Artificial Intelligence 326 (2024) 104034

Contents lists available at ScienceDirect

Artificial Intelligence

journal homepage: www.elsevier.com/locate/artint

Extending the description logic $\mathcal{EL}$ with threshold concepts induced by concept measures

Franz Baader *, Oliver Fernández Gil *

Institute for Theoretical Computer Science, Technische Universität Dresden, 01062 Dresden, Germany

A R T I C L E   I N F O

Keywords:
Description logic
Approximate concept definitions
Concept measures
Reasoning
Computational complexity

A B S T R A C T

In applications of AI systems where exact definitions of the important notions of the application domain are hard to come by, the use of traditional logic-based knowledge representation languages such as Description Logics may lead to very large and unintuitive definitions, and high complexity of reasoning. To overcome this problem, we define new concept constructors that allow us to define concepts in an approximate way. To be more precise, we present a family $\tau\mathcal{EL}(m)$ of extensions of the lightweight Description Logic $\mathcal{EL}$ that use threshold constructors for this purpose. To define the semantics of these constructors we employ graded membership functions m, which for each individual in an interpretation and concept yield a number in the interval $[0, 1]$ expressing the degree to which the individual belongs to the concept in the interpretation. Threshold concepts $C_{\bowtie t}$ for $\bowtie \in \{<, \leq, >, \geq\}$ then collect all the individuals that belong to C with degree $\bowtie t$. The logic $\tau\mathcal{EL}(m)$ extends $\mathcal{EL}$ with threshold concepts whose semantics is defined relative to a function m. To construct appropriate graded membership functions, we show how concept measures $\sim$ (which are graded generalizations of subsumption or equivalence between concepts) can be used to define graded membership functions $m_\sim$. Then we introduce a large class of concept measures, called *simi-d*, for which the logics $\tau\mathcal{EL}(m_\sim)$ have good algorithmic properties. Basically, we show that reasoning in $\tau\mathcal{EL}(m_\sim)$ is NP/coNP-complete without TBox, PSpace-complete w.r.t. acyclic TBoxes, and ExpTime-complete w.r.t. general TBoxes. The exception is the instance problem, which is already PSpace-complete without TBox w.r.t. combined complexity. While the upper bounds hold for all elements of *simi-d*, we could prove some of the hardness results only for a subclass of *simi-d*. This article considerably improves on and generalizes results we have shown in three previous conference papers and it provides detailed proofs of all our results.

1. Introduction

Truly intelligent behaviour usually relies on the availability of relevant knowledge and the ability to reason about this knowledge. For this reason, Knowledge Representation and Reasoning always has been and still is an important subfield of Artificial Intelligence, which introduces and investigates tailor-made representation formalisms with good algorithmic properties of their reasoning procedures. This paper defines a new family of such formalisms, which is motivated by the fact that, in some applications, it would be more intuitive and convenient to define certain notions of the domain in an approximate way. For example, in clinical diagnosis,

* Corresponding authors.
E-mail addresses: franz.baader@tu-dresden.de (F. Baader), oliver.fernandez@tu-dresden.de (O. Fernández Gil).

https://doi.org/10.1016/j.artint.2023.104034
Received 8 December 2021; Received in revised form 30 June 2023; Accepted 14 October 2023

Available online 20 October 2023
0004-3702/

diseases are often linked to a long list of medical signs and symptoms, but patients that have a certain disease rarely show all of them. Instead, one looks for the occurrence of sufficiently many of these signs and symptoms. Similarly, in match-making, people looking for a flat to rent, a bicycle to buy, or a movie to watch may have a long list of desired properties, but will also be satisfied if many, but not all, of them are met. Classical logic-based knowledge representation formalisms would need to resort to large disjunctions to express such conditions, which are not only inconvenient to write and comprehend, but also hard to reason about. In a nutshell, this paper introduces novel representation formalisms extending classical Description Logics that can describe such concepts in a compact and easy to comprehend way and have better reasoning complexity than classical formalisms using large disjunctions. We now give a more detailed introduction into the motivations for and contributions of this paper, and then put these contributions into the context of our previous work on this topic.

Description Logics (DLs) [1,2] are a well-investigated family of logic-based knowledge representation languages, which are frequently used to formalize ontologies for application domains such as biology and medicine [3]. To define the important notions of such an application domain as formal concepts, DLs state necessary and sufficient conditions for an individual to belong to a concept. These conditions can be atomic properties required for the individual (expressed by concept names) or properties that refer to relationships with other individuals and their properties (expressed as role restrictions). The expressivity of a particular DL is determined on the one hand by what sort of properties can be required and how they can be combined. On the other hand, DLs provide their users with ways of stating terminological axioms in a so-called TBox. The simplest kind of TBoxes are called acyclic TBoxes, which consist of concept definitions without cyclic dependencies among the defined concepts. Basically, such TBoxes introduce abbreviations for complex concepts. General TBoxes consist of general concept inclusions (GCI), which can be used to state inclusion constraints between concepts. Data (i.e., information about specific individuals) can be formulated in the ABox, which consists of concept assertions relating individuals to concepts and role assertions relating individuals with each other. Given a knowledge base consisting of a TBox and an ABox, one then wants to infer consequences such as implied subconcept-superconcept relationships between concepts (subsumption problem) or implied element relationships between individuals and concepts (instance problem).

In the DL $\mathcal{EL}$, concepts can be built using concept names as well as the concept constructors conjunction ($C \sqcap D$), existential restriction ($\exists r.C$), and the top concept ($\top$). Though mainly concentrating on $\mathcal{EL}$ in this paper, we will also consider the more expressive DL $\mathcal{ALC}$, which is obtained from $\mathcal{EL}$ by adding the concept constructor negation ($\neg C$), and thus implicitly values restriction ($\forall r.C$) and disjunction ($C \sqcup D$) [4]. The DL $\mathcal{EL}$ has drawn considerable attention in the last two decades since, on the one hand, important inference problems such as subsumption and instance are polynomial in $\mathcal{EL}$, not only w.r.t. acyclic TBoxes [5], but also w.r.t. GCIs [6,7]. On the other hand, though quite inexpressive, $\mathcal{EL}$ underlies the OWL 2 EL profile[1] and can be used to define biomedical ontologies, such as the large medical ontology SNOMED CT,[2] which is basically an acyclic $\mathcal{EL}$ TBox. In $\mathcal{EL}$ we can, for example, formalize the concept of a *good movie* as a movie that is uplifting, has a simple, but original plot, a likable, an evil, and a funny character, action and love scenes, an unobtrusive sound track, and a happy ending using the following concept description:

$$\begin{aligned} C_{Movie} \;:=\; & \mathsf{Movie} \sqcap \mathsf{Uplifting} \sqcap \exists \mathsf{plot}.(\mathsf{Simple} \sqcap \mathsf{Original}) \sqcap \exists \mathsf{character}.\mathsf{Likable} \sqcap \exists \mathsf{character}.\mathsf{Evil} \sqcap \\ & \exists \mathsf{character}.\mathsf{Funny} \sqcap \exists \mathsf{scene}.\mathsf{Action} \sqcap \exists \mathsf{scene}.\mathsf{Love} \sqcap \exists \mathsf{sound}.\mathsf{Unobtrusive} \sqcap \exists \mathsf{ending}.\mathsf{Happy}. \end{aligned} \tag{1}$$

Within an acyclic TBox, the concept definition $\mathsf{GM} \doteq C_{Movie}$ can then be used to introduce the abbreviation GM for the large concept description C_{Movie}. The GCI $\mathsf{GM} \sqsubseteq \exists \mathsf{projected\_in}.\mathsf{Film\_Festival}$ states that good movies are shown at film festivals. Finally, the assertions $\mathsf{GM}(\mathsf{P\&P})$ and $\mathsf{character}(\mathsf{P\&P}, \mathsf{EB})$ say that “Pride and Prejudice” is a good movie that has Elizabeth Bennet as a character [8].

Like all traditional DLs, $\mathcal{EL}$ is based on classical first-order logic, and thus its semantics is strict in the sense that all the stated properties need to be satisfied for an individual to belong to a concept. In applications where exact definitions are hard to come by, it would be useful to relax this strict requirement and allow for approximate definitions of concepts, where most, but not all, of the stated properties are required to hold. For example, people looking for a movie to watch may have a long list of desired properties (such as the ones stated in the concept C_{Movie} in (1)), but will also be satisfied if many, but not all, of them are met. As already mentioned, such situations also occur in other match-making applications [9,10] and in clinical diagnosis. In order to support defining concepts in such an approximate way, we introduce DLs extending $\mathcal{EL}$ with threshold concept constructors of the form $C_{\bowtie t}$, where C is an $\mathcal{EL}$ concept, $\bowtie \in \{<, \leq, >, \geq\}$, and t is a *rational* number in $[0, 1]$. The semantics of these new concept constructors is defined using a graded membership function m that, given a (possibly complex) $\mathcal{EL}$ concept C and an individual d of an interpretation $\mathcal{I}$, returns a value from the interval [0,1], rather than a Boolean value from $\{0, 1\}$. The concept $C_{\bowtie t}$ then collects all the individuals that belong to C with degree $\bowtie t$, where this degree is computed using the function m. In this way we can, for instance, require a good movie to belong to the $\mathcal{EL}$ concept C_{Movie} in (1) with degree at least .8. If we employ the simple graded membership function m_s that returns the percentage of the top-level conjuncts of C satisfied by d (formally introduced in Example 2.12), then $(C_{Movie})_{\geq .8}$ is actually equivalent to a disjunction of 45 $\mathcal{EL}$ concepts, each of which is a conjunction of 8 of the 10 top-level conjuncts of C_{Movie}. Even for a large class of more complicated graded membership functions m we can show (see Lemma 3.11) that threshold concepts of the form $C_{\geq t}$ or $C_{> t}$ are equivalent to disjunctions of $\mathcal{EL}$ concepts, whereas the ones of the form $C_{\leq t}$ or $C_{< t}$ are equivalent to conjunctions of negated $\mathcal{EL}$ concepts. However, the $\mathcal{ALC}$ concepts obtained this way may be very large (in the worst case, non-elementary in the size of C), which means that they are much harder to comprehend than the threshold concepts. In addition, using this translation into $\mathcal{ALC}$ for reasoning purposes does not yield good complexity results, due to the large sizes of these disjunctions and conjunctions as well as the high complexity (ExpTime) of reasoning in $\mathcal{ALC}$.

[1] See http://www.w3.org/TR/owl2-profiles/.

[2] See http://www.ihtsdo.org/snomed-ct/.

The DL $\tau\mathcal{EL}(m)$ is obtained from $\mathcal{EL}$ syntactically by adding the new threshold constructors $C_{\bowtie t}$ and semantically by interpreting these constructors using the graded membership function m. There are, of course, different possibilities for how to define a graded membership function m, and the semantics of the obtained new logic $\tau\mathcal{EL}(m)$ depends on m. Consequently, the complexity of reasoning in $\tau\mathcal{EL}(m)$ may also depend on which function m is used. Instead of investigating this complexity for a single, hand-crafted graded membership function (such as the function m_s from Example 2.12 or the function deg introduced in [11]), our goal is to obtain complexity results for a large class of graded membership functions. Our main idea for constructing such functions is to employ concept measures $\sim$, which generalize equivalence $C \equiv D$ (concept similarity measures) or subsumption $C \sqsupseteq D$ (directional measures) between concepts C, D by returning a value in the interval $[0, 1]$ rather than an element of $\{0, 1\}$. Such measures $\sim$ can be used to induce a graded membership function $m_\sim$ as follows: to determine the degree to which an element d of an interpretation $\mathcal{I}$ belongs to the $\mathcal{EL}$ concept C, $m_\sim$ considers the values $C \sim D$ for all $\mathcal{EL}$ concepts D such that d belongs to D in $\mathcal{I}$, and then returns the maximal value obtained this way. To ensure that this construction, which is inspired by the approach employed in [12] to relax instance queries, yields a well-defined graded membership function $m_\sim$, the concept measure $\sim$ needs to satisfy additional properties, which we collect under the name *standard measures*. In particular, these conditions ensure that the maximum employed in the construction always exists. The simple graded membership function m_s used in our movie example can actually be obtained in this way from a directional measure $\sim_{su}$. For $\mathcal{EL}$ concepts C, D, we know that C subsumes D iff every top-level conjunct of C subsumes D. Basically, the measure $\sim_{su}$ instead counts the top-level conjuncts of C that satisfy this and divides this number by the number of all top-level conjuncts of C (see Example 3.1 for details). The advantage of employing concept measures for defining graded membership functions is that there has been quite some work in the DL community on defining and investigating such measures [13–15,12,16]. In particular, a framework for constructing both directional and similarity measures with well-understood properties, called *simi*, has been introduced in [14]. This framework first defines directional measures, and then uses a fuzzy connector to combine the values obtained by comparing the concepts in both directions to obtain similarity measures.

