{e} = (e_1, \dots, e_n)$, we may write $\bar{e}^\dagger$ to denote $(e_1^\dagger, \dots, e_n^\dagger)$.

Rule 1. Choose a $d \in \Delta^{\mathcal{I}^i}$ such that $\rho_L^\mathcal{U}(d) = \bar{e}$ and $\rho_L^{\mathcal{I}^i}(d)$ is undefined.

Consider the $\mathcal{EL}$-interpretation $\mathcal{J}_{\bar{e}}$. Using the construction and structure of $\mathcal{U}$ and $\mathcal{I}^i$ and our choice of initial elements, it can be verified that none of the elements of $\Delta^{\mathcal{J}_{\bar{e}}}$ occur in $\Delta^{\mathcal{I}^i}$.[5]

Now define $\mathcal{I}^{i+1}$ to be $\mathcal{I}^i$, modified as follows:

- $\rho_L^{\mathcal{I}^{i+1}}(d) = \bar{e}$ and $\rho_L^{\mathcal{I}^{i+1}}(d^\dagger) = \bar{e}^\dagger$;
- $\mathcal{I}_L^{i+1}$ is the disjoint union of $\mathcal{I}_L^i$ and $\mathcal{J}_{\bar{e}}$.

Rule 2. Choose a tuple $\bar{e}$ over $\Delta^{\mathcal{I}_L^i}$, for some $L \in \mathbf{A}_\mathcal{O}$, such that $\rho_L^\mathcal{U}(d) = \bar{e}$ and there is no d' such that $\rho_L^{\mathcal{I}^i}(d') = \bar{e}$.

Consider the $\mathcal{EL}$-interpretation $\mathcal{J}_d$. It can be verified that none of the elements of $\Delta^{\mathcal{J}_d}$ occur in $\Delta^{\mathcal{I}^i}$. Now define $\mathcal{I}^{i+1}$ to be $\mathcal{I}^i$, modified as follows:

- $\rho_L^{\mathcal{I}^{i+1}}(d) = \bar{e}$ and $\rho_L^{\mathcal{I}^{i+1}}(d^\dagger) = \bar{e}^\dagger$;
- $\mathcal{I}_L^{i+1}$ is the disjoint union of $\mathcal{I}_L^i$ and $\mathcal{J}_d$.

The following lemma is a straightforward implication of the construction of $\mathcal{I}$.

Lemma 22. *$\mathcal{I}$ has the following properties:*

1. $\Delta^\mathcal{I} \subseteq \{d, d^\dagger \mid d \in \Delta^\mathcal{U}\}$;
2. $\mathsf{CN}_\mathcal{I}(d) = \mathsf{CN}_\mathcal{I}(d^\dagger) = \mathsf{CN}_\mathcal{U}(d)$ *for all* $d \in \Delta^\mathcal{I}$;
3. $\rho_L^\mathcal{I}(d) = \rho_L^\mathcal{U}(d)$ *and* $\rho_L^\mathcal{I}(d^\dagger) = \rho_L^\mathcal{U}(d)^\dagger$ *for all* $L \in \mathbf{A}_\mathcal{O}$ *and* $d \in \Delta^\mathcal{I}$;
4. $\mathcal{I}|_{\bar{e}}$ *and* $\mathcal{I}|_{\bar{e}^\dagger}$ *are isomorphic to* $\mathcal{U}|_{\bar{e}}$, *for all ensembles* $\bar{e}$ *in* $\mathcal{I}$.

Now we prove that $\mathcal{I}$ is in fact the desired pseudo-model. Recall that for tameness, we need to consider the graph $G_{h,\mathcal{I}}$ for a homomorphism h from some CQ q to $\mathcal{I}$. For readability, we often index h with the interpretation it maps to (for example $h_\mathcal{I}$) and will in such cases omit the $\mathcal{I}$ from the lower index of $G_{h,\mathcal{I}}$.

Lemma 23. *$\mathcal{I}$ is a L_0-pseudo-model of C_0 and $\mathcal{O}$.*

Proof. We have $C_0^{\mathcal{I}_{L_0}^0} \neq \emptyset$ and thus also $C_0^{\mathcal{I}_{L_0}} \neq \emptyset$. In the construction of $\mathcal{I}$, we already argued that each of the $\mathcal{J}_{\bar{e}}$, with $\bar{e}$ a tuple over $\Delta^{\mathcal{U}_L}$, is a model of $\mathcal{O}_L$. Hence it is easy to verify that $\mathcal{I}$ satisfies all concept inclusions, role inclusions, and range restrictions in $\mathcal{O}$. What remains to show is that concept refinements and abstractions are also satisfied. For simplicity, we consider only the elements of $\Delta^\mathcal{I}$ that are also elements of $\Delta^\mathcal{U}$, but not their $\cdot^\dagger$-companions. For the latter, all arguments are identical.

Assume that $d \in A^{\mathcal{I}_L}$ and there is a concept refinement $L'{:}\, q(\bar{x})$ refines $L{:}\, A$ in $\mathcal{O}$. Lemma 22 implies $d \in A^{\mathcal{U}_L}$ and since $\mathcal{U}$ is a model of $\mathcal{O}$, $\rho_{L'}^\mathcal{U}(d) = \bar{e}$ is defined and $\bar{e}$ is an answer to q on $\mathcal{U}$. The construction of $\mathcal{I}$ then implies that either Rule 1 or Rule 2 set $\rho_{L'}^\mathcal{I}(d) = \bar{e}$ and Lemma 22 implies that $\bar{e}$ is an answer to q on $\mathcal{I}$, as required.

Assume that $\bar{e}$ is a tame (w.r.t. $\mathcal{O}_L$) answer to a CQ $q(\bar{x})$ on $\mathcal{I}_L$ for some concept abstraction $L'{:}\, A$ abstracts $L{:}\, q(\bar{x})$ in $\mathcal{O}$. Let $h_\mathcal{I}$ be the tame homomorphism with $h_\mathcal{I}(\bar{x}) = \bar{e}$.

We differentiate between three cases. The ensemble $\bar{e}$ might coincide with an ensemble in $\mathcal{I}$, it might not share any constants at all with an ensemble in $\mathcal{I}$, or it might partially overlap with an ensemble in $\mathcal{I}$. We will show that only the first case occurs and satisfies the concept abstraction, the other two lead to a contradiction.

- *Case 1:* $\bar{e}$ coincides with an existing L-ensemble in $\mathcal{I}$. Then by the construction of $\mathcal{I}$, Rule 1 or Rule 2 have set $\rho_L^\mathcal{I}(d) = \bar{e}$ for some $d \in \Delta^\mathcal{I}$. For both rule applications, this implies $\rho_L^\mathcal{U}(d) = \bar{e}$. By Lemma 22, $\bar{e}$ is an answer to q on $\mathcal{U}_L$ and since $\mathcal{U}$ is a model of $\mathcal{O}$ we obtain $d \in A^{\mathcal{U}_{L'}}$. Lemma 22 then implies $d \in A^{\mathcal{I}_{L'}}$, as required.
- *Case 2:* $\bar{e}$ consists of only free elements in $\mathcal{I}_L$. By definition of tameness, $G_{h_\mathcal{I}}$ is then a tree and each node is a singleton class. Hence we assume for simplicity that the nodes are the variables themselves (instead of equivalence classes containing one variable). Let x_R be the root of $G_{h_\mathcal{I}}$. We argue that there is a homomorphism $h_\mathcal{U}$ from q to $\mathcal{U}_L$ with $h_\mathcal{U}(x_R) = h_\mathcal{I}(x_R)$.

  First, we prove a claim about the existence of roles in $\mathcal{U}$ that are similar to roles that span across ensembles in $\mathcal{I}$. It will be helpful here and in the next case.

  Claim 1. Let $e, e' \in \Delta^{\mathcal{I}_L}$ be two elements and $S = \{r \mid (e, e') \in r^{\mathcal{I}_L}\}$ the set of role names that go from e to e'. If e and e' are not part of the same L-ensemble in $\mathcal{I}$, then for any $f \in \Delta^{\mathcal{U}_L}$ with $\mathsf{CN}_\mathcal{I}(e) = \mathsf{CN}_\mathcal{U}(f)$, there is an $f' \in \Delta^{\mathcal{U}_L}$ such that $(f, f') \in r^{\mathcal{U}_L}$ for all $r \in S$ and $\mathsf{CN}_\mathcal{I}(e') = \mathsf{CN}_\mathcal{U}(f')$.

  Proof of the claim. By construction of $\mathcal{I}$, there is a tuple

[5] And even if they did, we could easily deal with this issue by renaming elements; we thus refrain from analyzing this in great depth.

$\bar{e}$ of elements such that $e, e' \in \Delta^{\mathcal{J}_{\bar{e}}}$. Since e and e' are not part of the same ensemble in $\mathcal{I}$, e' must have been introduced by Property 2 in the construction of $\mathcal{J}_{\bar{e}}$ (also in the case where $e = e'$). Consequently, there is a CI $A \sqsubseteq_L \exists r.B \in \mathcal{O}$ such that $e \in A^{\mathcal{I}_L}$, $r \in S$ and $e' = e_{r,B}$ with $e_{r,B}$ the distinguished element chosen for r and B in $\mathcal{J}_{\bar{e}}$.

All edges in S must have then been set in the second step of the construction of $\mathcal{J}_{\bar{e}}$. This implies $S = R_{r,\mathcal{O}_L}$. What remains is to find an f' such that $\mathsf{CN}_{\mathcal{I}}(e') = \mathsf{CN}_{\mathcal{U}}(f')$. For this, we need to argue that e' is 'universal'.

We call an element $d \in \Delta^{\mathcal{I}}$ *concept-name-universal* for a concept C, abstraction level L and ontology $\mathcal{O}$, if in any model $\mathcal{J}$ of $\mathcal{O}$ and element $e \in C^{\mathcal{J}_L}$, we have $\mathsf{CN}_{\mathcal{I}}(d) \subseteq \mathsf{CN}_{\mathcal{J}}(e)$. Note that we do not use the standard universality definition here, since $\mathcal{I}$ might contain reflexive edges. In the following we prove that e' is concept-name-universal for $B \sqcap C_{r,\mathcal{O}_L}$, L, and $\mathcal{O}$.

By Lemma 22, we have $e \in A^{\mathcal{U}_L}$. It is relatively straightforward to see that there cannot be an r-successor g of e in $\mathcal{U}_L$ that satisfies B and is part of the same ensemble as e. If there were such a g, then Property 3 of $S(\bar{e})$ and the definition of $\mathcal{J}_{\bar{e}}$ would have introduced $(e, g) \in r^{\mathcal{I}_L}$ instead of $(e, e') \in r^{\mathcal{I}_L}$.

But since $\mathcal{U}_L$ is a model of $\mathcal{O}$, there still must be an r-succesor $\widehat{e}$ of e that satisfies $B \sqcap C_{r,\mathcal{O}_L}$ (and is thus not part of the same ensemble as e). The construction (uniformity) of $\mathcal{U}$ then implies that $\widehat{e}$ is universal for $B \sqcap C_{r,\mathcal{O}_L}$ and L. Property 2 of $S(\bar{e})$ indicates that we choose as $e_{r,B}$ an element satisfying the fewest concept names (in addition to the other conditions). Since we just showed that there is an element that satisfies the necessary conditions for Property 2 and is also universal for $B \sqcap C_{r,\mathcal{O}_L}$, L, and $\mathcal{O}$, we obtain concept-name-universality of $e_{r,B}$ for $B \sqcap C_{r,\mathcal{O}_L}$, L, and $\mathcal{O}$.