Just because the graded membership function $m_\sim$ is well-defined does not mean that it is also computable. Intuitively, computability of $m_\sim$ is a prerequisite for reasoning in $\tau\mathcal{EL}(m_\sim)$ to be decidable. We will show that there exist standard concept measures $\sim$ such that $m_\sim$ is not computable and reasoning in $\tau\mathcal{EL}(m_\sim)$ is undecidable. This can, however, only be the case if $\sim$ is not computable. In fact, under appropriate computability assumptions for $\sim$, we will show that reasoning in $\tau\mathcal{EL}(m_\sim)$ is decidable. However, the proof of this result is based on the translation of the concepts of $\tau\mathcal{EL}(m_\sim)$ into the decidable DL $\mathcal{ALC}$ mentioned above, which may cause a non-elementary blow-up in the worst case. Thus, this translation does not yield satisfactory complexity upper bounds. To obtain logics of the form $\tau\mathcal{EL}(m_\sim)$ for which reasoning has lower complexity, we will then introduce a restricted class of concept measures, called *simi-d*, which is based on certain directional instances of the *simi* framework of [14], and show that the members of this class are standard concept measures. We will also explain why we restrict the attention to instances of the directional part of *simi*, though in previous work [17] we have used undirectional concept similarity measures. Basically, we can show that directional measures sometimes yield well-defined graded membership functions when the corresponding undirectional measures do not. And in many cases, threshold DLs obtained from the well-behaved concept similarity measures employed in [17] are equal to or can be simulated by threshold DLs induced by measures in the class *simi-d*.

The main contributions of the paper are results that determine the complexity of reasoning in threshold DLs of the form $\tau\mathcal{EL}(m_\sim)$ for $\sim \in$ *simi-d*. Just as for some classical DLs like $\mathcal{ALC}$ and $\mathcal{FL}_0$ [2,18], but in contrast to what is the case for pure $\mathcal{EL}$, this complexity depends on which kind of TBox formalism is used. Unlike the situation for classical DLs, defining appropriate notions of acyclic and general TBoxes is already a non-trivial task. For reasoning in $\tau\mathcal{EL}(m_\sim)$ w.r.t. acyclic TBoxes, we show that the standard inference problems (concept satisfiability and subsumption, ABox consistency and instance) are PSpace-complete. Without TBox, satisfiability and consistency are NP-complete and subsumption is coNP-complete, but the instance problem remains PSpace-complete, unless one considers data complexity, for which it is coNP-complete both with and without acyclic TBox.[3] For general TBoxes, we show that, for all measures $\sim \in$ *simi-d*, all the considered reasoning problems are ExpTime-complete w.r.t. general $\tau\mathcal{EL}(m_\sim)$ TBoxes. In the particular case of instance checking, the problem remains coNP-complete w.r.t. data complexity.

1.1. Relationship to our previous work

This paper is based on three previous conference publications [11,19,17], but extends the results obtained there considerably. Graded membership functions m and the DLs $\tau\mathcal{EL}(m)$ induced by them were first defined in [11]. In addition to introducing the general family of DLs $\tau\mathcal{EL}(m)$, we have also defined a concrete graded membership function deg there, which is obtained as a natural extension of the well-known homomorphism characterization [20] of crisp membership and subsumption in $\mathcal{EL}$. The paper [11] restricts the attention to the case without TBox, and shows that concept satisfiability and ABox consistency are NP-complete in $\tau\mathcal{EL}(deg)$, whereas the subsumption and the instance checking problem are co-NP complete (the latter w.r.t. data complexity).

These complexity results were restricted to the particular graded membership function deg. Our next goal was to prove such results not only for the single logic $\tau\mathcal{EL}(deg)$, but for a large class of such logics. For this purpose we considered graded membership functions that are defined using concept similarity measures (CSMs) [22,14,15,12,23,24]. Already in [11] it was shown that a CSM $\sim$ that is equivalence invariant, role-depth bounded, and equivalence closed can be used to define a corresponding graded membership function $m_\sim$. In particular, the graded membership function deg can be obtained in this way, i.e., there is a CSM $\sim^*$ such that

[3] For satisfiability, subsumption, and consistency, we could only show the complexity lower bounds for a restricted class of elements of *simi-d*, both for the case without TBox and w.r.t. acyclic TBoxes. The upper bounds hold without such a restriction.

$m_{\sim^*} = deg$. The complexity of reasoning in logics of the form $\tau\mathcal{EL}(m_\sim)$ for CSMs $\sim$ satisfying these three properties was investigated in more detail in [17], where such CSMs were called *standard* CSMs.[4] The decidability and undecidability results for reasoning in $\tau\mathcal{EL}(m_\sim)$ mentioned above were already shown in [17], but w.r.t. the restricted notion of standard measure employed there. In order to obtain threshold DLs $\tau\mathcal{EL}(m_\sim)$ of lower complexity, we identify in [17] a class of standard CSMs (called *simi-mon*) that can be defined using a restricted version of the simi framework introduced in [14]. For DLs $\tau\mathcal{EL}(m_\sim)$ induced by CSMs in *simi-mon*, we were able to show the same complexity upper bounds as for $\tau\mathcal{EL}(deg)$. Matching lower bounds could be proved only for a subclass of these logics.

The complexity results for the logics $\tau\mathcal{EL}(deg)$ and $\tau\mathcal{EL}(m_\sim)$ for $\sim \in$ *simi-mon* were respectively shown in [11] and [17] for the setting without a TBox. In [19], we investigate reasoning in $\tau\mathcal{EL}(deg)$ w.r.t. an acyclic TBox. Surprisingly, this is not as easy as might have been expected. On the one hand, one must be quite careful to define acyclic TBoxes such that they still just introduce abbreviations for complex concepts, and thus can be unfolded. On the other hand, we show in [19] that, in contrast to the case of $\mathcal{EL}$, adding acyclic TBoxes to $\tau\mathcal{EL}(deg)$ increases the complexity of reasoning.

The purpose of the present paper is, *first*, to provide detailed proofs of (generalizations of) the results stated in [11,19,17]. *Second*, we extend both [19] and [17] by considering reasoning in threshold logics $\tau\mathcal{EL}(m_\sim)$ w.r.t. acyclic TBoxes, and showing that the complexity results for $\tau\mathcal{EL}(deg)$ w.r.t. an acyclic TBox of [19] can be transferred to these logics, where $\sim$ belongs to our new class of directional measures *simi-d*, which is closely related to *simi-mon*. Actually, we improve on these results: whereas in [19] we show a PSpace upper bound, but only lower bounds on the second level of the polynomial hierarchy, here we close this gap by also proving a PSpace lower bound. In addition, we also investigate reasoning in logics of the form $\tau\mathcal{EL}(m_\sim)$ w.r.t. general TBoxes, and show that it is ExpTime-complete. *Third*, we generalize the overall framework of threshold logics and make adjustments that are necessitated by these generalizations. The first major generalization is that, in contrast to our previous work [11,19,17], in this paper we do not assume a fixed *finite* signature of concept and role names, but allow for countably infinite sets of such names. This is more in line with how DLs are usually introduced in the community, and it strengthens our complexity upper bounds since they no longer depend on the assumption that the number of concept and role names is a constant. However, it is then no longer clear from the outset that only a finite number of "relevant" concept and role names needs to be considered when evaluating threshold concepts in logics of the form $\tau\mathcal{EL}(m_\sim)$, in particular since we allow for measures $\sim$ that are more general than the ones investigated in [17].[5] This is the main reason why our definition of standard measures differs from the one employed in [17]. For a similar reason, we also need to adjust the definition of acyclic and general TBoxes by assuming that the name space for defined concepts comes from a separate set of concept names of which the graded membership function is agnostic. Regarding the measures $\sim$ used to define graded membership functions, we show that it is more appropriate to use directional measures (which generalize subsumption between concepts) rather than symmetric similarity measures (which generalize equivalence), though in the restricted setting considered in our previous work this did not make a difference.

1.2. Overview

In the next section, we introduce the DLs $\mathcal{EL}$ and $\tau\mathcal{EL}(m)$ and show some useful properties. Then, we extend the notions of general and acyclic TBoxes from $\mathcal{EL}$ to $\tau\mathcal{EL}(m)$, which turns out to be a non-trivial task. In Section 3 we define our new notion of standard concept measures $\sim$, show how they can be used to construct graded membership functions $m_\sim$, and prove the above mentioned undecidability and decidability results for DLs of the form $\tau\mathcal{EL}(m_\sim)$, depending on whether $\sim$ is computable or not. We also introduce a restricted class of concept measures, called *simi-d*, which is based on certain directional instances of *simi*, show that the members of this class are standard concept measures, and explain why we restrict the attention to directional measures. The rest of the paper is concerned with investigating the complexity of reasoning in threshold DLs of the form $\tau\mathcal{EL}(m_\sim)$ for $\sim \in$ *simi-d*. In Section 4 we consider reasoning in $\tau\mathcal{EL}(m_\sim)$ w.r.t. acyclic and empty TBoxes and show the complexity results mentioned above. In Section 5, we concentrate on general TBoxes and prove the ExpTime-completeness results as well as the lower complexity of coNP for data complexity mentioned above. Finally, we describe some related work in Section 6 and then finish the paper with a conclusion in Section 7, which sums up the contributions of the paper and gives ideas for future work.

2. The description logics $\mathcal{EL}$ and $\tau\mathcal{EL}$

We will first introduce $\mathcal{EL}$ and then show how it can be extended by threshold concepts to the DL $\tau\mathcal{EL}$. Afterwards, we will define general and acyclic TBoxes for $\tau\mathcal{EL}$.

2.1. The description logic $\mathcal{EL}$

Let NC, NR and NI be countably infinite sets of *concept*, *role* and *individual* names. Starting with NC, the set $\mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$ of $\mathcal{EL}$ concept descriptions over NC and NR is obtained by using the concept constructors *conjunction* ($C \sqcap D$), *existential restriction* ($\exists r.C$), and *top* ($\top$) in the following way:

[4] In the presented paper, we define standard CSMs in a more general way.

[5] Basically, this is due to the fact that we do not restrict the simi framework to the default primitive measure, for which different concept (role) names have similarity degree 0.

$$C ::= \top \mid A \mid C \sqcap C \mid \exists r.C$$

where $A \in \mathsf{NC}$, $r \in \mathsf{NR}$ and $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$. We use $\mathcal{C}_{\mathcal{EL}}(\mathsf{NC}^f, \mathsf{NR}^f)$ to denote the set of concept descriptions that use only concept and role names from two given finite sets $\mathsf{NC}^f \subseteq \mathsf{NC}$ and $\mathsf{NR}^f \subseteq \mathsf{NR}$.[6] If the sets of concept and role names used to define concept descriptions are clear from the context or irrelevant, we will often write $\mathcal{C}_{\mathcal{EL}}$ in place of $\mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$.

The semantics of $\mathcal{EL}$ is defined using standard first-order interpretations, i.e., relational structures. An *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ of NC and NR consists of a non-empty domain $\Delta^{\mathcal{I}}$ and an interpretation function $\cdot^{\mathcal{I}}$ that assigns subsets $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$ to concept names $A \in \mathsf{NC}$ and binary relations $r^{\mathcal{I}}$ over $\Delta^{\mathcal{I}}$ to role names $r \in \mathsf{NR}$. The function $\cdot^{\mathcal{I}}$ is inductively extended to all $\mathcal{EL}$ concept descriptions in $\mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$ as follows:

$$\top^{\mathcal{I}} := \Delta^{\mathcal{I}}, \quad (C \sqcap D)^{\mathcal{I}} := C^{\mathcal{I}} \cap D^{\mathcal{I}}, \quad (\exists r.C)^{\mathcal{I}} := \{x \in \Delta^{\mathcal{I}} \mid \exists y.((x,y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}})\}.$$

Given two $\mathcal{EL}$ concept descriptions C and D, we say that C is *subsumed* by D (written as $C \sqsubseteq D$) if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for all interpretations $\mathcal{I}$. These two concept descriptions are equivalent (written as $C \equiv D$) if $C \sqsubseteq D$ and $D \sqsubseteq C$. In addition, C is *satisfiable* if $C^{\mathcal{I}} \neq \emptyset$ for some interpretation $\mathcal{I}$.

Information about specific individuals (represented by the set of individual names NI) can be expressed in an ABox, which is a finite set of *assertions* of the form $C(a)$ or $r(a,b)$, where $C \in \mathcal{C}_{\mathcal{EL}}$, $r \in \mathsf{NR}$, and $a, b \in \mathsf{NI}$. In the presence of an ABox, an interpretation $\mathcal{I}$ additionally assigns domain elements $a^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$ to individual names $a \in \mathsf{NI}$. We say that $\mathcal{I}$ *satisfies* an assertion $C(a)$ if $a^{\mathcal{I}} \in C^{\mathcal{I}}$ and $r(a,b)$ if $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$, and that $\mathcal{I}$ is a *model* of the ABox $\mathcal{A}$ (denoted as $\mathcal{I} \models \mathcal{A}$) if it satisfies all the assertions of $\mathcal{A}$. The ABox $\mathcal{A}$ is *consistent* if $\mathcal{I} \models \mathcal{A}$ for some interpretation $\mathcal{I}$. Finally, the individual a is an *instance* of the concept description C in $\mathcal{A}$ (written as $\mathcal{A} \models C(a)$) if $a^{\mathcal{I}} \in C^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathcal{A}$.

An $\mathcal{EL}$ TBox $\mathcal{T}$ is a finite set of concept definitions of the form $E \doteq C_E$, where $E \in \mathsf{NC}$ and $C_E \in \mathcal{C}_{\mathcal{EL}}$. In addition, we require that no concept name occurs more than once on the left-hand side of a definition. Concept names occurring on the left-hand side of a definition in $\mathcal{T}$ are called *defined concepts* of $\mathcal{T}$, whereas the other elements of NC are called *primitive concepts* of $\mathcal{T}$. We denote as $\mathsf{NC}_d^{\mathcal{T}}$ and $\mathsf{NC}_{pr}^{\mathcal{T}}$, respectively, the sets of defined and primitive concepts of $\mathcal{T}$. A *general* $\mathcal{EL}$ TBox is a finite set of *general concept inclusions* (GCIs), which are axioms of the form $C \sqsubseteq D$ with $C, D \in \mathcal{C}_{\mathcal{EL}}$. We say that an interpretation $\mathcal{I}$ satisfies a concept definition $E \doteq C_E$ if $E^{\mathcal{I}} = (C_E)^{\mathcal{I}}$, and a GCI $C \sqsubseteq D$ if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. Note that GCIs generalize the notion of concept definitions, since a definition $E \doteq C_E$ can be equivalently expressed with two GCIs $E \sqsubseteq C_E$ and $C_E \sqsubseteq E$. We say that $\mathcal{I}$ is a model of a general $\mathcal{EL}$ TBox $\mathcal{T}$ (denoted as $\mathcal{I} \models \mathcal{T}$) iff $\mathcal{I}$ satisfies all GCIs $C \sqsubseteq D \in \mathcal{T}$. The relations $\sqsubseteq$ and $\equiv$ are now defined *modulo* the set of models of $\mathcal{T}$, and denoted as $\sqsubseteq_{\mathcal{T}}$ and $\equiv_{\mathcal{T}}$, respectively. The consistency and instance problem can be adapted accordingly to the presence of a TBox, and we can talk about consistency and instance w.r.t. $\mathcal{T}$

$\mathcal{EL}$ TBoxes can be classified into being *acyclic* or *cyclic*, based on how their defined concepts depend on each other. The following definition formalizes this classification.

Definition 2.1. Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox. A defined concept E_1 *directly* depends on a defined concept E_2 iff $E_1 \doteq C_{E_1} \in \mathcal{T}$ and E_2 occurs in C_{E_1}. We denote this *direct dependency* relation as $\rightarrow \subseteq \mathsf{NC}_d^{\mathcal{T}} \times \mathsf{NC}_d^{\mathcal{T}}$, and write it as $E_1 \rightarrow E_2$. Let now $\rightarrow^+$ be the transitive closure of $\rightarrow$. Then, $\mathcal{T}$ is called *cyclic* iff it contains a defined concept E that depends *directly* or *indirectly* on itself, i.e., $E \rightarrow^+ E$. Otherwise, it is called *acyclic*.

The unfolding $u_{\mathcal{T}}(C)$ of an $\mathcal{EL}$ concept description $C \in \mathcal{C}_{\mathcal{EL}}$ w.r.t. $\mathcal{T}$ is the concept description obtained by exhaustively replacing all occurrences of defined concepts E by their definitions C_E in $\mathcal{T}$. More formally,

$$u_{\mathcal{T}}(C) := \begin{cases} C & \text{if } C = \top \text{ or } C \in \mathsf{NC}_{pr}^{\mathcal{T}}, \\ u_{\mathcal{T}}(C_E) & \text{if } C = E \text{ and } E \doteq C_E \in \mathcal{T}, \\ u_{\mathcal{T}}(C_1) \sqcap u_{\mathcal{T}}(C_2) & \text{if } C = C_1 \sqcap C_2, \\ \exists r.u_{\mathcal{T}}(C') & \text{if } C = \exists r.C'. \end{cases} \tag{2}$$

Based on this, the meaning of a concept description C can always be determined from the meaning of its unfolded description, i.e., $C^{\mathcal{I}} = [u_{\mathcal{T}}(C)]^{\mathcal{I}}$ for all models of $\mathcal{I}$ of $\mathcal{T}$, which implies that $C \equiv_{\mathcal{T}} u_{\mathcal{T}}(C)$. From a model-theoretical point of view this is captured by the following proposition (see [25]).

Proposition 2.2. *Let $\mathcal{T}$ be an acyclic $\mathcal{EL}$ TBox. Any interpretation $\mathcal{I}$ of $\mathsf{NC}_{pr}^{\mathcal{T}} \cup \mathsf{NR}$ can be uniquely extended to a model of $\mathcal{T}$.*

We continue with some technical definitions related to $\mathcal{EL}$ concept descriptions, which will be used later on. Let C be an $\mathcal{EL}$ concept description. We denote as $top(C)$ the set of $\mathcal{EL}$ atoms occurring in the top-level conjunction of C, where an $\mathcal{EL}$ *atom* is either a concept name or an existential restriction. In addition, we use $\mathsf{sig}(C)$ to denote the *signature* of C, i.e., the set of all concept and role names occurring in C.

[6] The f in NC^f and NR^f stresses that these sets are finite subsets of NC and NR.

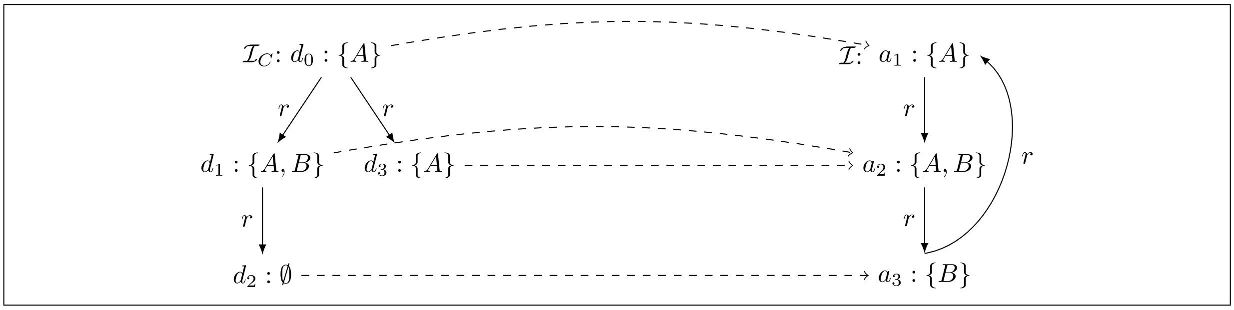


Fig. 1. Graphical representation of interpretations $\mathcal{I}_C$ and $\mathcal{I}$ for Example 2.5.

Example 2.3. Consider the $\mathcal{EL}$ concept description

$$C := (A \sqcap \exists r.((A \sqcap B) \sqcap \exists r.\top)) \sqcap \exists r.A.$$

We have $top(C) = \{A, \exists r.((A \sqcap B) \sqcap \exists r.\top), \exists r.A\}$ and $top(\top) = \emptyset$, as well as $\mathsf{sig}(C) = \{A, B, r\}$ and $\mathsf{sig}(\top) = \emptyset$. $\triangle$

As usual in DL, we will often use the fact that conjunction is associative to dispense with some of the parentheses. The concept description C in our example could thus be written as $A \sqcap \exists r.(A \sqcap B \sqcap \exists r.\top) \sqcap \exists r.A$. When applying inductive definitions based on binary conjunction to such descriptions, we assume without loss of generality that the parentheses have been put left-most, as in the concept C in the example.

The set $\mathsf{sub}(C)$ of *sub-descriptions of* C is defined as

$$\mathsf{sub}(C) := \begin{cases} \{C\} & \text{if } C = \top \text{ or } C \in \mathsf{NC}, \\ \{C\} \cup \mathsf{sub}(C_1) \cup \mathsf{sub}(C_2) & \text{if } C \text{ is of the form } C_1 \sqcap C_2, \\ \{C\} \cup \mathsf{sub}(D) & \text{if } C \text{ is of the form } \exists r.D. \end{cases}$$

In our example, we have $\mathsf{sub}(C) = \{C, A \sqcap \exists r.((A \sqcap B) \sqcap \exists r.\top), \exists r.A, A, \exists r.((A \sqcap B) \sqcap \exists r.\top), (A \sqcap B) \sqcap \exists r.\top, A \sqcap B, \exists r.\top, B, \top\}$. We say that a set S of $\mathcal{EL}$ concept descriptions is *closed under building sub-descriptions* if $C \in S$ implies that $\mathsf{sub}(C) \subseteq S$.

The *role depth* $\mathsf{rd}(C)$ and the *size* $\mathsf{s}(C)$ of an $\mathcal{EL}$ concept description C are inductively defined as

$$\begin{aligned} \mathsf{rd}(\top) = \mathsf{rd}(A) &:= 0 & \mathsf{s}(\top) = \mathsf{s}(A) &:= 1, \\ \mathsf{rd}(C_1 \sqcap C_2) &:= \max(\mathsf{rd}(C_1), \mathsf{rd}(C_2)), & \mathsf{s}(C_1 \sqcap C_2) &:= \mathsf{s}(C_1) + \mathsf{s}(C_2) + 1, \\ \mathsf{rd}(\exists r.C) &:= \mathsf{rd}(C) + 1, & \mathsf{s}(\exists r.C) &:= \mathsf{s}(C) + 1. \end{aligned}$$

In our example, we have $\mathsf{rd}(C) = 2$ and $\mathsf{s}(C) = 12$. Note that the number of sub-descriptions of C is linear in the size of C: $|\mathsf{sub}(C)| \leq \mathsf{s}(C)$.

Next, we introduce homomorphisms between interpretations and describe how they can be used to characterize subsumption and concept membership in $\mathcal{EL}$.

Definition 2.4. Let $\mathcal{I}$ and $\mathcal{J}$ be two interpretations. A *homomorphism* $\phi : \mathcal{I} \to \mathcal{J}$ from $\mathcal{I}$ into $\mathcal{J}$ is a mapping $\phi : \Delta^{\mathcal{I}} \to \Delta^{\mathcal{J}}$ such that

- $d \in A^{\mathcal{I}}$ implies $\phi(d) \in A^{\mathcal{J}}$ for all $d \in \Delta^{\mathcal{I}}$ and $A \in \mathsf{NC}$;
- $(d, e) \in r^{\mathcal{I}}$ implies $(\phi(d), \phi(e)) \in r^{\mathcal{J}}$ for all $d, e \in \Delta^{\mathcal{I}}$ and $r \in \mathsf{NR}$.

To better illustrate this and subsequent notions, we will often view interpretations as graphs.

Example 2.5. Fig. 1 depicts two interpretations $\mathcal{I}$ and $\mathcal{I}_C$, whose domain elements are named nodes, which are labeled by the set of concept names they belong to, and whose edges describe role relationships. For instance, in the case of $\mathcal{I}_C$ we have

$$\Delta^{\mathcal{I}_C} = \{d_0, d_1, d_2, d_3\}, \quad A^{\mathcal{I}_C} = \{d_0, d_1, d_3\}, \quad B^{\mathcal{I}_C} = \{d_1\}, \quad r^{\mathcal{I}_C} = \{(d_0, d_1), (d_0, d_3), (d_1, d_2)\}.$$

It is easy to see that the dashed lines describe a homomorphism ϕ from $\mathcal{I}_C$ into $\mathcal{I}$, i.e., the mapping ϕ defined as $\phi(d_0) = a_1$, $\phi(d_1) = a_2$, $\phi(d_3) = a_2$ and $\phi(d_2) = a_3$ satisfies the two conditions stated in Definition 2.4. $\triangle$

In [20], it was shown that every $\mathcal{EL}$ concept description C can be translated into a corresponding description tree T_C, and vice-versa. Such a description tree can alternatively be seen as a finite *tree-shaped* interpretation, which we denote as $\mathcal{I}_C$. For instance, the $\mathcal{EL}$ concept description C introduced in Example 2.3 is translated into the interpretation $\mathcal{I}_C$ depicted on the left-hand side in Fig. 1. The element d_0 is called the *root* of $\mathcal{I}_C$. Using this representation of $\mathcal{EL}$ concept descriptions, homomorphisms between interpretations can be used to characterize subsumption in $\mathcal{EL}$.

Theorem 2.6 *([20]). Let C and D be $\mathcal{EL}$ concept descriptions. Then, $C \sqsubseteq D$ iff there exists a homomorphism $\phi : \mathcal{I}_D \to \mathcal{I}_C$ such that $\phi(d_0) = c_0$, where d_0 and c_0 are the roots of $\mathcal{I}_D$ and $\mathcal{I}_C$, respectively.*

The proof of this result can be easily adapted to obtain a similar characterization for membership of elements of an interpretation in $\mathcal{EL}$ concepts.

Theorem 2.7. *Let $\mathcal{I}$ be an interpretation, $d \in \Delta^{\mathcal{I}}$ and C an $\mathcal{EL}$ concept description. Then, $d \in C^{\mathcal{I}}$ iff there exists a homomorphism $\phi : \mathcal{I}_C \to \mathcal{I}$ such that $\phi(c_0) = d$, where c_0 is the root of $\mathcal{I}_C$.*

Using closure under composition of homomorphisms, the following property is obtained as a direct consequence of the previous theorem.