Remember that for proving the claim, we need to prove that there exists an $f' \in \Delta^{\mathcal{U}_L}$ such that $(f, f') \in s^{\mathcal{U}_L}$ for all $s \in S$ and $\mathsf{CN}_{\mathcal{I}}(e') = \mathsf{CN}_{\mathcal{U}}(f')$. This has now become straightforward. First, $\mathsf{CN}_{\mathcal{I}}(e) = \mathsf{CN}_{\mathcal{U}}(f)$ implies that $f \in A^{\mathcal{U}_L}$. Then, $\mathcal{U}$ being a model of $\mathcal{O}$ implies that there is an r-successor f' of f that satisfies $D = B \sqcap C_{r,\mathcal{O}_L}$ (we showed that $A \sqsubseteq_L \exists r.B \in \mathcal{O}$). The construction of $\mathcal{U}$ implies that f is universal for D and L, and we already proved that $e' = e_{r,B}$ is concept-name universal for D, L, and $\mathcal{O}$. Hence $\mathsf{CN}_{\mathcal{I}}(e') = \mathsf{CN}_{\mathcal{U}}(f')$. And finally since $S = R_{r,\mathcal{O}_L}$, $\mathcal{U}$ being a model of $\mathcal{O}$ implies that $(f, f') \in s^{\mathcal{U}_L}$ for all $s \in S$. This finishes the proof of the claim.

Now let us get back to defining the homomorphism $h_{\mathcal{U}}$ from q to $\mathcal{U}_L$. We define it step by step, traversing $G_{h_{\mathcal{I}}}$ from the root x_R to its leaves while also proving that for all $x \in \mathsf{var}(q)$, we have $\mathsf{CN}_{\mathcal{I}}(h_{\mathcal{I}}(x)) = \mathsf{CN}_{\mathcal{U}}(h_{\mathcal{U}}(x))$. We set $h_{\mathcal{U}}(x_R) = h_{\mathcal{I}}(x_R)$. This is a partial homomorphism, since by Lemma 22, they satisfy the same concept names and by tameness there are no reflexive atoms in q. It is trivial to show that $\mathsf{CN}_{\mathcal{I}}(h_{\mathcal{I}}(x_R)) = \mathsf{CN}_{\mathcal{U}}(h_{\mathcal{U}}(x_R))$. Now assume that $h_{\mathcal{U}}(x)$ is already defined, $\mathsf{CN}_{\mathcal{I}}(h_{\mathcal{I}}(x)) = \mathsf{CN}_{\mathcal{U}}(h_{\mathcal{U}}(x))$, (x, y) an edge in $G_{h_{\mathcal{I}}}$ and $h_{\mathcal{U}}(y)$ is undefined. Then we can directly use Claim 1 on $h_{\mathcal{I}}(x), h_{\mathcal{I}}(y)$, $S = \{r \mid (h_{\mathcal{I}}(x), h_{\mathcal{I}}(y)) \in r^{\mathcal{I}}\}$, and $h_{\mathcal{U}}(x)$ to define $h_{\mathcal{U}}(y)$. Hence we obtain a homomorphism $h_{\mathcal{U}}$ from q to $\mathcal{U}_L$.

Obtaining $h_{\mathcal{U}}$ is, however, a contradiction to $\mathcal{U}$ being a model of $\mathcal{O}$. By assumption $h_{\mathcal{I}}(x_R)$ was free, making $h_{\mathcal{U}}(\bar{x})$ free by construction of $\mathcal{I}$ and hence contradicting $\mathcal{U}$ being a model of $\mathcal{O}$. Thus we have proven that $\bar{e}$ cannot consist of only free elements in $\mathcal{I}$.

- *Case 3:* $\bar{e}$ partially overlap with an least one ensemble in $\mathcal{I}$. This means that $\bar{e}$ contains at least one element that is part of an ensemble and another element that is free or part of another ensemble.

  Recall that we call an ensemble $\bar{e}$ in $\mathcal{U}$ an *abstraction ensemble* if it was introduced by a Rule 3 application. Otherwise, we call $\bar{e}$ a *refinement ensemble*. Let c_R be the root of $G_{h_{\mathcal{I}}}$.

  We define $h_{\mathcal{U}}$ step by step, traversing $G_{h_{\mathcal{I}}}$ from its root c_R to its leaves while also proving that for all $x \in \mathsf{var}(q)$, we have $\mathsf{CN}_{\mathcal{I}}(h_{\mathcal{I}}(x)) = \mathsf{CN}_{\mathcal{U}}(h_{\mathcal{U}}(x))$. Start by setting $h_{\mathcal{U}}(x) = h_{\mathcal{I}}(x)$ for all $x \in c_R$. Lemma 22 implies that this is a partial homomorphism. It is trivial to show that $\mathsf{CN}_{\mathcal{I}}(h_{\mathcal{I}}(x)) = \mathsf{CN}_{\mathcal{U}}(h_{\mathcal{U}}(x))$ for all $x \in c_R$.

  Now assume that (c_1, c_2) is an edge in $G_{h_{\mathcal{I}}}$, $h_{\mathcal{U}}(x_1)$ is already defined for all $x_1 \in c_1$ and $h_{\mathcal{U}}$ is undefined for the elements in c_2. Let $x_1 \in c_1$ and $x_2 \in c_2$ be variables such that $r(x_1, x_2) \in q$.

  The construction of $\mathcal{U}$ implies that either $h_{\mathcal{I}}(x_2)$ is free or it is the origin of an abstraction ensemble in $\mathcal{I}$. Let $C = \mathsf{CN}_{\mathcal{I}}(h_{\mathcal{I}}(x_2))$ and $S = \{r \mid (h_{\mathcal{I}}(x_1), h_{\mathcal{I}}(x_2)) \in r^{\mathcal{I}_L}\}$.

  Applying Claim 1 on $h_{\mathcal{I}}(x_1), h_{\mathcal{I}}(x_2)$, S, and $h_{\mathcal{U}}(x_1)$ lets us obtain an $f' \in \Delta^{\mathcal{U}_L}$ such that $(h_{\mathcal{U}}(x_1), f') \in s^{\mathcal{U}_L}$ for all $s \in S$ and $\mathsf{CN}_{\mathcal{I}}(h_{\mathcal{I}}(x_2)) = \mathsf{CN}_{\mathcal{U}}(f')$. Hence setting $h_{\mathcal{U}}(x_2) = f'$ is a partial homomorphism. If $h_{\mathcal{I}}(x_2)$ is free, then we are done.

  Otherwise, $h_{\mathcal{I}}(x_2)$ is the origin of an abstraction ensemble $\bar{f}$ in $\mathcal{I}$. The construction of $G_{h_{\mathcal{I}}}$ implies that all $h_{\mathcal{I}}(x_2')$ with $x_2' \in c_2$ are then part of $\bar{f}$. By Lemma 22, $h_{\mathcal{I}}(x_2)$ is then also the origin of an abstraction ensemble $\bar{f}'$ in $\mathcal{U}$ that is isomorphic to $\bar{f}$. Hence we can extend $h_{\mathcal{U}}$ to the variables in c_2 by applying Lemma 21 on $\bar{f}'$ as $\bar{d}$, $h_{\mathcal{I}}(x_2)$ as d, and $h_{\mathcal{U}}(x_2)$ as e respectively.

  Condition 2 of tameness ensures that there are no role-atoms in q that go from variables in c_1 to c_2 but map to different elements than $h_{\mathcal{I}}(x_1)$ and $h_{\mathcal{I}}(x_2)$.

  We have thus obtained a homomorphism $h_{\mathcal{U}}$ from q to $\mathcal{U}_L$. What remains to prove is that this contradicts $\mathcal{U}$ being a model of $\mathcal{O}$. We do this by proving that $h_{\mathcal{U}}$ spans across multiple ensembles in $\mathcal{U}$.

  By construction, $h_{\mathcal{U}}$ overlaps with the ensemble of the elements in c_R. Our assumption was that $h_{\mathcal{I}}(\bar{x})$ spans across multiple ensembles and hence $G_{h_{\mathcal{I}}}$ contains at least two nodes. Let c be a successor of c_R and $r(x, y) \in q$ with $x \in c_R$ and $y \in c$ the role atom that exists by definition of $G_{h_{\mathcal{I}}}$. Also let $e = h_{\mathcal{I}}(x)$ and $e' = h_{\mathcal{I}}(y)$.

By construction of $\mathcal{I}$, the element e' must satisfy some concept name B not satisfied by any r-successor of e that is part of the L-ensemble of e. Lemma 22 then implies that $h_{\mathcal{U}}(x)$ does not have an r-successor that is part of the same ensemble and satisfies B. Hence $h_{\mathcal{U}}(y)$ must be part of a different ensemble (including no ensemble at all) than $h_{\mathcal{U}}(x)$. This contradicts $\mathcal{U}$ being a model of $\mathcal{O}$. ❑

We analyze the size of $\mathcal{I}$ componentwise since our algorithm guesses these components in the form of mosaics.

Lemma 24. *Each MCC in $\mathcal{I}$ has at most $2 \cdot (||\mathcal{O}||^2 + ||\mathcal{O}||)$ elements.*

Proof. By construction of $\mathcal{I}$, for every MCC $\mathcal{J}$ of $\mathcal{I}$ there is an ensemble $\bar{e}$ in $\mathcal{I}$ such that $\Delta^{\mathcal{J}} \subseteq S(\bar{e})$, with $S(\bar{e})$ defined as in the construction of $\mathcal{J}$. The ensemble $\bar{e}$ may contain at most $||\mathcal{O}||$ elements. $S(\bar{d})$ might introduce an element $e_{r,B}$ for every role name r and concept name B that appear in $\mathcal{O}$. Each of these $e_{r,B}$ might also be part of an ensemble. This results in adding at most $||\mathcal{O}|| \cdot ||\mathcal{O}||$ elements. for a total of $||\mathcal{O}||^2 + ||\mathcal{O}||$ elements. Finally, we also introduce $\cdot^{\dagger}$ copies of each of these elements hence doubling the total number. ❑

Now we are ready to prove the correctness of our PSPACE algorithm.

C.3 Soundness

Assume that the algorithm returned 'true'. We need to prove that there is an L_0-model $\mathcal{I}$ of C_0 and $\mathcal{O}$.

We construct $\mathcal{I}$ in two steps. First, we define an unraveling of individual mosaics and then we add the refinements and abstractions to these unravelings. Let $M = (\mathcal{I}, L, E, \bar{e})$ be a mosaic. We use E^+ to denote E extended with all length one tuples (d) such that $d \in \Delta^{\mathcal{I}}$ is not part of any ensemble in E. Note that $\bar{e} \in E$ unless $\bar{e}$ is the empty tuple. A *path* in M is a sequence

$$p = \bar{e}_1\, d_2 r_2 f_2 \bar{e}_2\ d_3 r_3 f_3 \bar{e}_3 \cdots d_k r_k f_k \bar{e}_k$$

where $k \geq 1$ and for $1 \leq i \leq k$, $\bar{e}_i \in E^+$, d_i is an element in $\bar{e}_{i-1}$, f_i is an element in $\bar{e}_i$, and each r_i is a role name such that the following conditions are satisfied:

1. $(d_i, f_i) \in r_i^{\mathcal{I}}$ for $1 \leq i < k$;
2. if $\bar{e}_i = \bar{e}_{i+1}$, then $\bar{e}_i \notin E$.