Corollary 2.8. *Let $\mathcal{I}, \mathcal{J}$ be interpretations, $d \in \Delta^{\mathcal{I}}$, C an $\mathcal{EL}$ concept description and $\phi : \mathcal{I} \to \mathcal{J}$ a homomorphism. Then, $d \in C^{\mathcal{I}}$ implies $\phi(d) \in C^{\mathcal{J}}$.*

Theorem 2.6 also implies the following relationships between subsumption, role depth and signature:

$$C \sqsubseteq D \text{ implies } \mathsf{sig}(C) \supseteq \mathsf{sig}(D) \text{ and } \mathsf{rd}(C) \geq \mathsf{rd}(D). \tag{3}$$

In particular this shows that equivalent concept descriptions have the same signature and role depth. It is also well-known that, up to equivalence, there are only finitely many $\mathcal{EL}$ concept descriptions of role depth $\leq k$ containing only concept and role names from given finite sets of such names (see [26]). We will now show that a set of representatives of these finitely many equivalence classes can effectively be computed.

For this, we employ the reduced form of $\mathcal{EL}$ concept descriptions introduced by Küsters [27]: every $\mathcal{EL}$ concept description C can be transformed into an equivalent *reduced form* C^r, by removing redundant parts of the description. More precisely, this transformation applies the rewrite rule $C \sqcap D \longrightarrow C$ if $C \sqsubseteq D$ modulo associativity and commutativity of $\sqcap$ as long as possible, not only at the top level conjunction of the concept description, but also under the scope of existential restrictions. Obviously, this rule preserves equivalence, and thus also the role depth. In addition, it can be applied only a polynomial number of times to a given $\mathcal{EL}$ concept descriptions C, and the size of the concept description C^r obtained by an exhaustive application of this rule is bounded by $\mathsf{s}(C)$. We say that an $\mathcal{EL}$ concept description is *reduced* if this rule does not apply to it. Up to associativity and commutativity of $\sqcap$, equivalent $\mathcal{EL}$ concept descriptions have the same reduced form [27].

Lemma 2.9. *Let $\mathsf{NC}^f \subseteq \mathsf{NC}$ and $\mathsf{NR}^f \subseteq \mathsf{NR}$ be finite sets of concept and role names. For all $k \geq 0$ there exists a* finite *and* effectively computable *set $\mathcal{R}^k(\mathsf{NC}^f, \mathsf{NR}^f) \subseteq \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}^f, \mathsf{NR}^f)$ consisting of $\mathcal{EL}$ concept descriptions in reduced form and of role depth $\leq k$ that is complete in the following sense:*

- *for all $D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}^f, \mathsf{NR}^f)$ with $\mathsf{rd}(D) \leq k$ there is some $C \in \mathcal{R}^k(\mathsf{NC}^f, \mathsf{NR}^f)$ such that $D \equiv C$.*

Proof. We prove the lemma by induction over k. Concept descriptions of role depth $k = 0$ are conjunctions of concept names, where the empty conjunction corresponds to $\top$. The requirement of being reduced implies that each concept name occurs at most once in the conjunction. Thus, we define

$$\mathcal{R}^0(\mathsf{NC}^f, \mathsf{NR}^f) := \bigcup_{S \subseteq \mathsf{NC}^f} \{\prod_{A \in S} A\}.$$

When building these conjunctions, we assume that the concept names in the sets S are considered in an arbitrary (but fixed) linear order and that parentheses are put left-most. The set $\mathcal{R}^0(\mathsf{NC}^f, \mathsf{NR}^f)$ is obviously finite and, given the finite set NC^f, can easily be computed. In addition, modulo associativity and commutativity of $\sqcap$, C^r is in $\mathcal{R}^0(\mathsf{NC}^f, \mathsf{NR}^f)$ for all $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}^f, \mathsf{NR}^f)$ with $\mathsf{rd}(C) = 0$.

Up to equivalence, concept descriptions of role depth $\leq k$ for $k > 0$ are of the form

$$A_1 \sqcap \ldots \sqcap A_n \sqcap \exists s_1.D_1 \sqcap \ldots \sqcap \exists s_q.D_q \tag{4}$$

where $n, q \geq 0$, $\{A_1, \ldots, A_n\} \subseteq \mathsf{NC}^f$, and $D_i \in \mathcal{R}^{k-1}(\mathsf{NC}^f, \mathsf{NR}^f)$ for all $1 \leq i \leq q$. The requirement to be reduced imposes the additional constraint that two different conjuncts $\exists r.D_i$ and $\exists r.D_j$ of this conjunction must satisfy $D_i \not\sqsubseteq D_j$. Consequently, the elements of $\mathcal{R}^k(\mathsf{NC}^f, \mathsf{NR}^f)$ can be constructed as follows:

- consider all combinations of subsets $S \subseteq \mathsf{NC}^f$ and for every $r \in \mathsf{NR}^f$ of subsets $S_r \subseteq \mathcal{R}^{k-1}(\mathsf{NC}^f, \mathsf{NR}^f)$;
- test for all $r \in \mathsf{NR}^f$ whether S_r satisfies $C \not\sqsubseteq D$ for all distinct elements C, D of S_r;

- if these tests succeed, add

$$\bigsqcap_{A\in S} A \sqcap \bigsqcap_{r\in \mathsf{NR}^f} \bigsqcap_{C\in S_r} \exists r.C$$

to $\mathcal{R}^k(\mathsf{NC}^f, \mathsf{NR}^f)$. When building the above conjunctions, we again assume that the conjuncts are put in a fixed linear order, and that parentheses are put left-most.

This construction ensures that, modulo associativity and commutativity of $\sqcap$, the reduced form of (4) is in $\mathcal{R}^k(\mathsf{NC}^f, \mathsf{NR}^f)$. In addition, since subsumption in $\mathcal{EL}$ is decidable [20,6] and by induction $\mathcal{R}^{k-1}(\mathsf{NC}^f, \mathsf{NR}^f)$ is finite and computable, this shows that $\mathcal{R}^k(\mathsf{NC}^f, \mathsf{NR}^f)$ is also finite and computable. $\square$

We finish this subsection by defining the *restriction of an $\mathcal{EL}$ concept description to role depth k*:

$$C|_k := C \qquad \text{if } C \in \mathsf{NC} \text{ or } C = \top,$$
$$C|_k := C_1|_k \sqcap C_2|_k \qquad \text{if } C = C_1 \sqcap C_2,$$
$$(\exists r.C)|_k := \begin{cases} \top & \text{if } k=0,\\ \exists r.(C|_{k-1}) & \text{otherwise.}\end{cases}$$

Intuitively, $C|_k$ is obtained from C by removing all existential restrictions occurring at role depth k. For the concept C in Example 2.3 we have $C|_2 = C$ since $\mathsf{rd}(C) = 2$, $C|_1 = (A \sqcap \exists r.((A \sqcap B) \sqcap \top)) \sqcap \exists r.A$, and $C|_0 = (A \sqcap \top) \sqcap \top \equiv A$.

The following lemma is an easy consequence of the above definition and of the homomorphism characterization of subsumption in $\mathcal{EL}$ given in Theorem 2.6.

Lemma 2.10. *Let C, D be $\mathcal{EL}$ concept descriptions and k, ℓ natural numbers. Then $\mathsf{rd}(C|_k) \le k$ and $C \sqsubseteq C|_k$. In addition, $C \sqsubseteq D$ implies $C|_{\mathsf{rd}(D)} \sqsubseteq D$ and $k \le \ell$ implies $D|_k \sqsupseteq D|_\ell$.*

2.2. Adding threshold concepts to $\mathcal{EL}$

Threshold concepts are of the form $C_{\bowtie t}$ where C is an $\mathcal{EL}$ concept description, $\bowtie \in \{<, \le, >, \ge\}$, and t is a rational number in the interval $[0, 1]$. The underlying idea is that elements d of an interpretation $\mathcal{I}$ that are not elements of $C^{\mathcal{I}}$ may still satisfy some of the properties required by C, and we can express the degree of satisfaction of these properties by a number between 0 and 1. For instance, the threshold concept $C_{>.8}$ collects the individuals that belong to C with degree $> .8$. We formalize this intuition by introducing the notion of a graded membership function.

Definition 2.11. A *graded membership function* m is a family of functions that contains for every interpretation $\mathcal{I}$ a function $m^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \mathcal{C}_{\mathcal{EL}} \to [0, 1]$ satisfying the following conditions (for $C, D \in \mathcal{C}_{\mathcal{EL}}$):

M1: for all *interpretations* $\mathcal{I}$ and all $d \in \Delta^{\mathcal{I}}$: $d \in C^{\mathcal{I}}$ iff $m^{\mathcal{I}}(d, C) = 1$,

M2: $C \equiv D$ iff $m^{\mathcal{I}}(d, C) = m^{\mathcal{I}}(d, D)$ for all *interpretations* $\mathcal{I}$ and all $d \in \Delta^{\mathcal{I}}$,

M3: for all *interpretations* $\mathcal{I}, \mathcal{J}$, all *homomorphisms* $\phi : \mathcal{I} \to \mathcal{J}$ and all $d \in \Delta^{\mathcal{I}}$: $m^{\mathcal{I}}(d, C) \le m^{\mathcal{J}}(\phi(d), C)$.

Property *M1* requires that the value 1 is a distinguished value reserved for strict membership in a concept. It expresses the intuition that graded membership generalizes strict membership, which maps into $\{0, 1\}$. Property *M2* requires equivalence invariance. It expresses the intuition that the membership value should not depend on the syntactic form of a concept, but only on its semantics.[7] Property *M3* generalizes the property stated in Corollary 2.8 from crisp membership to graded membership. It says that membership functions should reward for properties (required by the concept) that hold, but not penalize if too many properties (more than required by the concept) are satisfied. This is motivated by the fact that $\mathcal{EL}$ can only ask for properties, but not require that properties are missing.

Example 2.12. A very simple family of functions satisfying *M1* and *M3* is obtained by looking at the top-level conjuncts of C, counting to how many of them d belongs, and then dividing by the number of all top-level conjuncts. More precisely,

$$m_{\mathsf{s}}^{\mathcal{I}}(d, C) = \frac{|\{\mathsf{At} \in top(C) \mid d \in \mathsf{At}^{\mathcal{I}}\}|}{|top(C)|}.$$

[7] Note that the right to left implication in *M2* is already a consequence of *M1*.

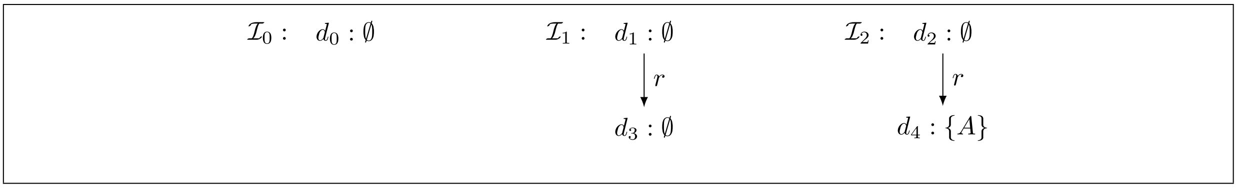


Fig. 2. Three interpretations used in the proof of Proposition 2.14.

However, this family would violate *M2* since C could contain redundant conjuncts. For instance, the concepts $C_1 = A \sqcap B \sqcap A$ and $C_2 = A \sqcap B$ are equivalent, but for an individual d satisfying only B we would have $m_s^{\mathcal{I}}(d, C_1) = 1/3 \neq m_s^{\mathcal{I}}(d, C_2) = 1/2$. This problem can be overcome by considering the top-level conjuncts of C^r rather than of C, i.e., by defining m_s as:

$$m_s^{\mathcal{I}}(d, C) := \frac{|\{\mathsf{At} \in top(C^r) \mid d \in \mathsf{At}^{\mathcal{I}}\}|}{|top(C^r)|}. \quad \triangle$$

There are many families of functions m satisfying *M1 – M3*, and the exact meaning of threshold concepts depends crucially on which of them is used to define their semantics. For this reason, we use m as a parameter in the name of the logic and write $\tau\mathcal{EL}(m)$ rather than just $\tau\mathcal{EL}$. Given the sets NC and NR of concept and role names, the set of $\tau\mathcal{EL}(m)$ *concept descriptions* is obtained as follows:

$$\widehat{C} ::= \top \mid A \mid \widehat{C} \sqcap \widehat{C} \mid \exists r.\widehat{C} \mid C_{\bowtie t}$$

where $A \in \mathsf{NC}$, $r \in \mathsf{NR}$, $\bowtie \in \{<, \leq, >, \geq\}$, $t \in [0,1] \cap \mathbb{Q}$, $C \in \mathcal{C}_{\mathcal{EL}}$ and $\widehat{C}$ is a $\tau\mathcal{EL}(m)$ concept description. Concepts of the form $C_{\bowtie t}$ are called *threshold concepts*. We denote by $\widehat{\mathsf{N}}_{\mathsf{C}}$ the set of all threshold concepts. A $\tau\mathcal{EL}(m)$ ABox generalizes the notion of ABox in $\mathcal{EL}$, by allowing the use of assertions of the form $\widehat{C}(a)$. Extending the notion of (general) TBoxes to $\tau\mathcal{EL}(m)$ is more involved and requires additional care. For this reason, we choose to defer the introduction of their syntax, semantics and corresponding notions to Section 2.3. Syntactic notions like role depth, size, and sub-descriptions can be extended from $\mathcal{EL}$ to $\tau\mathcal{EL}(m)$ as follows: $\mathsf{rd}(C_{\bowtie t}) := \mathsf{rd}(C)$, $\mathsf{sub}(C_{\bowtie t}) := \{C_{\bowtie t}\} \cup \mathsf{sub}(C)$, $\mathsf{s}(C_{\bowtie t}) := \mathsf{s}(C) + 1$. The notion of closure under building sub-descriptions extends to $\tau\mathcal{EL}(m)$ by using the extended definition of sub-descriptions.

The semantics of $\tau\mathcal{EL}(m)$ concept descriptions and ABoxes is defined as for $\mathcal{EL}$, with the addition that threshold concepts are interpreted as

$$(C_{\bowtie t})^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid m^{\mathcal{I}}(d, C) \bowtie t\}.$$

For example, if C is the concept description (1) in the introduction and m_s is the simple graded membership function from Example 2.12, then the movie *Pride and Prejudice* belongs to $C_{\geq 8/9}$ since the only property missing is the action scene.

Requiring property *M1* has the following consequences for the semantics of threshold concepts.

Proposition 2.13. *For every $\mathcal{EL}$ concept description C we have $C_{\geq 1} \equiv C$ and $C_{<1} \equiv \neg C$.*

Here the semantics of negation is defined as usual, i.e., $[\neg C]^{\mathcal{I}} := \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$. The second equivalence basically says that $\tau\mathcal{EL}(m)$ can express negation of $\mathcal{EL}$ concept descriptions. Consequently, unlike $\mathcal{EL}$ concept descriptions, not all $\tau\mathcal{EL}(m)$ concept descriptions are satisfiable (i.e., can be interpreted by a non-empty set). A simple example is the concept description $A_{\geq 1} \sqcap A_{<1}$, which is equivalent to the obviously unsatisfiable concept $A \sqcap \neg A$. This, however, does not imply that $\tau\mathcal{EL}(m)$ is closed under negation since the threshold constructors can only be applied to $\mathcal{EL}$ concept descriptions. Thus, negation cannot be nested using these constructors. In fact, as we will now prove, property *M3* implies that full negation is not expressible in $\tau\mathcal{EL}(m)$.

If we extend $\mathcal{EL}$ with full negation, then we obtain the well-known DL $\mathcal{ALC}$ [4]. Thus, to show that a threshold logic $\tau\mathcal{EL}(m)$ cannot express full negation, it is sufficient to show that there are $\mathcal{ALC}$ concept descriptions that cannot be expressed in $\tau\mathcal{EL}(m)$. Recall that, in $\mathcal{ALC}$, we can express value restrictions $\forall r.C$ with the semantics

$$(\forall r.C)^{\mathcal{I}} := \left\{d \in \Delta^{\mathcal{I}} \mid \forall e \in \Delta^{\mathcal{I}}.(d,e) \in r^{\mathcal{I}} \Rightarrow e \in C^{\mathcal{I}}\right\}$$

since $\forall r.C \equiv \neg\exists r.\neg C$.

Proposition 2.14. *For all graded membership functions m, full negation cannot be expressed in $\tau\mathcal{EL}(m)$.*

Proof. We show that the $\mathcal{ALC}$ concept description $\forall r.A$ for a concept name A cannot be expressed in $\tau\mathcal{EL}(m)$. Assume to the contrary that $\widehat{C}$ is a $\tau\mathcal{EL}(m)$ concept description satisfying $\forall r.A \equiv \widehat{C}$, i.e., $(\forall r.A)^{\mathcal{I}} = \widehat{C}^{\mathcal{I}}$ for all interpretations $\mathcal{I}$. Consider the interpretations $\mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_2$ depicted in Fig. 2. Obviously, $d_0 \in (\forall r.A)^{\mathcal{I}_0}$, and thus $d_0 \in \widehat{C}^{\mathcal{I}_0}$. Since d_0 has no role successors and does not belong to the interpretation of any concept name, this means that $\widehat{C}$ must be a conjunction of threshold concepts, i.e., be of the form $(E^1)_{\bowtie t_1} \sqcap \ldots \sqcap$

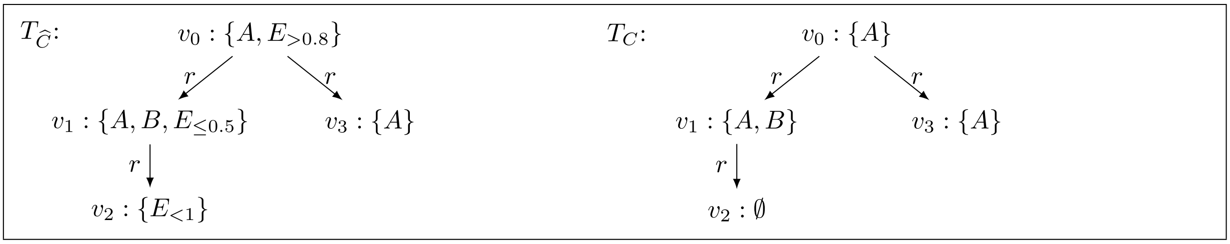


Fig. 3. $\tau\mathcal{EL}(m)$ and $\mathcal{EL}$ description trees for Example 2.16.

$(E^q)_{\bowtie t_q}$. Consequently, $d_0 \in [(E^i)_{\bowtie t_i}]^{\mathcal{I}_0}$ for all $i, 1 \leq i \leq q$. From $d_2 \in (\forall r.A)^{\mathcal{I}_2}$ we can also deduce $d_2 \in \widehat{C}^{\mathcal{I}_2}$, and thus $d_2 \in [(E^i)_{\bowtie t_i}]^{\mathcal{I}_2}$ for all $i, 1 \leq i \leq q$.

Since there is a homomorphism from $\mathcal{I}_0$ into $\mathcal{I}_1$ that maps d_0 to d_1 and a homomorphism from $\mathcal{I}_1$ into $\mathcal{I}_2$ that maps d_1 to d_2, property *M3* yields

$$m^{\mathcal{I}_0}(d_0, E^i) \leq m^{\mathcal{I}_1}(d_1, E^i) \leq m^{\mathcal{I}_2}(d_2, E^i) \text{ for all } i, 1 \leq i \leq q.$$

Together with $d_0 \in [(E^i)_{\bowtie t_i}]^{\mathcal{I}_0}$ and $d_2 \in [(E^i)_{\bowtie t_i}]^{\mathcal{I}_2}$, these inequalities yield $d_1 \in [(E^i)_{\bowtie t_i}]^{\mathcal{I}_1}$ for all $i, 1 \leq i \leq q$. This implies $d_1 \in \widehat{C}^{\mathcal{I}_1}$. However, this contradicts our assumption that $\widehat{C} \equiv \forall r.A$ since obviously $d_1 \notin (\forall r.A)^{\mathcal{I}_1}$. $\square$

Homomorphisms can also be used to characterize membership in $\tau\mathcal{EL}(m)$ concepts, i.e., give necessary and sufficient conditions under which an element in an interpretation belongs to a $\tau\mathcal{EL}(m)$ concept description. This characterization generalizes the one given in Theorem 2.7 for membership in $\mathcal{EL}$ concepts.

Definition 2.15. A $\tau\mathcal{EL}(m)$ *description tree* is a finite tree of the form $T = (V_T, E_T, \ell_T, v_T)$ where:

- V_T is the set of nodes of the tree,
- $E_T \subseteq V_T \times \mathsf{NR} \times V_T$ is the set of edges of the tree, which are labeled by role names,
- $\ell_T : V_T \to 2^{\mathsf{NC} \cup \widehat{\mathsf{N}}_\mathsf{C}}$ is a function that labels nodes with subsets of $\mathsf{NC} \cup \widehat{\mathsf{N}}_\mathsf{C}$, and
- $v_T \in V_T$ is the root of the tree.

Instead of giving a formal definition of the translation of $\tau\mathcal{EL}(m)$ concept descriptions $\widehat{C}$ into $\tau\mathcal{EL}(m)$ description trees $T_{\widehat{C}}$, we illustrate it by an example.

Example 2.16. Let E be an $\mathcal{EL}$ concept description. The $\tau\mathcal{EL}(m)$ concept description

$$\widehat{C} := A \sqcap E_{>0.8} \sqcap \exists r.(A \sqcap B \sqcap E_{\leq 0.5} \sqcap \exists r.E_{<1}) \sqcap \exists r.A$$

corresponds to the $\tau\mathcal{EL}(m)$ description tree $T_{\widehat{C}}$ depicted on the left-hand side of Fig. 3. Furthermore, the right-hand side of the figure depicts the $\mathcal{EL}$ description tree that results from ignoring the threshold concepts in the labels of $T_{\widehat{C}}$, namely, the one corresponding to the $\mathcal{EL}$ concept description $A \sqcap \exists r.(A \sqcap B \sqcap \exists r.\top) \sqcap \exists r.A$. $\triangle$

Let $T = (V_T, E_T, \ell_T, v_T)$ be a $\tau\mathcal{EL}(m)$ description tree and $\mathcal{I}$ an interpretation. The mapping $\phi : V_T \to \Delta^{\mathcal{I}}$ is a τ-*homomorphism* from T to $\mathcal{I}$ if

- $B \in \ell_T(u)$ implies $\phi(u) \in B^{\mathcal{I}}$ for all $u \in V_T$;
- $(u, r, v) \in E_T$ implies $(\phi(u), \phi(v)) \in r^{\mathcal{I}}$.

Note that, in the first condition, B is either a concept name or a threshold concept.

The following characterization of membership in $\tau\mathcal{EL}(m)$ concept descriptions is an easy consequence of the same for $\mathcal{EL}$ (see Theorem 2.7).

Theorem 2.17. *Let $\mathcal{I}$ be an interpretation, $d \in \Delta^{\mathcal{I}}$, and $\widehat{C}$ a $\tau\mathcal{EL}(m)$ concept description. Then, $d \in \widehat{C}^{\mathcal{I}}$ iff there exists a τ-homomorphism ϕ from $T_{\widehat{C}}$ to $\mathcal{I}$ such that $\phi(v_{T_{\widehat{C}}}) = d$.*

Proof. The idea is to treat threshold concepts as atomic concepts and then apply Theorem 2.7. More precisely,

- to each threshold concept $E_{\bowtie t}$ occurring in $\widehat{C}$, we assign a new concept name $E_t^{\bowtie}$. Then, an $\mathcal{EL}$ concept description C is obtained as the result of replacing each $E_{\bowtie t}$ occurring in $\widehat{C}$ by $E_t^{\bowtie}$;

- the interpretation $\mathcal{I}$ is extended into an interpretation $\widehat{\mathcal{I}}$, by defining $(E_t^{\bowtie})^{\widehat{\mathcal{I}}} := (E_{\bowtie t})^{\mathcal{I}}$.

Notice that $T_{\widehat{C}}$ is exactly $\mathcal{I}_C$, if we replace the labels $E_{\bowtie t}$ by $E_t^{\bowtie}$. The following two equivalences are easy to prove.

- $d \in \widehat{C}^{\mathcal{I}}$ iff $d \in C^{\widehat{\mathcal{I}}}$, and
- $\phi : V_T \to \Delta^{\mathcal{I}}$ is a τ-homomorphism from $T_{\widehat{C}}$ to $\mathcal{I}$ with $\phi(v_{T_{\widehat{C}}}) = d$ iff ϕ is a homomorphism from $\mathcal{I}_C$ into $\widehat{\mathcal{I}}$ with $\phi(c_0) = d$, where c_0 is the root of $\mathcal{I}_C$.

Thus, our claim follows by an application of Theorem 2.7. □

If the interpretation $\mathcal{I}$ is finite and $m^{\mathcal{I}}$ is computable, then the existence of a τ-homomorphism is decidable. Algorithm 1 is an extension of the polynomial time algorithm introduced in [20] to decide the existence of a homomorphism between two $\mathcal{EL}$ description trees. Basically, it needs to additionally verify the condition $\phi(u) \in B^{\mathcal{I}}$ in the definition of a τ-homomorphism not only for the case where B is a concept name, but also for the case where B is a threshold concept of the form $E_{\bowtie t}$. This can be done by computing $m^{\mathcal{I}}(d, E)$ and then comparing this number with t. Except for the possible higher complexity of this computation, the algorithm runs in polynomial time. Thus, if $m^{\mathcal{I}}$ is computable in polynomial time, then the overall run time is polynomial. Basically, the algorithm computes for every element $e \in \Delta^{\mathcal{I}}$ the nodes of the tree that can be mapped to e by a homomorphism. The computation is done in a bottom-up manner, i.e., when the algorithm checks to which elements e a node v can be mapped, it has already computed this information for all successors of v.

Algorithm 1 τ-homomorphism from a $\tau\mathcal{EL}(m)$ description tree to a finite interpretation $\mathcal{I}$.

Input: A $\tau\mathcal{EL}(m)$ description tree T, a finite interpretation $\mathcal{I}$, and $d \in \Delta^{\mathcal{I}}$.
Output: "yes", if there exists a τ-homomorphism ϕ from T to $\mathcal{I}$ with $\phi(v_T) = d$; "no", otherwise.