The last condition ensures that we do not unravel edges inside of ensembles. We cannot generally demand $\bar{e}_i \neq \bar{e}_{i+1}$ because reflexive loops of non-ensemble elements must be unraveled. We use $\mathsf{headE}(p)$ to denote the ensemble $\bar{e}_1$ and likewise for $\mathsf{tailE}(p)$ and $\bar{e}_k$. Now the *unraveling* of M *at a tuple* $\bar{d} \in E^+$ is the interpretation $\mathcal{J}$ defined by setting

$$\begin{aligned} \Delta^{\mathcal{J}} &= \{(p,d) \mid p \text{ path in } M \text{ with } \mathsf{headE}(p) = \bar{d} \\ &\qquad \text{and } d \in \mathsf{tailE}(p)\} \\ A^{\mathcal{J}} &= \{(p,d) \mid d \in A^{\mathcal{I}}\} \\ r^{\mathcal{J}} &= \{(p,d),(p,e) \mid (d,e) \in r^{\mathcal{I}} \text{ and } \mathsf{tailE}(p) \in E\} \cup \\ &\qquad \{((p,d),(p',f)) \mid p' = p\, d r f \bar{e} \text{ for some } \bar{e}\}. \end{aligned}$$

The condition $\mathsf{tailE}(p) \in E$ is there to prevent reflexive loops of non-ensemble elements to be included.

Let $\mathcal{M}$ be the set of all mosaics that our algorithm guessed during its successful run.

Lemma 25. *Any unraveling $\mathcal{J}$ of a mosaic $M \in \mathcal{M}$ at some tuple is a model of $\mathcal{O}_{L^M}$.*

Proof. Let $M \in \mathcal{M}$ with $M = (\mathcal{I}, L, E, \bar{e})$ be such a mosaic. By definition of mosaics, $\mathcal{I}$ is a model of $\mathcal{O}_L$. What remains to be proved is that the same holds for $\mathcal{J}$. We first consider the different forms of CIs that any element $d \in \Delta^{\mathcal{J}}$ has to satisfy:

- it is straightforward to see that CIs of the form $\top \sqsubseteq_L A \in \mathcal{O}$ and $A_1 \sqcap \ldots, \sqcap A_n \sqsubseteq_L A \in \mathcal{O}$ are satisfied, by definition of $\mathcal{J}$ and the fact that $\mathcal{I}$ is a model of $\mathcal{O}_L$;
- if $(p,d) \in A^{\mathcal{J}_L}$ and $A \sqsubseteq_L \exists r.B \in \mathcal{O}$, then there is an $e \in B^{\mathcal{I}_L}$ and $(d,e) \in r^{\mathcal{I}_L}$ by $\mathcal{I}$ being a model of $\mathcal{O}_L$.

  If both d and e are part of the same ensemble in E, then $((p,d),(p,e)) \in r^{\mathcal{J}}$ and $(p,e) \in B^{\mathcal{J}}$ by definition of $\mathcal{J}$.

  If $d = e$ and d is not part of an ensemble in E, then $((p,d),(pdrd(d),d)) \in r^{\mathcal{J}}$ and $(pdrd(d),d) \in B^{\mathcal{J}}$ by definition of $\mathcal{J}$.

  If d and e are not part of the same ensemble in E and $d \neq e$, then let $\bar{f} \in E^+$ be the tuple with e in $\bar{f}$. Again the definition of $\mathcal{J}$ implies that $((p,d),(pdre\bar{f},e)) \in r^{\mathcal{J}}$ and $(pdre\bar{f},e) \in B^{\mathcal{J}}$, as required;
- if $(p,d) \in (\exists r.A)^{\mathcal{J}}$ and $\exists r.A \sqsubseteq_L B \in \mathcal{O}$, then by semantics, there is an $(p',e) \in A^{\mathcal{J}}$ with $((p,d),(p',e)) \in r^{\mathcal{J}}$. The definition of $\mathcal{J}$ and paths then imply that $(d,e) \in r^{\mathcal{I}}$ and $e \in A^{\mathcal{I}}$. Since $\mathcal{I}$ is a model of $\mathcal{O}_L$, we have $d \in B^{\mathcal{I}}$ and thus $(p,d) \in B^{\mathcal{J}}$, as required.

What remains to be proved is that role inclusions and range restrictions are satisfied. Let $((p,d),(p',f)) \in r^{\mathcal{J}}$. By definition of $\mathcal{J}$, there are two cases. If $p = p'$, then $(d,f) \in r^{\mathcal{I}}$. Since $\mathcal{I}$ is a model of $\mathcal{O}$, it is then easy to see that any role inclusion and range restriction with regards to r is satisfied in $\mathcal{J}$.

Otherwise, we have $p' = pdrf\bar{e}$ for some $\bar{e} \in E^+$. The definition of paths implies that $(d,f) \in r^{\mathcal{I}}$ and clearly any role inclusion and range restriction is again satisfied. ❑

Next, we show that for homomorphisms of connected CQs into the unraveling, we can find tame homomorphisms that match into the pseudo-model.

Lemma 26. *Let $\mathcal{J}$ be the unraveling of any $M \in \mathcal{M}$ at some tuple and with $M = (\mathcal{I}, L, E, \bar{e})$. For every connected CQ q and homomorphism $h_{\mathcal{J}}$ from q to $\mathcal{J}$, there is a tame homomorphism $h_{\mathcal{I}}$ from q to $\mathcal{I}$.*

Proof. For all $x \in \mathsf{var}(q)$ and $h_{\mathcal{J}}(x) = (p,d)$, we set $h_{\mathcal{I}}(x) = d$. The definition of $\mathcal{J}$ implies that if $(p,d) \in A^{\mathcal{J}}$, then $d \in A^{\mathcal{I}}$ and that if $((p,d),(p',e)) \in r^{\mathcal{J}}$, then $(d,e) \in r^{\mathcal{I}}$. Thus $h_{\mathcal{I}}$ is an answer to q on $\mathcal{I}$.

What remains to be proved is that $h_{\mathcal{I}}$ is tame. We first prove that $h_{\mathcal{J}}$ is tame and then use this to show that $h_{\mathcal{I}}$ must be tame as well. Note, however, that currently $\mathcal{J}$ is not equipped with a set of ensembles. Hence let $E_{\mathcal{J}}$ be the set of ensembles of $\mathcal{J}$, defined by

$$E_{\mathcal{J}} = \{((p, d_1), \dots, (p, d_n)) \mid (p, d_i) \in \Delta^{\mathcal{J}} \text{ for } 1 \le i \le n \text{ and } (d_1, \dots, d_n) \in E\}.$$

Let us consider the directed graph $G_{h_\mathcal{J},\mathcal{J}}$ as defined in the main part of the paper. We must show that Conditions 1 and 2 of tameness are satisfied. For Condition 1, we have to show that $G_{h_\mathcal{J},\mathcal{J}}$ is a tree, possibly with self-loops on ensemble nodes.

By assumption, q is connected which implies that $G_{h_\mathcal{J},\mathcal{J}}$ is connected. Now take any edge (c_1, c_2) in $G_{h_\mathcal{J},\mathcal{J}}$ with $c_1 \neq c_2$. Then q contains an atom $r(x_1, x_2)$ with $x_1 \in c_1$ and $x_2 \in c_2$, and we have $(h_\mathcal{J}(x_1), h_\mathcal{J}(x_2)) \in r^\mathcal{J}$. By construction of $G_{h_\mathcal{J},\mathcal{J}}$, $c_1 \neq c_2$ implies that $h_\mathcal{J}(x_1)$ and $h_\mathcal{J}(x_2)$ are not part of the same ensemble in $\mathcal{J}$. Consequently, if $h_\mathcal{J}(x_1) = (p_1, d_1)$ and $h_\mathcal{J}(x_2) = (p_2, d_2)$, then $p_1 \neq p_2$. From $(h_\mathcal{J}(x_1), h_\mathcal{J}(x_2)) \in r^\mathcal{J}$ and the definition of $\mathcal{J}$ we obtain that p_2 is a *longer* path than p_1. Clearly, this property precludes the existence of cycles in $G_{h_\mathcal{J},\mathcal{J}}$.

Regarding self-loops, consider any edge (c, c) in $G_{h_\mathcal{J},\mathcal{J}}$. We have to show that c is an ensemble node. Assume that it is not. As mentioned in the main part of the paper, then $c = \{x\}$ is a singleton. Thus q contains an atom $r(x, x)$ with $x \in c$. We have $(h_\mathcal{J}(x), h_\mathcal{J}(x)) \in r^\mathcal{J}$. By definition of $\mathcal{J}$, this implies that $h_\mathcal{J}(x) = (p, d)$ with $\mathsf{tailE}(p) \in E$. But this means that c is an ensemble node, which is a contradiction.

Now for Condition 2. Let (c_1, c_2) be an edge in $G_{h_\mathcal{J},\mathcal{J}}$. Then there are ensembles $\bar{e}_1, \bar{e}_2 \in E_\mathcal{J}$ such that for all $x \in c_i$, we have $h_\mathcal{J}(x) \in \bar{e}_i$, for $i \in \{1, 2\}$. By definition of $E_\mathcal{J}$, there are thus paths p_1, p_2 such that for all $x \in c_i$, $h_\mathcal{J}(x)$ is of the form (p_i, d) for some d and for $i \in \{1, 2\}$. Since (c_1, c_2) is an edge in $G_{h_\mathcal{J},\mathcal{J}}$, q must contain an atom $r(x_1, x_2)$ with $x_1 \in c_1$ and $x_2 \in c_2$. Thus p_2 has the form $p_1 d r f \bar{e}$. To see that Condition 2 is satisfied, take any $r(x_1, x_2) \in q$ with $x_1 \in c_1$ and $x_2 \in c_2$. Then $r(h_\mathcal{J}(x_1), h_\mathcal{J}(x_2)) \in r^\mathcal{J}$ and by definition of $\mathcal{J}$ we must have $h_\mathcal{J}(x_1) = (p_1, d)$ and $h_\mathcal{J}(x_2) = (p_2, f)$. Consequently $d_1 = d$ and $d_2 = f$ are the elements required to witness Condition 2.

At this point, we have shown that $h_\mathcal{J}$ is tame and now prove that also $h_\mathcal{I}$ is. To show that Condition 1 is satisfied, it suffices to prove that $G_{h_\mathcal{I},\mathcal{I}} = G_{h_\mathcal{J},\mathcal{J}}$. For this, in turn, it is enough to show that $x \sim_{h_\mathcal{I}} y$ if and only if $x \sim_{h_\mathcal{J}} y$ for all $x, y \in \mathsf{var}(q)$. Assume that $x \sim_{h_\mathcal{I}} y$. Then G_q contains a path $x = z_1, \dots, z_n = y$ such that $h_\mathcal{I}(z_1), \dots, h_\mathcal{I}(z_n)$ are all part of the same ensemble in E. Now consider $h_\mathcal{J}(z_1), \dots, h_\mathcal{J}(z_n)$ and recall that q contains an atom $r_i(z_i, z_{i+1})$ for $1 \le i < n$. We have $(h_\mathcal{J}(z_i), h_\mathcal{J}(z_{i+1})) \in r_i^\mathcal{J}$. If $h_\mathcal{J}(z_i) = (p, d)$ and $h_\mathcal{J}(z_{i+1}) = (p', d')$, then the definition of $r_i^\mathcal{J}$ yields $p = p'$: otherwise p' takes the form $pdrf\bar{e}$ and Condition 2 of the definition of paths ensures that $\mathsf{tailE}(p) \neq \bar{e}$. This contradicts the assumption that $h_\mathcal{I}(z_i)$ and $h_\mathcal{I}(z_{i+1})$ are part of the same ensemble in E, since by definition of paths, $d \in \mathsf{tailE}(p)$ and $d' \in \bar{e}$. By definition of $E_\mathcal{J}$, $p = p'$ implies that $h_\mathcal{J}(z_i)$ and $h_\mathcal{J}(z_{i+1})$ are in the same ensemble in $E_\mathcal{J}$ and by transitivity so are all of $h_\mathcal{J}(z_1), \dots, h_\mathcal{J}(z_n)$. Consequently, $x \sim_{h_\mathcal{J}} y$.