Let $T = (V_T, E_T, \ell_T, v_T)$ and $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$. Assume that $v_1, \ldots, v_n$ is an enumeration of V_T in post-order, i.e., where each node occurs after all its successors
Define a labeling $\delta : \Delta^{\mathcal{I}} \to \mathcal{P}(V_T)$ as follows
Initialize δ by $\delta(e) := \emptyset$ for all $e \in \Delta^{\mathcal{I}}$
for all $1 \leq i \leq n$ **do**
    for all $e \in \Delta^{\mathcal{I}}$ **do**
        if $[B \in \ell_T(v_i) \Rightarrow e \in B^{\mathcal{I}}]$ **and** $[(v_i, r, v) \in E_T \Rightarrow \exists e' \in \Delta^{\mathcal{I}}$ **s.t.** $(e, e') \in r^{\mathcal{I}}$ **and** $v \in \delta(e')]$ **then**
            $\delta(e) := \delta(e) \cup \{v_i\}$
if $v_T \in \delta(d)$ **then** return "yes", **else** return "no" **end if**

Together with Theorem 2.17, this algorithm shows the following decidability and complexity results for membership in $\tau\mathcal{EL}(m)$ concept descriptions. We call a graded membership function *(polynomial time) computable* if $m^{\mathcal{I}}(d, C)$ is computable (in polynomial time in the size of $\mathcal{I}$ and C) for all finite interpretations $\mathcal{I}$ and $\mathcal{EL}$ concepts C.

Theorem 2.18. *Let m be a (polynomial time) computable graded membership function. Then, for a given finite interpretation $\mathcal{I}$, an element $d \in \Delta^{\mathcal{I}}$, and a $\tau\mathcal{EL}(m)$ concept description $\widehat{C}$, it can be decided (in polynomial time) whether $d \in \widehat{C}^{\mathcal{I}}$ or not.*

As an easy consequence of this theorem we obtain also a (polynomial time) procedure that can decide whether a given finite interpretation is a model of an ABox $\mathcal{A}$. In fact, one just needs to check whether $\mathcal{I}$ satisfies all role assertions $r(a, b)$ (by simply looking up whether $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$ holds) and all concept assertions $\widehat{C}(a)$ (by checking whether $a^{\mathcal{I}} \in \widehat{C}^{\mathcal{I}}$ using Algorithm 1).

Corollary 2.19. *Let m be a (polynomial time) computable graded membership function. Then, for a given finite interpretation $\mathcal{I}$ and a $\tau\mathcal{EL}(m)$ ABox $\mathcal{A}$, it can be decided (in polynomial time) whether $\mathcal{I} \models \mathcal{A}$ or not.*

2.3. TBoxes for $\tau\mathcal{EL}(m)$

We will first introduce general $\tau\mathcal{EL}(m)$ TBoxes, and then present the special case of acyclic $\tau\mathcal{EL}(m)$ TBoxes

2.3.1. General $\tau\mathcal{EL}(m)$ TBoxes

For each graded membership function m, the logic $\tau\mathcal{EL}(m)$ is a DL with well-defined syntax and semantics. Thus, we could define general $\tau\mathcal{EL}(m)$ TBoxes as usual as finite sets of $\tau\mathcal{EL}(m)$ GCIs, where a $\tau\mathcal{EL}(m)$ GCI is of the form $\widehat{C} \sqsubseteq \widehat{D}$ for $\tau\mathcal{EL}(m)$ concept descriptions $\widehat{C}$ and $\widehat{D}$. Given the semantics for $\tau\mathcal{EL}(m)$ concept descriptions defined in the previous subsection, the notion of models of such TBoxes can then also be defined in the usual way. We can say that the interpretation $\mathcal{I}$ satisfies the GCI $\widehat{C} \sqsubseteq \widehat{D}$ if $\widehat{C}^{\mathcal{I}} \subseteq \widehat{D}^{\mathcal{I}}$ holds, and it is a model of a general $\tau\mathcal{EL}(m)$ TBox if it satisfies all the GCIs occurring in it. However, though this yields a formally well-defined semantics for TBoxes in $\tau\mathcal{EL}(m)$, it does not always lead to the expected behaviour.

For instance, in our movie example from the introduction, we may want to state that movies satisfying a high number of the properties required by the concept description C_{Movie} in (1) are projected at a film festival, while this is never the case for movies satisfying a low number of these properties, using the GCIs

$$(C_{Movie})_{>.8} \sqsubseteq \exists \mathsf{projected\_in.Film\_Festival} \text{ and } (C_{Movie})_{<.2} \sqsubseteq (\exists \mathsf{projected\_in.Film\_Festival})_{<1}. \tag{5}$$

Since C_{Movie} is a rather large concept description, one may not want to use it directly within the above GCIs, but rather introduce a concept name, say GM, as an abbreviation. We can try to do this by introducing the GCIs

$$\mathsf{GM} \sqsubseteq C_{Movie} \text{ and } C_{Movie} \sqsubseteq \mathsf{GM}, \tag{6}$$

and then replacing the GCIs in (5) with the GCIs

$$\mathsf{GM}_{>.8} \sqsubseteq \exists \mathsf{projected\_in.Film\_Festival} \text{ and } \mathsf{GM}_{<.2} \sqsubseteq (\exists \mathsf{projected\_in.Film\_Festival})_{<1}. \tag{7}$$

Obviously, this makes sense only if, in all models of the GCIs (6), the concept descriptions $(C_{Movie})_{>.8}$ and $\mathsf{GM}_{>.8}$ are interpreted in the same way. Unfortunately, this need not be the case since our graded membership function m, which was introduced when defining the logic $\tau\mathcal{EL}(m)$, is of course agnostic of the TBox. For example, consider the simple graded membership function m_s introduced in Example 2.12, and assume that $\mathcal{I}$ is a model of the GCIs (6) and d an element of this model that satisfies all but one of the top-level conjuncts of C_{Movie}. Then $d \in ((C_{Movie})_{>.8})^{\mathcal{I}}$ since $8/9 > .8$, and $d \notin \mathsf{GM}^{\mathcal{I}}$ since $\mathcal{I}$ is a model of (6) and $d \notin (C_{Movie})^{\mathcal{I}}$. But then we have $d \notin (\mathsf{GM}_{>.8})^{\mathcal{I}}$ since $0 \not> .8$.

To avoid this problem, we allow for a separate acyclic $\mathcal{EL}$ TBox that can be used to introduce abbreviations that can then be employed in the $\tau\mathcal{EL}(m)$ GCIs. The defined concepts of this acyclic TBox are new concept names from a set ND of concept names disjoint with NC and NR. The reason for employing new names is as follows. We assume that the graded membership function m is defined only on the $\mathcal{EL}$ concept descriptions in $\mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR})$. For a given acyclic $\mathcal{EL}$ TBox $\mathcal{T}$ we want to extend m to a function $\widehat{m}$ that can deal with concepts descriptions in $\mathcal{C}_{\mathcal{EL}}(\mathsf{NC} \cup \mathsf{NC}_d^{\mathcal{T}},\mathsf{NR})$ in a way that preserves equivalence w.r.t. $\mathcal{T}$. If we were to use elements of NC as defined concepts in $\mathcal{T}$, this could not be achieved since the value of the membership function on such concepts would already be fixed. This justifies the following more involved definition of general TBoxes for $\tau\mathcal{EL}(m)$.

Definition 2.20. A general $\tau\mathcal{EL}(m)$ TBox is a pair $\mathfrak{T} = (\widehat{\mathcal{T}},\mathcal{T})$ where $\mathcal{T}$ is an acyclic $\mathcal{EL}$ TBox such that:

1. for all $E \doteq C_E \in \mathcal{T}$ we have $E \in \mathsf{ND}$ and $C_E \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC} \cup \mathsf{NC}_d^{\mathcal{T}},\mathsf{NR})$;

and $\widehat{\mathcal{T}}$ is a set of GCIs of the form $\widehat{C} \sqsubseteq \widehat{D}$ satisfying the following conditions:

2. $\widehat{C}$ and $\widehat{D}$ are $\tau\mathcal{EL}(m)$ concept descriptions using concept names from $\mathsf{NC} \cup \mathsf{ND}$ and role names from NR,
3. for all threshold concepts $C_{\bowtie t}$ occurring in $\widehat{C}$ and $\widehat{D}$ we have $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC} \cup \mathsf{NC}_d^{\mathcal{T}},\mathsf{NR})$.

In the above example, we can now assume that $\mathsf{GM} \in \mathsf{ND}$ and the GCIs in (6) (written as a concept definition $\mathsf{GM} \doteq C_{Movie}$) constitute the acyclic $\mathcal{EL}$ TBox $\mathcal{T}$. Extended with the TBox $\widehat{\mathcal{T}}$ consisting of the GCIs in (7), this yields a general $\tau\mathcal{EL}(m)$ TBox.

At the moment, however, the semantics of such TBoxes is not well-defined. In fact, it is not clear how to evaluate threshold concepts of the form $C_{\bowtie t}$ containing concept names $E \in \mathsf{NC}_d^{\mathcal{T}}$ since the graded membership function m is not defined on such concepts. In order to extend m to them, we unfold the names E occurring in C to $u_{\mathcal{T}}(E)$, and then apply m to the unfolded concept, which is an element of $\mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR})$, on which m is defined. Before we can make this idea more formal, we need to introduce some notation.

Definition 2.21. We denote the set of acyclic $\mathcal{EL}$ TBoxes satisfying condition 1. of Definition 2.20 with $\mathfrak{S}_{\mathsf{D}}$. Given $\mathcal{T} \in \mathfrak{S}_{\mathsf{D}}$, we call a $\tau\mathcal{EL}(m)$ concept description $\widehat{D}$ *correctly defined w.r.t.* $\mathcal{T}$ if $\widehat{D}$ satisfies conditions 2. and 3. of Definition 2.20. In the set $\mathfrak{P}_{\mathsf{D}}$ we collect pairs consisting of $\mathcal{EL}$ concept descriptions C and TBoxes $\mathcal{T}$ in $\mathfrak{S}_{\mathsf{D}}$ such that C can only contain elements of ND that are defined concepts in $\mathcal{T}$, i.e.,

$$\mathfrak{P}_{\mathsf{D}} := \{(C,\mathcal{T}) \mid \mathcal{T} \in \mathfrak{S}_{\mathsf{D}} \text{ and } C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC} \cup \mathsf{NC}_d^{\mathcal{T}},\mathsf{NR})\}.$$

The following lemma is an easy consequence of this definition.

Lemma 2.22. *Let $\mathcal{T} \in \mathfrak{S}_{\mathsf{D}}$ and $\widehat{D}$ be a $\tau\mathcal{EL}(m)$ concept description that is correctly defined w.r.t. $\mathcal{T}$. Then $(C,\mathcal{T}) \in \mathfrak{P}_{\mathsf{D}}$ for all threshold concepts $C_{\bowtie t}$ occurring in $\widehat{D}$, and $u_{\mathcal{T}}(C) \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR})$.*

We are now ready to extend a given graded membership function m such that it takes an acyclic $\mathcal{EL}$ TBox in $\mathfrak{S}_{\mathsf{D}}$ into account.

Definition 2.23. For all graded membership functions m (according to Definition 2.11), the extension of m computing membership degrees w.r.t. acyclic $\mathcal{EL}$ TBoxes in $\mathfrak{S}_{\mathsf{D}}$ consists of a family of functions $\widehat{m}$ containing for every interpretation $\mathcal{I}$ a function $\widehat{m}^{\mathcal{I}}$: $\Delta^{\mathcal{I}} \times \mathfrak{P}_{\mathsf{D}}(\mathcal{I}) \to [0,1]$ satisfying

$$\widehat{m}^{\mathcal{I}}(d,C,\mathcal{T}) := m^{\mathcal{I}}(d,u_{\mathcal{T}}(C)),$$

where $\mathfrak{P}_{\mathsf{D}}(\mathcal{I})$ is the set of all pairs $(C,\mathcal{T}) \in \mathfrak{P}_{\mathsf{D}}$ such that $\mathcal{I} \models \mathcal{T}$.

The family of functions $\hat{m}$ is well-defined for the following reasons. Since $\mathcal{T}$ is acyclic and $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC} \cup \mathsf{NC}_d^{\mathcal{T}}, \mathsf{NR})$, the concept description $u_{\mathcal{T}}(C)$ is well-defined and belongs to $\mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$ by Lemma 2.22. Hence, well-definedness of m on concepts in $\mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$ implies well-definedness of $\hat{m}$. Moreover, m is the restriction of $\hat{m}$ to the case $\mathcal{T} = \emptyset$ since then $\mathsf{NC}_d^{\mathcal{T}} = \emptyset$ and $u_{\emptyset}(C) = C$ for all $\mathcal{EL}$ concept descriptions $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$.

The following lemma states that the graded membership functions $\hat{m}$ defined above satisfy a generalization of the conditions stated in Definition 2.11.

Lemma 2.24. *Let m be a graded membership function. Then, for all $\mathcal{EL}$ TBoxes $\mathcal{T}$ in $\mathfrak{S}_{\mathsf{D}}$ and all $\mathcal{EL}$ concept descriptions $C, D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC} \cup \mathsf{NC}_d^{\mathcal{T}}, \mathsf{NR})$ we have:*

$M1^{\mathcal{T}}$: for all models $\mathcal{I}$ of $\mathcal{T}$ and all $d \in \Delta^{\mathcal{I}}$: $d \in C^{\mathcal{I}}$ iff $\hat{m}^{\mathcal{I}}(d, C, \mathcal{T}) = 1$,

$M2^{\mathcal{T}}$: $C \equiv_{\mathcal{T}} D$ iff $\hat{m}^{\mathcal{I}}(d, C, \mathcal{T}) = \hat{m}^{\mathcal{I}}(d, D, \mathcal{T})$ for all models $\mathcal{I}$ of $\mathcal{T}$ and all $d \in \Delta^{\mathcal{I}}$,

$M3^{\mathcal{T}}$: for all models $\mathcal{I}, \mathcal{J}$ of $\mathcal{T}$, all homomorphisms $\phi : \mathcal{I} \to \mathcal{J}$ and all $d \in \Delta^{\mathcal{I}}$: $\hat{m}^{\mathcal{I}}(d, C, \mathcal{T}) \leq \hat{m}^{\mathcal{J}}(\phi(d), C, \mathcal{T})$.