The converse direction is similar but simpler, using the definition of $E_\mathcal{J}$.

Now for Condition 2. Take any edge (c_1, c_2) in $G_{h_\mathcal{J},\mathcal{J}}$ and let (p_1, d_1) and (p_2, d_2) be the elements of $\Delta^\mathcal{J}$ which witness that Conditions 2 is satisfied. Then for the edge (c_1, c_2) in $G_{h_\mathcal{I},\mathcal{I}}$ we can take d_1 and d_2 as witnesses. ❑

In the following, we only need one specific unraveling for each mosaic $M \in \mathcal{M}$, depending on the part of the algorithm in which M was guessed:

- if M was guessed in Line 4, we choose $d \in C_0^{\mathcal{I}^M}$ and unravel at (d);
- if M was guessed in Line 7 for some $T \in X_L$, then we choose $d \in (\bigsqcap T)^{\mathcal{I}^M}$ and unravel at (d);
- if M was guessed in Line 13, then we unravel at $\bar{e}^M$.

Note that each $M \in \mathcal{M}$ is guessed at only one of these places: mosaics guessed at Line 4 have level L_0 while the mosaics guessed at Line 7 have a level $L \neq L_0$; moreover, e^M is the empty tuple for mosaics guessed at Lines 4 and 7, but not for mosaics guessed at Line 13. When we speak about the *unraveling of a mosaic* $M \in \mathcal{M}$, we mean the unraveling at the tuple defined above.

We use the unraveled mosaics to construct an L_0-model $\mathcal{I}$ of C_0 and $\mathcal{O}$. More precisely, we construct a sequence of A-interpretations $\mathcal{I}^0, \mathcal{I}^1, \mathcal{I}^2, \dots$ in which more and more concept abstractions and refinements from $\mathcal{O}$ are satisfied. The desired model $\mathcal{I}$ is obtained in the limit. Along with the sequence $\mathcal{I}^0, \mathcal{I}^1, \dots$, we define mappings $M_0, M_1, \dots$ and $\omega_0, \omega_1, \dots$ such that M_i associates with every $d \in \Delta^{\mathcal{I}_i}$ a mosaic $M_i(d) \in \mathcal{M}$ and ω_i associates with every $d \in \Delta^{\mathcal{I}_i}$ an element $\omega(d) \in \Delta^{\mathcal{I}^{M(d)}}$. For the sake of readability, we treat the functions from the sequence $M_0, M_1, \dots$ as a single function M with growing domain, and likewise for $\omega_0, \omega_1, \dots$. Additionally we may write $M(\bar{d})$ or $\omega(\bar{d})$ to mean $M(\bar{d}) = M$ iff $M(d) = M$ for all $d \in \bar{d}$ and likewise for ω.

Now for constructing $\mathcal{I}$, we start with defining $\mathcal{I}^0$ by taking, for every $L \in \mathbf{A}_\mathcal{O}$, the interpretation $\mathcal{I}^0_L$ to be the disjoint union of the unravelings of all mosaics $M \in \mathcal{M}$ with $L^M = L$. The refinement function ρ of $\mathcal{I}^0$ is empty. We set $M(d) = M$ if d is an element of the unraveling of M and $\omega(d) = e$ if the element d in the unraveling of M *originated* from the element $e \in \Delta^{\mathcal{I}^M}$, that is, if d has the form (p, e).

In the following we might use $\rho_L^{-1}(\bar{d})$ to denote the $d \in \Delta^\mathcal{I}$ with $\rho_L(d) = \bar{d}$. Note that inversing ρ returns a distinct value or is undefined since we make sure that it is an injective function.

To obtain $\mathcal{I}^{i+1}$, we start from $\mathcal{I}^i$ and apply the following two rules (in all possible ways):

R1 If $d \in A^{\mathcal{I}^i_L}$, $Q^{\mathsf{ref}}_{M(d),L'}(\omega(d)) \neq \emptyset$, and $\rho_{L'}(d)$ is undefined, then, for the mosaic M' that our algorithm guesses for this choice of $M = M(d)$, d, and L' in Line 12, do the following:

1. add a disjoint copy $\mathcal{J}$ of the unraveling of M' to $\mathcal{I}_{L'}$;

2. set $\rho_{L'}(d)$ to be the tuple $\bar{e}$ over $\mathcal{J}$ with $\omega(\bar{e}) = \bar{e}^{M'}$.[6]

For all elements $d \in \Delta^{\mathcal{J}}$, set $M(d) = M'$ and $\omega(d) = e$ if d is a copy of an element f in the unraveling of M' such that f that originated from $e \in \Delta^{\mathcal{I}^{M'}}$.

R2 If $\bar{d} \in \mathcal{I}_L^i$, $T^{\mathsf{abs}}_{M(\bar{d}),L'}(\omega(\bar{d})) \neq \emptyset$, $\omega(\bar{d}) \neq \bar{e}^{M(\bar{d})}$, and $\rho_L^{-1}(\bar{d})$ is undefined, then, for the set of concept names T' that our algorithm chooses for this choice of $M = M(d)$, $\bar{d}$, and L' in Line 17, and for the mosaic M' that our algorithm guesses for the choice of $L = L'$ and $T = T'$ in Line 7, do the following:

1. add a disjoint copy $\mathcal{J}$ of the unraveling of M' to $\mathcal{I}_{L'}$;
2. set $\rho_L(d) = \bar{d}$ with $d \in \Delta^{\mathcal{J}}$ the element such that the unraveling of M' unravels at $\omega(d)$.

For all elements $d \in \Delta^{\mathcal{J}}$, set $M(d) = M'$ and $\omega(d) = e$ if d is a copy of an element f in the unraveling of M' such that f that originated from $e \in \Delta^{\mathcal{I}^{M'}}$.

In the limit of this sequence of interpretations, we obtain the A-interpretation $\mathcal{I}^* = \bigcup_{i \geq 0} \mathcal{I}_i$ where the union $\mathcal{J} = (\mathbf{A}_{\mathcal{O}}, \prec, (\mathcal{J}_L)_{L \in \mathbf{A}_{\mathcal{O}}}, \rho^{\cup})$ of two A-interpretations $\mathcal{I} = (\mathbf{A}_{\mathcal{O}}, \prec, (\mathcal{I}_L)_{L \in \mathbf{A}_{\mathcal{O}}}, \rho)$ and $\mathcal{I}' = ((\mathbf{A}_{\mathcal{O}}, \prec, (\mathcal{I}'_L)_{L \in \mathbf{A}_{\mathcal{O}}}, \rho'))$ is defined as:

$$\Delta^{\mathcal{J}} = \Delta^{\mathcal{I}_L} \cup \Delta^{\mathcal{I}'_L};$$
$$A^{\mathcal{J}} = A^{\mathcal{I}_L} \cup A^{\mathcal{I}'_L};$$
$$R^{\mathcal{J}} = R^{\mathcal{I}_L} \cup R^{\mathcal{I}'_L};$$
$$\rho^{\cup}_L(d) = \bar{e}, \text{ if } \rho_L(d) = \bar{e} \text{ or } \rho'_L(d) = \bar{e}$$

with the added condition that if both $\rho_L(d)$ and $\rho'_L(d)$ are defined, then $\rho_L(d) = \rho'_L(d)$. Our chase satisfies this condition since the rules only apply to elements where ρ was undefined.

Lemma 27. *$\mathcal{I}^*$ is an A-interpretation.*

Proof. We need to show the following:

1. The directed graph $(\mathbf{A}_{\mathcal{I}}, \prec)$ is a tree.
   Clear by definition of $\mathcal{I}^*$.
2. ρ is a partial function. Both R1 and R2 only define ρ for elements for which it was undefined previously.
3. No element in ρ is part of two distinct ensembles.
   First, we argue that if $\bar{e}$ is an L-ensemble in $\mathcal{I}^*$, then $\omega(\bar{e}) \in E^{M(\bar{e})}$. When we achieve this we are done, since overlapping ensembles are of course from the same mosaic M and the definition of mosaics does not allow for overlapping ensembles in E^M.
   Now assume that $\rho_L(d) = \bar{e}$. By the construction of $\mathcal{I}^*$, it was either R1 or R2 that set $\rho_L(d)$. R1 specifies that $\omega(\bar{e}) = \bar{e}^{M'}$ for the mosaic $M' \in \mathcal{M}$ with $M(\bar{e}) = M'$. The definition of mosaics then implies $\bar{e}^{M'} \in E^{M'}$, as required.
   If R2 set $\rho_L(d) = \bar{e}$, then let M be the mosaic with $M(\bar{e}) = M$ and $\bar{f}$ the tuple over $\mathcal{I}^M$ with $\omega(\bar{e}) = \bar{f}$. Since $T^{\mathsf{abs}}_{M,L'}(\omega(\bar{e})) \neq \emptyset$, Line 16 in the algorithm immediately implies $\bar{f} \in E$.

[6]Note that $\bar{e}^{M'}$ is preserved by the unraveling

❏

What remains for proving the soundness is to show that $\mathcal{I}^*$ is indeed the desired model.

Lemma 28. *$\mathcal{I}^*$ is an L_0-model of C_0 and $\mathcal{O}$.*

Proof. All concept inclusions, role inclusions, and range restrictions in $\mathcal{O}$ are satisfied because of Lemma 25 and the fact that we only add disjoint copies of unraveled mosaics in the construction of $\mathcal{I}^*$.

Let $d \in A^{\mathcal{I}^*_L}$ be an element such that there is a concept refinement $L'\colon q(\bar{x})$ refines $L\colon A \in \mathcal{O}$. Then $Q^{\mathsf{ref}}_{M(d),L'}(d) \neq \emptyset$ and thus either R1 or R2 must have defined $\rho_L(d)$ in the construction of $\mathcal{I}^*$. If $\rho_L(d) = \bar{e}$ was set by some R1 (resp. R2) application, then Line 13 (resp. Line 18) in the algorithm implies that $\bar{e}$ is an answer to q.

Let $h : \bar{x} \mapsto \bar{e}$ be a homomorphism from q to $\mathcal{I}_L$ for some concept abstraction $L'{:}A$ abstracts $L{:}q(\bar{x})$ in $\mathcal{O}$. Lemma 26 implies that there is then also a tame homomorphism h' from q to $M(\bar{e})$. Hence $T^{\mathsf{abs}}_{M(\bar{e}),L'}(\omega(\bar{e})) \neq \emptyset$ and now there are two cases.

Case 1: $\omega(\bar{e}) = \bar{e}^{M(\bar{e})}$. By the definition of our algorithm we then know that $M(\bar{e})$ and $\bar{e}$ were guessed in Line 12 to satisfy some refinement statements and hence R1 defined $\rho_L^-(\bar{e})$. Line 13 together with the definition of unravelings then guarantees that $\rho_L^-(\bar{e})$ satisfies A.

Case 2: $\omega(\bar{e}) \neq \bar{e}^{M(\bar{e})}$. This implies R2 defined $\rho_L^-(\bar{e})$. Line 18 together with the definition of unravelings then guarantees that $\rho_L^-(\bar{e})$ satisfies A.

What remains to be shown is that $C_0^{\mathcal{I}^*_{L_0}} \neq \emptyset$. This is straightforward to see since our algorithm guesses a mosaic M with $C_0^{\mathcal{I}^M} \neq \emptyset$ in Line 4, and the unraveling of that mosaic is of course part of $\mathcal{I}^*$. ❏

C.4 Completeness

Assume that C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$. We prove that our algorithm accepts by showing how to take the non-deterministic choices towards a successful run. Recall that these choices are as follows: the algorithm guesses sets of concept names X_L in Line 3, mosaics in Lines 4, 7, and 12, and a $T' \in X_{L'}$ in Line 17.

Since C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$, by Lemma 4 there is an L_0-pseudo-model $\mathcal{I}$ of C_0 and $\mathcal{O}$. We use $\mathcal{I}$ to guide the non-deterministic choices of the algorithm.

- For Line 3, recall that that $\mathcal{I}$ is constructed out of a universal model $\mathcal{U}$ that satisfies Lemma 20. This lemma provides us a set Y_L of sets of concept names for the abstraction elements of level L. It also gives the upper bound of $|Y_L| \leq ||\mathcal{O}||^3 + ||\mathcal{O}||^2$ and thus we use Y_L as X_L in the algorithm. This bound then also holds for $\mathcal{I}$ by Lemma 22.
- For Line 4, choose an MCC $\mathcal{J}$ of $\mathcal{I}_{L_0}$ with $C_0^{\mathcal{I}_{L_0}} \neq \emptyset$ and such that $\mathcal{J}$ does not contain a refinement ensemble. A straightforward analysis of the construction of $\mathcal{I}$ implies that such an MCC exists. By Lemma 24, any MCC we choose contains at most $2 \cdot (||\mathcal{O}||^2 + ||\mathcal{O}||)$ elements and hence it is straightforward to convert $\mathcal{J}$ to a mosaic M:

Recall that Δ is the domain of mosaics. Let $\mathcal{J}'$ be an isomorphic copy of $\mathcal{J}$ such that $\Delta^{\mathcal{J}'} \subseteq \Delta$ by some isomorphism $\iota\colon \Delta^{\mathcal{J}} \to \Delta^{\mathcal{J}'}$. We define M as follows: $M = (\mathcal{J}', L_0, \{\iota(\bar{e}) \mid \bar{e} \text{ is } L\text{-ensemble in } \mathcal{I}|_{\Delta^{\mathcal{J}}}\}, ())$;

- For Line 7, let $L \in \mathbf{A}_{\mathcal{O}}$ and $T \in X_L$ be the abstraction level and set of elements chosen in Line 6. We choose an MCC $\mathcal{J}$ of $\mathcal{I}_L$ with $(\bigsqcap T)^{\mathcal{I}_L} \neq \emptyset$ which exists by definition of X_L. Let $\iota\colon \Delta^{\mathcal{J}} \to \Delta^{\mathcal{J}'}$ be an isomorphism to an isomorphic copy $\mathcal{J}'$ of $\mathcal{J}$, as before. We then convert $\mathcal{J}$ to a mosaic $M = (\mathcal{J}', L, \{\iota(\bar{e}) \mid \bar{e} \text{ is } L\text{-ensemble in } \mathcal{I}|_{\Delta^{\mathcal{J}}}\}, ())$;
- For Line 12, let M, $d \in \Delta^{\mathcal{I}^M}$, and $L' \in \mathbf{A}_{\mathcal{O}}$ be the mosaic, element, and abstraction level chosen in Lines 10 and 11. Let ι be the isomorphism that was used to construct $\mathcal{I}^M$ from some MCC of $\mathcal{I}$.

  Line 11 implies that $Q^{\mathsf{ref}}_{M,L'}(d) \neq \emptyset$ and hence $\rho^{\mathcal{I}}_{L'}(\iota^{-1}(d)) = \bar{e}$ is defined since $\mathcal{I}$ is a pseudo-model of $\mathcal{O}$. Let $\mathcal{J}$ be the MCC containing $\bar{e}$ and $\mathcal{J}'$ the isomorphic copy of $\mathcal{J}$ by some isomorphism ι'. Note that $\bar{e}$ is the one and only refinement ensemble in $\mathcal{J}$. We then convert this to a mosaic $M = (\mathcal{J}', L', \{\iota'(\bar{f}) \mid \bar{f} \text{ is } L\text{-ensemble in } \mathcal{I}|_{\Delta^{\mathcal{J}}}\}, \iota'(\bar{e}))$;
- For Line 17, let M, $L'{:}A$ <u>abstracts</u> $L{:}q(\bar{x}) \in \mathcal{O}$, and $\bar{d}$ be the mosaic, concept abstraction, and tame answer to q on $\mathcal{I}$ chosen in Lines 10 and 15. Let ι be the isomorphism that was used to construct $\mathcal{I}^M$ from some MCC of $\mathcal{I}$. Since $\mathcal{I}$ is a pseudo-model of $\mathcal{O}$, $\rho^{\mathcal{I}}_L(d) = \iota^{-1}(\bar{d})$ is defined such that $d \in A^{\mathcal{I}_{L'}}$.

  What remains to be shown is that $\iota^{-1}(\bar{d})$ is an abstraction ensemble in $\mathcal{I}$, since then we can choose a $T \in Y_L$ with $\mathsf{CN}_{\mathcal{I}}(d) = T$, by Lemma 20. This is trivial since our choice of mosaics implies that if $\iota^{-1}(\bar{d})$ were a refinement ensemble, then $\bar{e}^M = \bar{d}$ and the check $\bar{d} \neq \bar{e}^M$ in Line 15 ensures that this cannot be the case.

What remains to be shown is our algorithm does not return false in Lines 13, 16 or 18. Since $\mathcal{I}$ is a pseudo-model, concept refinements and concept abstractions are of course satisfied (w.r.t. to tame answers). Hence our choice of mosaics implies that Line 13 does not return false, since we choose the MCC that satisfies the concept refinement of d and convert it to a mosaic. Line 16 is similar in that every tame answer $\bar{d}$ to a CQ of a concept abstraction must be an abstraction ensemble in $\mathcal{I}$, and hence $\bar{d} \in E^M$ by our choice of mosaics. For Line 18, $\mathcal{I}$ being a pseudo-model implies that there must be an abstraction element d for the abstraction ensemble $\bar{d}$ satisfying all the concept refinement and concept abstraction statements. We chose as T the set of concept names satisfied by d in $\mathcal{I}$ and thus Line 18 does also not return false.

D Proofs for Section 5

In this section we will give more details on the proof of Theorem 3 and Theorem 4. Let us first consider Theorem 3 which we repeat there for the reader's convenience.

Theorem 3. *Satisfiability is*

1. CONP*-hard in* $\mathcal{EL}^{\mathsf{abs}}[\mathrm{cr}]$ *and*
2. PSPACE*-hard in* $\mathcal{EL}^{\mathsf{abs}}[\mathrm{cr}, \mathrm{ca}]$.

In the main part of the paper we presented a reduction from unsatisfiability in propositional logic to satisfiability in $\mathcal{EL}^{\mathsf{abs}}[\mathrm{cr}]$ which resulted in Lemma 6 and proves the first point of the theorem. Let us first present a proof of this lemma.

Lemma 6. *φ is unsatisfiable iff $\top$ is L_0-satisfiable w.r.t. $\mathcal{O}$.*

Proof. "$\Rightarrow$". Assume that φ is unsatisfiable. Recall that φ contains the variables $p_1, \ldots, p_n$. We use words of length $m \leq n$ over the alphabet $\{0,1\}$ to represent valuations for the first m of these variables. For a word $w \in \{0,1\}^m$ and $i \leq m$, we use $w[i]$ to denote the i-th symbol in w. Define an AR-interpretation $\mathcal{I} = (\mathbf{A}_{\mathcal{I}}, \prec, (\mathcal{I}_L)_{L\in\mathbf{A}_{\mathcal{I}}}, \rho)$ of $\mathcal{O}$ with $\mathbf{A}_{\mathcal{I}} = \{L_i \mid 1 \leq i \leq n\}$ and $\prec = \{(L_i, L_{i+1}) \mid 1 \leq i < n\}$, and with $\mathcal{I}_{L_i}$ and ρ_{L_i} defined as follows, for $0 \leq i \leq n$:

$$\begin{aligned}\Delta^{\mathcal{I}_{L_i}} &= \{d^i_w \mid w \in \{0,1\}^i\};\\ P_j^{\mathcal{I}_{L_i}} &= \{d^i_w \mid w[j] = 1\} \text{ for } 1 \leq j \leq i;\\ \overline{P}_j^{\mathcal{I}_{L_i}} &= \{d^i_w \mid w[j] = 0\} \text{ for } 1 \leq j \leq i;\\ \rho_{L_{i+1}}(d^i_w) &= (d^{i+1}_{w0}, d^{i+1}_{w1}) \text{ for all } d^i_w \in \Delta^{\mathcal{I}_{L_i}}.\end{aligned}$$

On abstraction level L_n, we additionally set

$$\begin{aligned}T_\psi^{\mathcal{I}_{L_n}} &= \{d^n_w \mid w \models \psi\} \text{ for all } \psi \in \mathsf{sub}(\varphi);\\ F_\psi^{\mathcal{I}_{L_n}} &= \{d^n_w \mid w \not\models \psi\} \text{ for all } \psi \in \mathsf{sub}(\varphi).\end{aligned}$$

It is straightforward to see that this satisfies the Refinements (1) to (3) in $\mathcal{O}$ and thus $\top$ is L_0 satisfiable w.r.t. $\mathcal{O}$.

"$\Leftarrow$". Assume that $\top$ is L_0 satisfiable w.r.t. $\mathcal{O}$. Take any valuation w for φ. We have to show that $w \not\models \varphi$. Due to Refinements (1) to (3) in $\mathcal{O}$, we find a $d \in \Delta^{\mathcal{I}_{L_n}}$ such that $d \in P_i^{\mathcal{I}_{L_n}}$ if and only if $w[i] = 1$, for $1 \leq i \leq n$. As a consequence of CIs (4) to (6) in $\mathcal{O}$, $d \in T_\psi^{\mathcal{I}_{L_n}}$ if and only if $w \models \psi$, for all $\psi \in \mathsf{sub}(\varphi)$. It now follows from CI (7) in $\mathcal{O}$ that $w \not\models \varphi$, as required. ❏

Next, we give the proof for Point 2 of Theorem 3 which was captured by Lemma 7 in the main part of the paper.

Lemma 7. *φ_0 is valid iff $\top$ is L_0-satisfiable w.r.t. $\mathcal{O}$.*

Proof. "$\Rightarrow$". Assume that φ_0 is valid. We need to construct a model $\mathcal{I}$ of $\mathcal{O}$. We start with an interpretation $\mathcal{I}$ constructed as in the proof of Lemma 6 and extend it as follows. We add the following to $\mathcal{I}$ for $1 \leq i \leq n$:

$$s^{\mathcal{I}_{L_i}} = \{(d^i_{w0}, d^i_{w1}) \mid w \in \{0,1\}^{i-1}\}.$$

It is clear that the Statements (1) to (6) of $\mathcal{O}$ are now satisfied. Statement (7) is not part of $\mathcal{O}$ and for CI (8) we add the following to $\mathcal{I}$:

$$F^{\mathcal{I}_{L_n}} = \{d^n_w \mid w \not\models \varphi\}$$

For the Abstractions (9) and (10) we add the following to $\mathcal{I}$ for each $i \in \{0, \ldots, n-1\}$ if $Q_i = \forall$:

$$F^{\mathcal{I}_{L_i}} = \{d^i_w \mid d^{i+1}_{w0} \in F^{\mathcal{I}_{L_{i+1}}} \text{ or } d^{i+1}_{w1} \in F^{\mathcal{I}_{L_{i+1}}}\}$$

and for $i \in \{0, \ldots, n-1\}$ with $Q_i = \exists$ we add the following:

$$F^{\mathcal{I}_{L_i}} = \{d^i_w \mid d^{i+1}_{w0} \in F^{\mathcal{I}_{L_{i+1}}} \text{ and } d^{i+1}_{w1} \in F^{\mathcal{I}_{L_{i+1}}}\}.$$

It is straightforward to see that $d_\varepsilon \notin F$ and thus $\top$ is L_0-satisfiable w.r.t. $\mathcal{O}$.