Proof. Since m is a graded membership function in the sense of Definition 2.11, it satisfies the properties *M1*, *M2*, and *M3*. Hence, the facts that $C \equiv_{\mathcal{T}} u_{\mathcal{T}}(C)$, $\mathcal{I} \models \mathcal{T}$, and Definition 2.23 imply that $\hat{m}$ satisfies $M1^{\mathcal{T}}$: $d \in C^{\mathcal{I}} \Leftrightarrow d \in u_{\mathcal{T}}(C)^{\mathcal{I}} \Leftrightarrow m(d, u_{\mathcal{T}}(C)) = 1 \Leftrightarrow \hat{m}^{\mathcal{I}}(d, C, \mathcal{T}) = 1$. Regarding $M2^{\mathcal{T}}$, it is again easy to see that the implication from right to left in $M2^{\mathcal{T}}$ is a consequence of $M1^{\mathcal{T}}$. The implication in the other direction follows from the fact that m satisfies *M2* and that $C \equiv_{\mathcal{T}} D$ iff $u_{\mathcal{T}}(C) \equiv u_{\mathcal{T}}(D)$. Finally, $M3^{\mathcal{T}}$ it is a direct consequence of Definition 2.23 and *M3*. $\square$

We are finally ready to define the semantics of general $\tau\mathcal{EL}(m)$ TBoxes.

Definition 2.25. Let $(\widehat{\mathcal{T}}, \mathcal{T})$ be a general $\tau\mathcal{EL}(m)$ TBox. Given an interpretation $\mathcal{I}$ of $\mathsf{NC} \cup \mathsf{ND}$ and NR that is a model of $\mathcal{T}$, we define the extension of $\mathcal{I}$ to $\tau\mathcal{EL}(m)$ concept descriptions $\widehat{D}$ that are correctly defined w.r.t. $\mathcal{T}$ as for $\mathcal{EL}$ when dealing with the concept constructors top concept, conjunction, and existential restriction, and interpret threshold concepts occurring in $\widehat{D}$ using $\hat{m}$ rather than m, i.e.,

$$(C_{\bowtie t})^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid \hat{m}^{\mathcal{I}}(d, C, \mathcal{T}) \bowtie t\}. \tag{8}$$

The interpretation $\mathcal{I}$ is a *model* of $(\widehat{\mathcal{T}}, \mathcal{T})$ if it is a model of $\mathcal{T}$ and satisfies the GCIs in $\widehat{\mathcal{T}}$, i.e., $\widehat{C}^{\mathcal{I}} \subseteq \widehat{D}^{\mathcal{I}}$ holds for all GCIs $\widehat{C} \sqsubseteq \widehat{D}$ in $\widehat{\mathcal{T}}$, where the threshold concepts in $\widehat{C}$ and $\widehat{D}$ are interpreted using $\hat{m}$, as defined in (8).

Based on this semantics, satisfiability, subsumption, instance, etc. w.r.t. general $\tau\mathcal{EL}(m)$ TBoxes are then defined in the usual way.

2.3.2. Acyclic $\tau\mathcal{EL}(m)$ TBoxes

As with acyclic TBoxes in $\mathcal{EL}$, the purpose of acyclic $\tau\mathcal{EL}(m)$ TBoxes is to introduce abbreviations for composite $\tau\mathcal{EL}(m)$ concept descriptions. In principle, we can introduce such TBoxes as the special case of general $\tau\mathcal{EL}(m)$ TBoxes $(\widehat{\mathcal{T}}, \mathcal{T})$ where $\widehat{\mathcal{T}}$ consists of acyclic concept definitions.

For instance, using the acyclic $\mathcal{EL}$ TBox $\mathcal{T}$ that consists of the concept definitions $E \doteq \exists r.A \sqcap \exists r.B$ and $F \doteq A \sqcap B$, the threshold concepts $(\exists r.A \sqcap \exists r.B)_{\geq 1/2}$ and $(A \sqcap B)_{>0.6}$ can be abbreviated as $E_{\geq 1/2}$ and $F_{>0.6}$, respectively. On top of this, we can also introduce the abbreviations β for $E_{\geq 1/2} \sqcap B$ and γ for $F_{>0.6}$, and can then employ them in other concept definitions, as done in the following pair of TBoxes:

$$\widehat{\mathcal{T}} := \begin{Bmatrix} \alpha \doteq \exists s.\gamma \sqcap \exists r.\beta \\ \beta \doteq E_{\geq 1/2} \sqcap B \\ \gamma \doteq F_{>0.6} \end{Bmatrix} \qquad \mathcal{T} := \begin{Bmatrix} E \doteq \exists r.A \sqcap \exists r.B \\ F \doteq A \sqcap B \end{Bmatrix} \tag{9}$$

Overall, the concept name α is then supposed to abbreviate the $\tau\mathcal{EL}(m)$ concept description

$$\exists s.((A \sqcap B)_{>0.6}) \sqcap \exists r.((\exists r.A \sqcap \exists r.B)_{\geq 1/2} \sqcap B),$$

which is obtained from α by unfolding.

The question is now from which name space the defined concepts in $\widehat{\mathcal{T}}$ (i.e., α, β, γ in our example) should come from. The following example illustrates why these names should again come from ND rather than from NC.

Example 2.26. Consider the general $\tau\mathcal{EL}(m)$ TBox $(\widehat{\mathcal{T}}, \mathcal{T})$ introduced in (9), and assume that the concept name γ belongs to NC. Let $\mathcal{J}$ be a model of $\mathcal{T}$ such that $\Delta^{\mathcal{J}} = \{e\}$, $r^{\mathcal{J}} = s^{\mathcal{J}} = \emptyset$, $e \in A^{\mathcal{J}}$ and $e \notin B^{\mathcal{J}}$. Now, assume that m is a graded membership function such that $m^{\mathcal{J}}(e, A \sqcap B) > 0.6$ if $e \in \gamma^{\mathcal{J}}$, but $m^{\mathcal{J}}(e, A \sqcap B) = 0.5$ if $e \notin \gamma^{\mathcal{J}}$.[8] Hence, there are two different ways to extend $\mathcal{J}$ into a model of

[8] In the next section, we will introduce graded membership functions that can behave like this. Intuitively, the function assumes that γ is similar to B and thus belonging to γ can partially outweigh the missing B.

$(\widehat{\mathcal{T}}, \mathcal{T})$. Namely, by setting $\gamma^{\mathcal{J}} = \beta^{\mathcal{J}} = \alpha^{\mathcal{J}} = \emptyset$ or by setting $\beta^{\mathcal{J}} = \alpha^{\mathcal{J}} = \emptyset$ and $\gamma^{\mathcal{J}} = \{e\}$. This clearly contradicts the intuition underlying concept definitions, which should uniquely determine the interpretation of the defined concept.

Even worse, if $\gamma \doteq F_{>0.6}$ is replaced with $\gamma \doteq F_{<0.6}$, then there is actually no way to extend $\mathcal{J}$ into a model of $\gamma \doteq F_{<0.6}$. In this case, the definition is not ambiguous, but inconsistent, which again should not be the case. △

The problem illustrated by this example goes away if we assume that the defined concepts in $\widehat{\mathcal{T}}$ are elements of ND. In fact, since the graded membership function m is agnostic of such concept names, whether an element of the interpretation domain belongs to such a name or not cannot influence the value assigned by m.

Definition 2.27. An *acyclic $\tau\mathcal{EL}(m)$ TBox* is a general $\tau\mathcal{EL}(m)$ TBox $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$ where $\widehat{\mathcal{T}}$ is an acyclic set of concept definitions of the form $\alpha \doteq \widehat{C}_\alpha$ such that:

- $\alpha \in \mathsf{ND} \setminus \mathsf{NC}_d^{\mathcal{T}}$ and $\widehat{C}_\alpha$ is a $\tau\mathcal{EL}(m)$ concept description satisfying conditions 2 and 3 of Definition 2.20,
- α does not occur as the left-hand side of another concept definition in $\widehat{\mathcal{T}}$.

For example, the pair of TBoxes in (9) form an acyclic $\tau\mathcal{EL}(m)$ TBox $(\widehat{\mathcal{T}}, \mathcal{T})$ if we assume that $\alpha, \beta, \gamma, E, F \in \mathsf{ND}$. Note that the semantics of acyclic $\tau\mathcal{EL}(m)$ TBoxes is given by Definition 2.25 since we have introduced such TBoxes as a special case of general $\tau\mathcal{EL}(m)$ TBoxes.

Given an acyclic $\tau\mathcal{EL}(m)$ TBox $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$, we define the set $\mathsf{NC}_d^{\mathfrak{T}}$ of defined concepts in $\mathfrak{T}$ as $\mathsf{NC}_d^{\widehat{\mathcal{T}}} \cup \mathsf{NC}_d^{\mathcal{T}}$, where $\mathsf{NC}_d^{\widehat{\mathcal{T}}}$ is the set of defined concepts in $\widehat{\mathcal{T}}$. We denote the set $(\mathsf{NC} \cup \mathsf{ND}) \setminus \mathsf{NC}_d^{\mathfrak{T}}$ as $\mathsf{NC}_{pr}^{\mathfrak{T}}$. The unfolding $u_{\mathfrak{T}}(\widehat{D})$ of a $\tau\mathcal{EL}(m)$ concept description $\widehat{D}$ w.r.t. $\mathfrak{T}$ is defined by considering the rules in (2) w.r.t. the primitive concepts in $\mathsf{NC}_{pr}^{\mathfrak{T}}$ and the set of definitions $\widehat{\mathcal{T}} \cup \mathcal{T}$, and introducing the following additional rule:

$$u_{\mathfrak{T}}(\widehat{D}) := [u_{\mathcal{T}}(C)]_{\bowtie t} \ \text{ if } \widehat{D} = C_{\bowtie t}.$$

It is not hard to see that Definition 2.27 and the restrictions imposed on $\mathcal{T}$ in Definition 2.20 guarantee that no defined concept in $\widehat{\mathcal{T}}$ occurs in $\mathcal{T}$. Hence, since both $\widehat{\mathcal{T}}$ and $\mathcal{T}$ are sets of acyclic definitions, unfolding w.r.t. acyclic $\tau\mathcal{EL}(m)$ TBoxes terminates. The following lemma states that unfolding behaves as expected also for acyclic $\tau\mathcal{EL}(m)$ TBoxes.

Lemma 2.28. *Let $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$ be an acyclic $\tau\mathcal{EL}(m)$ TBox and $\widehat{D}$ a $\tau\mathcal{EL}(m)$ concept description that is correctly defined w.r.t. $\mathcal{T}$. Then $u_{\mathfrak{T}}(\widehat{D})$ is a well-defined $\tau\mathcal{EL}(m)$ concept description such that*

- *threshold concepts $C_{\bowtie t}$ occurring in $u_{\mathfrak{T}}(\widehat{D})$ satisfy $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$,*
- *$\widehat{D}^{\mathcal{I}} = u_{\mathfrak{T}}(\widehat{D})^{\mathcal{I}}$ holds for all models $\mathcal{I}$ of $\mathfrak{T}$.*

Using this lemma, we can show the analogon of Proposition 2.2 also for acyclic $\tau\mathcal{EL}(m)$ TBoxes.

Proposition 2.29. *Let $\mathfrak{T}$ be an acyclic $\tau\mathcal{EL}(m)$ TBox. Any interpretation $\mathcal{I}$ of $\mathsf{NC}_{pr}^{\mathfrak{T}} \cup \mathsf{NR}$ can be uniquely extended into a model of $\mathfrak{T}$.*

Proof. The basic idea is to interpret defined concepts E with definition $E \doteq C_E$ in $\mathcal{T}$ as $u_{\mathcal{T}}(C_E)^{\mathcal{I}}$, and defined concepts α with definition $\alpha \doteq \widehat{D}_\alpha$ in $\widehat{\mathcal{T}}$ as $u_{\widehat{\mathcal{T}}}(\widehat{D}_\alpha)^{\mathcal{I}}$. Since E and α do not belong to NC, this does not affect the evaluation of the graded membership functions, and thus of the thresholds concepts. □

3. Concept measures and graded membership functions

In this section, we introduce concept measures, which extend the well-known notion of concept similarity measures [14]. We show how a given concept measure $\sim$ satisfying certain additional properties can be used to define a graded membership function $m_\sim$. First, we describe a mechanism that transforms $\sim$ into a function $m_\sim$, and show that $m_\sim$ is a well-defined graded membership function provided that $\sim$ satisfies appropriate properties collected under the name standard concept measures. Afterwards, we prove that, in general, reasoning in logics of the form $\tau\mathcal{EL}(m_\sim)$ may be undecidable, but becomes decidable if $\sim$ is computable and some additional properties are satisfied. Finally, we introduce a concrete family of concept measures $\sim$, whose induced logics $\tau\mathcal{EL}(m_\sim)$ are the subject of study in the remaining sections of the paper. These measures are based on the simi framework proposed in [14], but in contrast to our previous work in [11,17], we use directional instances of simi rather than undirectional ones. The reasons for this decision are explained in detail in Subsection 3.4.