"$\Leftarrow$". Assume that $\top$ is L_0-satisfiable w.r.t. $\mathcal{O}$. We want to prove that φ_0 is valid. By assumption, there is a model $\mathcal{I}$ of $\mathcal{O}$. To prove that φ_0 is valid, we construct a boolean circuit. (Papadimitriou 2003) Let $T = (V, E, \ell)$ be this boolean circuit with V the set of nodes, E the set of edges and $\ell \colon V \to \{0, 1\}$ a labeling function that assigns a truth value to each node.

Choose any element $d \in \Delta^{\mathcal{I}_{L_0}}$ as the root. We now obtain a full binary tree of depth 2^n by taking the maximally connected component of d, viewing the refinement function as an edge relation and domain elements as nodes. Assume w.l.o.g. that the names of the elements reflect their position in the tree (the root is ε and has children 0 and 1 and so on). Then our circuit T can be defined as follows:

$$\begin{aligned} V &= \{w \mid w \in \{0,1\}^i\} \text{ for } 0 \leq i \leq n \\ E &= \{(w, w0), (w, w1) \mid w \in \{0,1\}^i\} \text{ for } 0 \leq i < n \\ \ell &= \{(w, 1) \mid w \in V \text{ and } w \notin F^{\mathcal{I}_{L_{|w|}}}\} \cup \\ &\quad\ \{(w, 0) \mid w \in V \text{ and } w \in F^{\mathcal{I}_{L_{|w|}}}\}. \end{aligned}$$

As part of the proof for previous reduction we have shown that there is a input gate w (leaf node in the tree) corresponding to each valuation of φ, and that $w \in T_\varphi^{\mathcal{I}_{L_n}}$ iff $w \models \varphi$. Due to the CI (8) and the Abstractions (9) and (10), the inner gates in T get assigned the correct truth value (corresponding to the quantifiers in φ_0). Finally, the CI (11) implies that $\ell(\varepsilon) = 1$ which proves that φ_0 is valid. ❏

Now we will prove Theorem 4 which we repeat here for the reader's convenience.

Theorem 4. *Satisfiability in $\mathcal{EL}^{\mathsf{abs}}[\mathrm{rr}]$ is* 2EXPTIME*-hard.*

This proof is a slight variation of the one for the 2EXPTIME-lower bound of $\mathcal{ALC}^{\mathsf{abs}}[\mathrm{rr}]$ presented in (Lutz and Schulze 2023). The main difference is that since we are in $\mathcal{EL}$ we do not have disjunction or $\forall$-quantification available. Hence we always have to add 'inverse'-roles for any role we introduce so that we can use them to simulate $\forall$-concepts.

For the definition of ATMs used in the reduction, we refer to (Lutz and Schulze 2023). Only note here that our ATMs have a one-side infinite tape and a dedicated accepting state q_a and rejecting state q_r, no successor configuration if its state is q_a or q_r, and exactly two successor configurations otherwise.

It is well-known that there is an exponentially space-bounded alternating Turing machine (ATM) that decides a 2EXPTIME-complete problem and on any input w makes at most $2^{|w|}$ steps (Chandra *et al.* 1981).

Let $M = (Q, \Sigma, \Gamma, q_0, \Delta)$ be a concrete such ATM with $Q = Q_\exists \uplus Q_\forall \uplus \{q_a, q_r\}$. We may assume w.l.o.g that M never attempts to move left when the head is positioned on the left-most tape cell. Let $w = \sigma_1 \cdots \sigma_n \in \Sigma^*$ be an input for M. We want to construct an $\mathcal{ALC}^{\mathsf{abs}}[\mathrm{rr}]$-ontology $\mathcal{O}$ and choose a concept name S and abstraction level L_1 such that S is L_1-satisfiable w.r.t. $\mathcal{O}$ iff $w \in L(M)$. Apart from S, which indicates the starting configuration, we use the following concept names:

- A_σ, for each $\sigma \in \Gamma$, to represent tape content;
- A_q, for each $q \in Q$, to represent state and head position;
- $B_{q,\sigma,M}$ for $q \in Q, \sigma \in \Gamma, M \in \{L, R\}$, serving to choose a transition;
- $H_\leftarrow, H_\rightarrow$ indicating whether a tape cell is to the right or left of the head.

plus some auxiliary concept names whose purpose shall be obvious. We use the role name t for next tape cell c_1, c_2 for successor configurations, and $\widehat{c}_1, \widehat{c}_2$ and $\widehat{t}$ as inverses of c_1, c_2 and t.

The ontology $\mathcal{O}$ uses the abstraction levels $\mathbf{A} = \{L_1, \ldots, L_n\}$ with $L_{i+1} \prec L_i$ for $i \in \{1, \ldots, n-1\}$. While we are interested in L_1-satisfiability of S, the computation of M is simulated on level L_n. We start with generating an infinite computation tree on level L_1:

$$S \sqsubseteq_{L_1} \exists c_1.N \sqcap \exists c_2.N \quad N \sqsubseteq_{L_1} \exists c_1.N \sqcap \exists c_2.N.$$

In the generated tree, each configuration is represented by a single object. On levels $L_2, \ldots, L_n$, we generate similar trees where, however, configurations are represented by t-paths. The length of these paths doubles with every level and each node on a path is connected via c_1 to the corresponding node in the path that represents the first successor configuration, and likewise for c_2 and the second successor configuration. This is illustrated in Figure 3 where for simplicity we only show a first successor configuration and three abstraction levels.

To generate this kind of structure, we introduce the following role refinements for $0 \leq i < n$ and $j \in \{1, 2\}$:

$$L_{i+1} \colon q_1(\bar{x}, \bar{y}) \ \underline{\mathsf{refines}}\ L_i \colon t(x, y)$$
$$L_{i+1} \colon q_2(\bar{x}, \bar{y}) \ \underline{\mathsf{refines}}\ L_i \colon c_j(x, y)$$

for $\bar{x} = x_1 x_2$, $\bar{y} = y_1 y_2$, and

$$\begin{aligned} q_1(\bar{x}, \bar{y}) = {} & t(x_1, x_2) \wedge t(x_2, y_1) \wedge t(y_1, y_2) \wedge \\ & \widehat{t}(x_2, x_1) \wedge \widehat{t}(y_1, x_2) \wedge \widehat{t}(y_2, y_1) \\ q_2(\bar{x}, \bar{y}) = {} & t(x_1, x_2) \wedge t(y_1, y_2) \wedge c_j(x_1, y_1) \wedge c_j(x_2, y_2) \wedge \\ & \widehat{t}(x_2, x_1) \wedge \widehat{t}(y_2, y_1) \wedge \widehat{c}_j(y_1, x_1) \wedge c_j(y_2, x_2) \end{aligned}$$

These two types of refinement statements are depicted in Figure 2, where we for simplicity only show the refinement for the first successor configuration. Note that whenever we refine to a role t or c_i, we also add $\widehat{t}$ and $\widehat{c}_i$ as the inverse of t and c_i.

To make more precise what we want to achieve, let the *m-computation tree*, for $m > 0$, be the interpretation $\mathcal{I}_m$ with

$$\begin{aligned} \Delta^{\mathcal{I}_m} &= \{c_0, c_1\}^* \cdot \{1, \ldots, m\} \\ t^{\mathcal{I}_m} &= \{(wi, wj) \mid w \in \{c_0, c_1\}^*, 1 \leq i < m, j = i + 1\} \\ c_\ell^{\mathcal{I}_m} &= \{(wj, wc_ij) \mid w \in \{c_0, c_1\}^*, 1 \leq j \leq m, i \in \{0, 1\}\} \end{aligned}$$

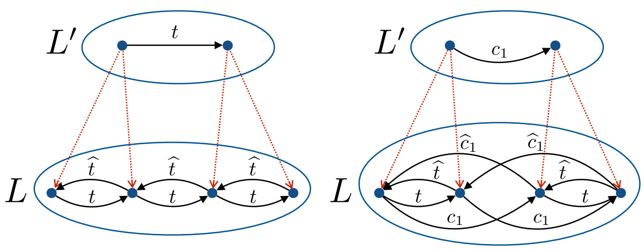

Figure 2: Two types of refinement statements. Dotted lines indicate refinement.

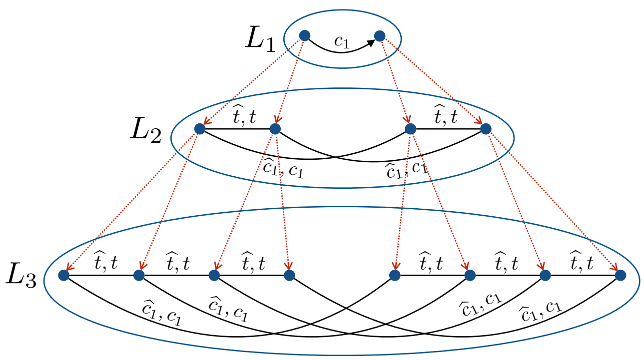

Figure 3: Example interpretation. Edge labels $\widehat{t}, t$ indicate that there is a $\widehat{t}$-edge pointing left and a t-edge pointing right, see Figure 2. Same for $\widehat{c}_1, c_1$-edges.

for $\ell \in \{1, 2\}$. It can be shown that for any model $\mathcal{I}$ of the $\mathcal{ALC}^{\mathsf{abs}}[\mathrm{rr}]$-ontology $\mathcal{O}$ constructed so far and for all $i \in \{1, \dots, n\}$, we must find a (homomorphic image of a) 2^i-computation tree in the interpretation $\mathcal{I}_{L_i}$. This crucially relies on the fact that ensembles cannot overlap. In Figure 3, for example, the role refinements for t and c_1 both apply on level L_2, and for attaining the structure displayed on level L_3 it is crucial that in these applications each object on level L_2 refines into the same ensemble on level L_3.

On level L_n, we thus find a 2^n-computation tree which we use to represent the computation of M on input w. To start, the concept name S is copied down from the root of the 1-computation tree on level L_1 to that of the 2^n-computation tree on level L_n. To achieve this, we add the following role refinement to $\mathcal{O}$ for $0 \leq i < n$, $\bar{x} = x_1 x_2$ and $\bar{y} = y_1 y_2$:

$L_{i+1}\colon q(\bar{x}, \bar{y})$ <u>refines</u> $L_i\colon S(x) \wedge c_1(x, y)$ where
$q(\bar{x}, \bar{y}) = S(x_1) \wedge \top(x_2) \wedge \top(y_1) \wedge \top(y_2)$.

We next describe the initial configuration:

$$
\begin{aligned}
S &\sqsubseteq_{L_n} A_{q_0} \sqcap A_{\sigma_1} \\
\exists \widehat{t}.S &\sqsubseteq_{L_n} A_{\sigma_2} \\
\exists \widehat{t}.A_{\sigma_2} &\sqsubseteq_{L_n} A_{\sigma_3} \\
&\vdots \\
\exists \widehat{t}.A_{\sigma_{n-1}} &\sqsubseteq_{L_n} A_{\sigma_n} \\
\exists \widehat{t}.A_{\sigma_n} &\sqsubseteq_{L_n} A_{\Box} \\
\exists \widehat{t}.A_{\Box} &\sqsubseteq_{L_n} A_{\Box}
\end{aligned}
$$

We add the transitions for successor configurations by first adding marker concepts:

$$
\begin{aligned}
\exists \widehat{c}_1.(A_q \sqcap A_\sigma) &\sqsubseteq_{L_n} B_{q',\sigma',M'} \\
\exists \widehat{c}_2.(A_q \sqcap A_\sigma) &\sqsubseteq_{L_n} B_{q'',\sigma'',M''}
\end{aligned}
$$

for all $q \in Q$ and $\sigma \in \Gamma$ such that $\Delta(q, \sigma) = \{(q', \sigma', M'), (q'', \sigma'', M'')\}$.

Next, we implement the correct configuration by using these marker concepts:

$$
\begin{aligned}
B_{q,\sigma,M} &\sqsubseteq_{L_n} A_\sigma \\
\exists t.B_{q,\sigma,L} &\sqsubseteq_{L_n} A_q \\
\exists \widehat{t}.B_{q,\sigma,R} &\sqsubseteq_{L_n} A_q
\end{aligned}
$$

for all $q \in Q$, $\sigma \in \Gamma$, and $M \in \{L, R\}$. Next, we want to evaluate the existential and universal states of the ATM. First, we mark configurations that are rejecting:

$$A_{q_r} \sqsubseteq C_{\mathsf{rej}}$$

Now we propagate computation results through the computation tree. For existential states, one of the successor states has to be accepting:

$$A_q \sqcap \exists c_1.C_{\mathsf{rej}} \sqcap \exists c_2.C_{\mathsf{rej}} \sqsubseteq_{L_n} C_{\mathsf{rej}}$$

for all $q \in Q_\exists$. For universal states, both successor states have to be accepting:

$$
\begin{aligned}
A_q \sqcap \exists c_1.C_{\mathsf{rej}} &\sqsubseteq_{L_n} C_{\mathsf{rej}} \\
A_q \sqcap \exists c_2.C_{\mathsf{rej}} &\sqsubseteq_{L_n} C_{\mathsf{rej}}
\end{aligned}
$$

for all $q \in Q_\forall$. Lastly, we have to prohibit illegal configurations or changes in the configuration. We mark cells that are not under the head:

$$
\begin{aligned}
\exists t.A_q &\sqsubseteq_{L_n} H_\rightarrow & \qquad \exists \widehat{t}.A_q &\sqsubseteq_{L_n} H_\leftarrow \\
\exists t.H_\rightarrow &\sqsubseteq_{L_n} H_\rightarrow & \qquad \exists \widehat{t}.H_\leftarrow &\sqsubseteq_{L_n} H_\leftarrow
\end{aligned}
$$

for all $q \in Q$. Cells not under the head do not change:

$$
\begin{aligned}
\exists \widehat{c}_i.(H_\rightarrow \sqcap A_\sigma) &\sqsubseteq_{L_n} A_\sigma \\
\exists \widehat{c}_i.(H_\leftarrow \sqcap A_\sigma) &\sqsubseteq_{L_n} A_\sigma
\end{aligned}
$$

for all $\sigma \in \Sigma$. State, content of tape, and head position must be unique:

$$
\begin{aligned}
A_q \sqcap A_{q'} &\sqsubseteq_{L_n} \bot & \qquad A_\sigma \sqcap A_{\sigma'} &\sqsubseteq_{L_n} \bot \\
H_\rightarrow \sqcap A_q &\sqsubseteq_{L_n} \bot & \qquad H_\leftarrow \sqcap A_q &\sqsubseteq_{L_n} \bot
\end{aligned}
$$

for all $q, q' \in Q$ and $\sigma, \sigma' \in \Gamma$ with $q \neq q'$ and $\sigma \neq \sigma'$. Finally, we want an accepting computation:

$$S \sqcap C_{\mathsf{rej}} \sqsubseteq_{L_n} \bot$$

This finishes the construction of $\mathcal{O}$ and it is not hard to verify the following.

Lemma 29. *S is L_1-satisfiable w.r.t. $\mathcal{O}$ iff $w \in L(M)$.*

E Proofs for Section 6

We want to prove Lemma 8 which we repeat here for the reader's convenience.

Lemma 8. *C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$ under set ensemble semantics iff $C_0 \sqcap L_0$ is satisfiable w.r.t. $\mathcal{O}'$.*

Soundness (“$\Rightarrow$”)**.** We prove the two directions of Lemma 8 separately starting with soundness.

Assume that C_0 is L_0-satisfiable w.r.t. $\mathcal{O}$ under set ensemble semantics. We want to show that $C_0 \sqcap L_0$ is then satisfiable w.r.t. $\mathcal{O}'$ (with $\mathcal{O}'$ as defined in the main part of the paper). By our assumption, there is an L_0-model $\mathcal{I} = (\mathbf{A}_\mathcal{I}, \prec, (\mathcal{I}_L)_{L \in \mathbf{A}_\mathcal{I}}, \rho)$ of C_0 and $\mathcal{O}$. We will use it to construct a model $\mathcal{I}'$ of $C_0 \sqcap L_0$ and $\mathcal{O}'$.

W.r.t. $\mathcal{O}$, we call a CQ q an *L-CQ* if q is part of a concept refinement $L{:}q(\bar{x})$ refines $L'{:}A$ in $\mathcal{O}$ or role refinement $L{:}q(\bar{x}, \bar{y})$ refines $L'{:}q_r(x, y)$ in $\mathcal{O}$ that refine to level L. We further call q *active*, if $A^{\mathcal{I}_{L'}} \neq \emptyset$ in the case that q is from the concept refinement or $q_r(\mathcal{I}_{L'}) \neq \emptyset$ if q is from the role refinement.

When constructing a model for $\mathcal{O}'$, by semantics we can ignore nominals that are part of a CI, where the left side is never satisfied. Hence we use $\mathsf{ANom}(\mathcal{O}') = \{a_x \mid x \in \mathsf{var}(q) \text{ and } q \text{ is an active } L\text{-CQ for some } L\}$ to denote the *active nominals*.

We define a function $f\colon \Delta^\mathcal{I} \cup \mathsf{ANom}(\mathcal{O}') \to \Delta^\mathcal{I}$, that maps $\Delta^\mathcal{I}$ and the active nominals in $\mathcal{O}'$ to elements in $\Delta^\mathcal{I}$ as follows:

- $f(d) = d$, for all $d \in \Delta^\mathcal{I}$;
- for all concept refinements $L{:}q(\bar{x})$ refines $L'{:}A$ in $\mathcal{O}$ with q being active, choose an $d \in A^{\mathcal{I}_{L'}}$. Then $\rho_L(d) = \bar{e}$ is defined, since $\mathcal{I}$ is a model of $\mathcal{O}$. We set $f(a_{x_i}) = \bar{e}[i]$ for all $x_i \in \bar{x}$; [7]
- for all role refinements $L{:}q(\bar{x}, \bar{y})$ refines $L'{:}q_r(x, y)$ in $\mathcal{O}$ with q being active, choose $d, d' \in \Delta^{\mathcal{I}_{L'}}$ such that $x \mapsto d$ and $y \mapsto d'$ is an answer to q_r. Then $\rho_L(d) = \bar{e}$ and $\rho_L(d') = \bar{e}'$ are defined, since $\mathcal{I}$ is a model of $\mathcal{O}$. We set $f(a_{x_i}) = \bar{e}[i]$ for all $x_i \in \bar{x}$ and $f(a_{y_i}) = \bar{e}'[i]$ for all $y_i \in \bar{y}$.

Intuitively, nominals represent the variables of a CQ, and we map active nominals to an answer of that CQ in $\mathcal{I}$. Note that f is a function since we assumed that the variables in all CQs have distinct names.

Now we define our model $\mathcal{I}'$ of $\mathcal{O}$.

$$\Delta^{\mathcal{I}'} = \mathsf{ANom}(\mathcal{O}') \cup \Delta^\mathcal{I}$$
$$L^{\mathcal{I}'} = \{d \mid f(d) \in \Delta^{\mathcal{I}_L}\}$$
$$u^{\mathcal{I}'} = \Delta^{\mathcal{I}'} \times \Delta^{\mathcal{I}'}$$
$$A^{\mathcal{I}'} = \{d \mid f(d) \in A^{\mathcal{I}_{L(f(d))}}\}$$
$$r_L^{\mathcal{I}'} = \{(d, e) \mid (f(d), f(e)) \in r^{\mathcal{I}_L}\} \text{ for all } L \in \mathbf{A}_\mathcal{O}$$

Lemma 30. *$\mathcal{I}'$ is a model of $C_0 \sqcap L_0$ and $\mathcal{O}'$.*

Proof. $\mathcal{I}$ being an L_0-model of C_0 and $\mathcal{O}$ implies that there is an element $d_0 \in C_0^{\mathcal{I}_{L_0}}$. The definition of $\mathcal{I}'$ then yields $d_0 \in (C_0 \sqcap L_0)^{\mathcal{I}'}$. Next, we go through the CIs of $\mathcal{O}'$:

- It is straightforward to see that the $\top \sqsubseteq \exists u.L$ is satisfied for any L since u works like a universal role in $\mathcal{I}'$;
- let there be an element $d \in (\exists r_L.A)^{\mathcal{I}'}$ and $L \sqcap \exists r_L.A \sqsubseteq B$ in $\mathcal{O}'$. Semantics imply that there is an $e \in A^{\mathcal{I}'}$ and $(d, e) \in r_L^{\mathcal{I}'}$. The definition of $\mathcal{I}'$ implies that then $(f(d), f(e)) \in r^{\mathcal{I}_L}$ and $f(e) \in A^{\mathcal{I}_L}$ and consequently $f(d) \in (\exists r.A)^{\mathcal{I}_L}$. If $L \sqcap \exists r_L.A \sqsubseteq B$ in $\mathcal{O}'$ then $\exists r.A \sqsubseteq_L B$ in $\mathcal{O}$ by construction of $\mathcal{O}'$. Since $\mathcal{I}$ is a model of $\mathcal{O}$, we hence obtain $f(d) \in B^{\mathcal{I}_L}$. Thus $d \in B^{\mathcal{I}'}$ by definition of $\mathcal{I}'$, as required.
  A similar argument can be made for CIs of the form $L \sqcap A_1 \sqcap \cdots \sqcap A_n \sqsubseteq_L B$ and $L \sqcap A \sqsubseteq_L \exists r_L.(L \sqcap B)$, as well as role inclusions and range restrictions.
- let there be an element $d \in (L' \sqcap A^{\mathcal{I}'})$ for a concept refinement $L{:}q(\bar{x})$ refines $L'{:}A$ in $\mathcal{O}$. Hence $f(d) \in A^{\mathcal{I}_L}$ by definition of $\mathcal{I}'$. Since $\mathcal{I}$ is a model of $\mathcal{O}$, $\rho_L(f(d)) = \bar{e}$ is defined and an answer to q on $\mathcal{I}_L$.
  Assume that there is a concept atom $B(x) \in q$ for some $x \in \bar{x}$. Then for every answer $\bar{e}$ to q on $\mathcal{I}$ by some homomorphism h, we have $h(x) \in B^{\mathcal{I}_L}$. The definition of $f(\cdot)$ then implies that $f(a_x) \in B^{\mathcal{I}_L}$ and thus $a_x \in (L \sqcap B)^{\mathcal{I}'}$, as required for CIs of the form $L' \sqcap A \sqsubseteq \exists u.(L \sqcap B \sqcap \{a_x\})$.
  Recall that $f(\cdot)$ chooses for each active L-CQ q *one* answer to that CQ on $\mathcal{I}$ and then maps the nominals corresponding to the variables of the CQ to that answer. Hence if $r(x, y) \in q$, then $(a_x, a_y) \in r_L^{\mathcal{I}'}$. It is then straightforward to see that CIs of the form $L' \sqcap A \sqsubseteq \exists u.(L \sqcap \{a_x\} \sqcap \exists r_L.(L \sqcap \{a_y\}))$ are satisfied.
- let there be an element $d \in (A_1 \sqcap \exists r_{L'}.A_2)^{\mathcal{I}'}$ for a role refinement $L{:}q(\bar{x}, \bar{y})$ refines $L'{:}q_r(x, y)$ in $\mathcal{O}$ with $q_r = A_1(x) \wedge r(x, y) \wedge A_2(y)$. By semantics, there is an $e \in A_2^{\mathcal{I}'}$ and $(d, e) \in r_{L'}^{\mathcal{I}'}$. Hence $(f(d), f(e)) \in r^{\mathcal{I}_L}$ and $f(e) \in A_2^{\mathcal{I}_L}$, by definition of $\mathcal{I}'$. Now we can argue in the same way as for concept refinements that CIs of the form $A_1 \sqcap \exists r_{L'}.A_2 \sqsubseteq \exists u.(L \sqcap B \sqcap \{a_x\})$ for concept atoms $B(x) \in q$ and CIs $A_1 \sqcap \exists r_{L'}.A_2 \sqsubseteq \exists u.(L \sqcap \{a_x\} \sqcap \exists s_L.(L \sqcap \{a_y\}))$ for role atoms $r(x, y) \in q$ in $\mathcal{O}$ are satisfied.

❑

Completeness (“$\Leftarrow$”)**.** For the other direction assume that $C_0 \sqcap L_0$ is satisfiable w.r.t. $\mathcal{O}'$. We want to show that then C_0 is L_0 satisfiable w.r.t. $\mathcal{O}$. By our assumption, there is a model $\mathcal{I}'$ of $C_0 \sqcap L_0$ and $\mathcal{O}'$. We will use it to construct an L_0-model $\mathcal{I}$ of C_0 and $\mathcal{O}$.

Recall that the nominals in $\mathcal{O}'$ represent an answer to L-CQs of a refinement statement. For example, for a concept refinement $L{:}q(\bar{x})$ refines $L'{:}A$, we have the nominals a_x for all $x \in \bar{x}$. When constructing a model of $\mathcal{O}$, this poses a slight difficulty. If there are multiple elements $d, e \in A^{\mathcal{I}'}$, then we use the same nominals to satisfy

[7] Recall that we always assume a naming scheme of $x_1, x_2, \ldots, x_n$ for the variables in $\bar{x}$

the CIs in $\mathcal{O}'$ regarding concept refinements (for example $L' \sqcap A \sqsubseteq \exists u.(L \sqcap B \sqcap \{a_x\})$) for d and for e.

If we were to naively construct $\mathcal{I}$ from $\mathcal{I}'$ by just copying the elements and setting the refinements in the obvious way, we would thus violate the condition that each element is part of at most one ensemble. Hence we use *unraveling* to define $\mathcal{I}$.

For a CQ q and level $L \in \mathbf{A}_\mathcal{O}$, we use $q[L]$ to denote the CQ obtained from q by replacing every role atom $r(x, y)$ with $r_L(x, y)$ for all role names $r \in \mathbf{R}$. This will be helpful when considering matches of q in models of $\mathcal{O}'$.

A *path* in $\mathcal{I}'$ is a sequence:

$$p = d_1 d_2 \cdots d_k$$

where $k \geq 1$, and for $1 \leq i \leq k$, $d_i \in \Delta^{\mathcal{I}'}$ and for $1 \leq i < k$, d_i and d_{i+1} satisfy one of the following conditions:

1. $(d_i, d_{i+1}) \in r_L^{\mathcal{I}'}$ for some $r \in \mathbf{R}$ and $L \in \mathbf{A}_\mathcal{O}$, or
2. there is a concept refinement $L{:}q(\bar{x})$ refines $L'{:}A$ in $\mathcal{O}$ with $d_i \in A^{\mathcal{I}'}$ and $d_{i+1} = a_x$ with $x \in \bar{x}$, or
3. there is a role refinement $L{:}q(\bar{x}, \bar{y})$ refines $L'{:}q_r(x, y)$ in $\mathcal{O}$ and homomorphism h from $q_r[L']$ to $\mathcal{I}'$ such that one of the following holds:
   - $h(x) = d_i$ and $d_{i+1} = a_x$ with $x \in \bar{x}$ or
   - $h(y) = d_i$ and $d_{i+1} = a_y$ with $y \in \bar{y}$.

For a path p, we use $\mathsf{t}(p)$ to denote the last element (tail) of p. We can now define $\mathcal{I}$.

$$\begin{aligned}
\Delta^{\mathcal{I}_L} &= \{(p, L) \mid p \text{ path in } \mathcal{I}' \text{ and } \mathsf{t}(p) \in L^{\mathcal{I}'}\} \\
A^{\mathcal{I}_L} &= \{(p, L) \mid \mathsf{t}(p) \in A^{\mathcal{I}'}\} \\
r^{\mathcal{I}_L} &= \{((p, L), (p', L)) \mid (\mathsf{t}(p), \mathsf{t}(p')) \in r_L^{\mathcal{I}'}\} \\
\rho_{L'}((p, L)) &= \{(p', L') \mid p' = pd \text{ and} \\
&\qquad \mathsf{t}(p) \text{ and } d \text{ satisfy Condition 2 or 3 of paths}\}
\end{aligned}$$

Lemma 31. *$\mathcal{I}$ is an A-interpretation.*

Proof. We need to show the following:

1. The directed graph $(\mathbf{A}_\mathcal{I}, \prec)$ is a tree. The definition of ρ and paths implies that if $(\mathbf{A}_\mathcal{I}, \prec)$ were not a tree, then $G_\mathcal{O}$ is also not a tree. This is a contradiction, since then $\mathcal{O}'$ would be unsatisfiable as explicitly stated in our reduction.
2. ρ is a partial function. Clear by definition of $\mathcal{I}$.
3. No element in ρ is part of two distinct ensembles.
   Follows from $(\mathbf{A}_\mathcal{I}, \prec^{-1})$ being a tree and the definition of ρ.

❏

Lemma 32. *$\mathcal{I}$ is an L_0-model of C_0 and $\mathcal{O}$.*

Proof. The first set of CIs introduced to $\mathcal{O}'$ ensures that for each $L \in \mathbf{A}_\mathcal{O}$, there is an element $d \in L^{\mathcal{I}'}$, and hence $(d, L) \in \Delta^{\mathcal{I}_L}$ by construction of $\mathcal{I}$, making it non-empty. Additionally, since there is an element $d_0 \in (C_0 \sqcap L_0)^{\mathcal{I}'}$, we have $(d_0, L_0) \in \mathcal{I}_{L_0}$, as required.

Next, let us consider the CIs, role inclusions and range restrictions in $\mathcal{O}$ (recall that $\mathcal{O}$ is normalized).

- let there be an element $(p, L) \in (A_1 \sqcap \cdots \sqcap A_n)^{\mathcal{I}_L}$ and $A_1 \sqcap \cdots \sqcap A_n \sqsubseteq_L B$ in $\mathcal{O}$. The construction of $\mathcal{I}$ implies that $t(p) \in (A_1 \sqcap \cdots \sqcap A_n)^{\mathcal{I}'}$ and $t(p) \in L^{\mathcal{I}'}$. By CI 2 in $\mathcal{O}'$, we then obtain $t(p) \in B^{\mathcal{I}'}$. And then again by construction of $\mathcal{I}$ we have $(p, L) \in B^{\mathcal{I}_L}$, as required;
  This also proves the case for CIs of the form $\top \sqsubseteq_L A$ in $\mathcal{O}$.
- let there be an $(p, L) \in (\exists r.A)^{\mathcal{I}_L}$ and $\exists r.A \sqsubseteq_L B$ in $\mathcal{O}$. Semantics imply that there is an $(p', L) \in A^{\mathcal{I}_L}$ with $((p, L), (p', L)) \in r^{\mathcal{I}_L}$. The construction of $\mathcal{I}$ then implies that $t(p) \in (L \sqcap \exists r_L.A)^{\mathcal{I}'}$. Hence we have $t(p) \in B^{\mathcal{I}'}$ by CI 3 and hence $(p, L) \in B^{\mathcal{I}_L}$, as required;
- let there be an element $(p, L) \in A^{\mathcal{I}_L}$ and $A \sqsubseteq_L \exists r.B$. By construction of $\mathcal{I}$, we have $t(p) \in (L \sqcap A)^{\mathcal{I}'}$. CI 4 then implies that there is an $d \in (L \sqcap B)^{\mathcal{I}'}$ with $(t(p), d) \in r_L^{\mathcal{I}'}$. Hence $(pd, L) \in B^{\mathcal{I}_L}$ and $((p, L), (pd, L)) \in r^{\mathcal{I}_L}$, as required;
- it is straightforward to see that any role inclusion is satisfied by our construction of $\mathcal{I}$ and Point 7 in the construction of $\mathcal{O}'$ and analogously for range restrictions and Point 8 in the construction of $\mathcal{O}'$.

Now, we consider the refinement statements in $\mathcal{O}$:

- let $(p, L') \in A^{\mathcal{I}_{L'}}$ be an element such that there is a concept refinement $L{:}\,q(\bar{x})$ refines $L'{:}\,A$ in $\mathcal{O}$. As always let $\bar{x} = x_1 \cdots x_n$. Then $t(p) \in (L' \sqcap A)^{\mathcal{I}'}$. It is easy to see that then Point 5 in the construction of $\mathcal{O}'$ implies that $a_{x_1} \cdots a_{x_n}$ is an answer to $q[L]$ on $\mathcal{I}'$ and that $a_{x_i} \in L^{\mathcal{I}'}$ for $1 \leq i \leq n$.
  Condition 2 of paths ensures that, for $1 \leq i \leq n$, pa_{x_i} is a path and consequently $(pa_{x_i}, L) \in \Delta^{\mathcal{I}_L}$. The definition of ρ then implies that $\{(pa_{x_1}, L), \ldots, (pa_{x_n}, L)\} \subseteq \rho_L((p, L'))$, as required;
- let $((p_1, L')(p_2, L')) \in q_r(\mathcal{I}_{L'})$ for a role refinement $L{:}q(\bar{x}, \bar{y})$ refines $L'{:}q_r(x, y)$ in $\mathcal{O}$. The proof follows the same structure as with concept refinements.
  Let $\bar{x} = x_1 \cdots x_n$ and $\bar{y} = y_1 \cdots y_m$. The definition of $\mathcal{I}$ implies that then $(t(p_1)t(p_2)) \in q_r[L'](\mathcal{I}')$. It is easy to see that then Point 6 in the construction of $\mathcal{O}'$ implies that $a_{x_1} \cdots a_{x_n} a_{y_1} \cdots a_{y_m}$ is an answer to $q[L]$ on $\mathcal{I}'$ and that $a_{x_i}, a_{y_j} \in L^{\mathcal{I}'}$ for $1 \leq i \leq n$ and $1 \leq j \leq m$.
  By Condition 3 of paths, we obtain that for $1 \leq i \leq n$, $p_1 a_{x_i}$ is a path and for $1 \leq j \leq m$, $p_2 a_{y_j}$ is a path. Consequently $(p_1 a_{x_i}, L) \in \Delta^{\mathcal{I}_L}$ and $(p_2 a_{y_j}, L) \in \Delta^{\mathcal{I}_L}$. The definition of ρ then implies that $\{(p_1 a_{x_1}, L), \ldots, (p_1 a_{x_n}, L)\} \subseteq \rho_L((p_1, L'))$ and $\{(p_2 a_{y_1}, L), \ldots, (p_2 a_{y_m}, L)\} \subseteq \rho_L((p_2, L'))$, as required.

❏

This finishes the completeness direction and thus we are done proving Lemma 8.