3.1. From concept measures to graded membership functions

In their most general form, *concept measures* (CMs) are functions that map pairs of concept descriptions C and D to values in the unit interval. In this paper, we restrict the attention to $\mathcal{EL}$ concept descriptions, and thus a concept measure is a mapping

$\sim\, : \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR}) \times \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR}) \to [0,1]$. We write $\sim$ in infix notation, i.e., use $C \sim D = t$ rather than $\sim(C,D) = t$ to indicate that the value of applying $\sim$ to C and D is $t \in [0,1]$.

Such measures are supposed to generalize crisp relations between concepts (which hold or do not hold) to graded relations, which hold to a certain degree. For example, *concept similarity measures* (CSiMs) generalize equivalence between concepts whereas *concept subsumption measures* (CSuMs) generalize subsumption. Intuitively, for a CSiM $\sim$, the larger the value $C \sim D$ is, the more similar the concept descriptions C and D are supposed to be. For a CSuM $\sim$, the value of $C \sim D$ expresses to what degree C subsumes D (i.e., it generalizes the crisp relation $C \sqsupseteq D$). Examples of such measures for $\mathcal{EL}$ and other DLs as well as properties they should satisfy can, for instance, be found in [13–15,12,16]. In Subsection 3.3 we will introduce a class of CMs that are inspired by the framework for defining CSiMs in [14]. To illustrate the definitions in the present subsection, we introduce a simple measure that is akin to the graded membership function defined in Example 2.12.

Example 3.1. Given $C, D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR})$, we have $C \sqsupseteq D$ iff $\mathsf{At} \sqsupseteq D$ holds for all $\mathsf{At} \in top(C)$ [28]. Basically, we define a CSuM $\sim_{\mathsf{su}}$ that generalizes this characterization of subsumption in $\mathcal{EL}$ by counting how many of the atoms of C satisfy this property. However, to make the counting independent of the occurrence of redundant atoms, we consider the reduced form C^r of C rather than C itself, i.e., we set

$$C \sim_{\mathsf{su}} D := \frac{|\{\mathsf{At} \in top(C^r) \mid \mathsf{At} \sqsupseteq D\}|}{|top(C^r)|}.$$

Note that, since $D \equiv D^r$, it is irrelevant whether we use D or D^r in this definition, i.e., we have $C \sim_{\mathsf{su}} D = C \sim_{\mathsf{su}} D^r$. Taking into account that equivalence corresponds to subsumption in both directions, we can obtain a CSiM from $\sim_{\mathsf{su}}$ by combining the values of $C \sim_{\mathsf{su}} D$ and $D \sim_{\mathsf{su}} C$. For example, we can take the average or the minimum as combining function. $\triangle$

Following an idea in [12,29] for how to relax instance queries using concept similarity measures, we now use CMs to define graded membership functions.

Definition 3.2. Let $\sim$ be a CM. Then, for all interpretations $\mathcal{I}$, the function $m_\sim^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR}) \to [0,1]$ is defined as

$$m_\sim^{\mathcal{I}}(d, C) := \max\{C \sim D \mid D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR}) \text{ and } d \in D^{\mathcal{I}}\}.$$

For an arbitrary CM $\sim$, the maximum in this definition need not exist since D ranges over infinitely many concept descriptions. To obtain functions $m_\sim$ that are well-defined and satisfy the properties required in Definition 2.11, we restrict the attention to *standard* CMs, which are measures satisfying several properties that we introduce in the following.

To achieve well-definedness of the maximum in Definition 3.2, we try to ensure that it is sufficient to restrict the attention to concept descriptions D that range over a set of the form $\mathcal{R}^k(\mathsf{NC}^f, \mathsf{NR}^f)$ for finite sets $\mathsf{NC}^f \subseteq \mathsf{NC}$ and $\mathsf{NR}^f \subseteq \mathsf{NR}$ (see Lemma 2.9). To this end, we require $\sim$ to satisfy the following three properties.

1. The CM $\sim$ is *equivalence invariant* if $C \equiv C'$ and $D \equiv D'$ imply $C \sim D = C' \sim D'$ for all $C, C', D, D' \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR})$.

The CM $\sim_{\mathsf{su}}$ introduced in Example 3.1 is equivalence invariant. This is an easy consequence of the fact that $C \equiv C'$ implies that the sets $top(C^r)$ and $top(C'^r)$ are equal up to equivalence.

2. The CM $\sim$ is *role-depth reducing* if there is a function $\mathfrak{r}_\sim : \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR}) \to \mathbb{N}$ such that, for all $C, D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR})$, there exists $D' \in \mathcal{C}_{\mathcal{EL}}$ satisfying:

   - $C \sim D \le C \sim D'$, $\mathsf{rd}(D') \le \mathfrak{r}_\sim(C)$ and $D \sqsubseteq D'$.

   We call such a function $\mathfrak{r}_\sim$ an *r-reducing function* for $\sim$.

The CM $\sim_{\mathsf{su}}$ is also role-depth reducing. In fact, if we set $\mathfrak{r}_{\sim_{\mathsf{su}}}(C) := \mathsf{rd}(C)$, then Lemma 2.10 implies that this function is an r-reducing function for $\sim_{\mathsf{su}}$ since we can use $D|_k$ for $k = \mathsf{rd}(C)$ as D'. Note that in this case we even have $C \sim D = C \sim D'$ since $C \sim D \ge C \sim D'$ follows from $D \sqsubseteq D'$.

The two properties introduced until now ensure that $m_\sim^{\mathcal{I}}(d, C)$ can be expressed in terms of $\mathcal{EL}$ concept descriptions in reduced form of role depth at most $\mathfrak{r}_\sim(C)$.

Lemma 3.3. *If a CM $\sim$ is equivalence invariant and role-depth reducing, then*

$$m_\sim^{\mathcal{I}}(d, C) = \max\{C \sim D' \mid D' \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR}),\ \mathsf{rd}(D') \le \mathfrak{r}_\sim(C),\ D' \textit{ is reduced, and } d \in D'^{\mathcal{I}}\}.$$

However, since NC and NR are infinite sets, the number of such concept descriptions may still be infinite. For this reason we require a third property called *signature reduction*:

3. The CM $\sim$ is *signature reducing* if there exists a function $\mathfrak{s}_\sim : \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR}) \to \mathcal{P}(\mathsf{NC} \cup \mathsf{NR})$ such that for all $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR})$:

   - $\mathfrak{s}_\sim(C)$ is a finite set,

   and, for all $C, D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR})$, there exists $D' \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR})$ satisfying the following properties:

   - $C \sim D \leq C \sim D'$, $\mathsf{sig}(D') \subseteq \mathfrak{s}_\sim(C)$ and $D \sqsubseteq D'$.

   We call such a function $\mathfrak{s}_\sim$ an *s-reducing function* for $\sim$. In addition, we denote as $\mathfrak{s}^{\mathsf{C}}_\sim(C)$ and $\mathfrak{s}^{\mathsf{R}}_\sim(C)$ the sets $\mathfrak{s}_\sim(C) \cap \mathsf{NC}$ and $\mathfrak{s}_\sim(C) \cap \mathsf{NR}$, respectively.

The CM $\sim_{\mathsf{su}}$ is signature reducing as well. In fact, we can set $\mathfrak{s}_\sim(C) := \mathsf{sig}(C)$ and construct D' from C and D by replacing in D concept names $A \notin \mathsf{sig}(C)$ and existential restrictions $\exists r.E$ for $r \notin \mathsf{sig}(C)$ with $\top$. Again, we even have $C \sim D = C \sim D'$ in this case.

Signature reduction guarantees that, to compute $m^{\mathcal{I}}(d, C)$, only concept descriptions D defined over $\mathfrak{s}^{\mathsf{C}}_\sim(C)$ and $\mathfrak{s}^{\mathsf{R}}_\sim(C)$ need to be considered. Since $\mathfrak{s}_\sim(C)$ is required to be finite, this implies the existence of the maximum in the definition of $m^{\mathcal{I}}(d, C)$. Let us now formally show that CMs satisfying the three properties introduced above induce well-defined functions $m_\sim$.

Proposition 3.4. *If $\sim$ is an equivalence invariant, role-depth reducing and signature reducing CM, then $m_\sim$ is well-defined.*

Proof. We need to show that the maximum in the definition of $m^{\mathcal{I}}(d, C)$ exists. Signature reduction tells us that, if $d \in D^{\mathcal{I}}$ for some $D \in \mathcal{C}_{\mathcal{EL}}$, then there exists D' such that $C \sim D \leq C \sim D'$, $d \in D'^{\mathcal{I}}$ and D' is defined over $\mathfrak{s}^{\mathsf{C}}_\sim(C)$ and $\mathfrak{s}^{\mathsf{R}}_\sim(C)$. This means that one can restrict the attention to concept descriptions defined over $\mathfrak{s}^{\mathsf{C}}_\sim(C)$ and $\mathfrak{s}^{\mathsf{R}}_\sim(C)$, i.e.:

$$m^{\mathcal{I}}_\sim(d, C) = \max\{C \sim D \mid D \in \mathcal{C}_{\mathcal{EL}}(\mathfrak{s}^{\mathsf{C}}_\sim(C), \mathfrak{s}^{\mathsf{R}}_\sim(C)) \text{ and } d \in D^{\mathcal{I}}\}.$$

Now, let $k = \mathfrak{r}_\sim(C)$. Since $\sim$ is role-depth reducing, one can always find for each $D \in \mathcal{C}_{\mathcal{EL}}(\mathfrak{s}^{\mathsf{C}}_\sim(C), \mathfrak{s}^{\mathsf{R}}_\sim(C))$ a concept $D' \sqsupseteq D$ such that $C \sim D \leq C \sim D'$ and $\mathsf{rd}(D') \leq k$. In addition, by (3), $D \sqsubseteq D'$ implies that D' is also defined over $\mathfrak{s}^{\mathsf{C}}_\sim(C)$ and $\mathfrak{s}^{\mathsf{R}}_\sim(C)$. Hence, using equivalence invariance of $\sim$, we can further restrict the search for an appropriate D to finitely many concept descriptions:

$$m^{\mathcal{I}}_\sim(d, C) = \max\{C \sim D \mid D \in \mathcal{R}^k(\mathfrak{s}^{\mathsf{C}}_\sim(C), \mathfrak{s}^{\mathsf{R}}_\sim(C)) \text{ and } d \in D^{\mathcal{I}}\}. \tag{10}$$

Since $\mathfrak{s}^{\mathsf{C}}_\sim(C)$ and $\mathfrak{s}^{\mathsf{R}}_\sim(C)$ are finite, the application of Lemma 2.9 yields that $\mathcal{R}^k(\mathfrak{s}^{\mathsf{C}}_\sim(C), \mathfrak{s}^{\mathsf{R}}_\sim(C))$ is also a finite set. Hence, the maximum in the previous equation always exists and $m_\sim$ is well-defined. □

Since we have seen that the CM $\sim_{\mathsf{su}}$ introduced in Example 3.1 satisfies the three properties required by the above proposition, we thus know that the function $m_{\sim_{\mathsf{su}}}$ is well-defined. We claim that $m_{\sim_{\mathsf{su}}}$ is actually equal to the simple graded membership function m_{s} introduced in Example 2.12, i.e., that for every interpretation $\mathcal{I}$ and every $d \in \Delta^{\mathcal{I}}$ we have

$$\frac{|\{\mathsf{At} \in top(C^r) \mid d \in \mathsf{At}^{\mathcal{I}}\}|}{|top(C^r)|} = \max\left\{ \frac{|\{\mathsf{At} \in top(C^r) \mid \mathsf{At} \sqsupseteq D\}|}{|top(C^r)|} \mid D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR}) \text{ and } d \in D^{\mathcal{I}} \right\}. \tag{11}$$

To prove this, we consider the $\mathcal{EL}$ concept description $D_m := \bigsqcap\{\mathsf{At} \in top(C^r) \mid d \in \mathsf{At}^{\mathcal{I}}\}$. First, note that $d \in D^{\mathcal{I}}_m$ and that $\{\mathsf{At} \in top(C^r) \mid d \in \mathsf{At}^{\mathcal{I}}\} = \{\mathsf{At} \in top(C^r) \mid \mathsf{At} \sqsupseteq D_m\}$. The inclusion from left to right is an immediate consequence of the definition of D_m, and the inclusion in the other direction follows from the fact that the elements of $top(C^r)$ are not comparable w.r.t. subsumption since C^r is reduced. Thus, it is sufficient to show that D_m yields the maximal value on the right-hand side of identity (11). This clearly follows from the fact that $d \in D^{\mathcal{I}}$ for an $\mathcal{EL}$ concept description D implies $\{\mathsf{At} \in top(C^r) \mid \mathsf{At} \sqsupseteq D\} \subseteq \{\mathsf{At} \in top(C^r) \mid d \in \mathsf{At}^{\mathcal{I}}\}$. This completes the proof that $m_{\sim_{\mathsf{su}}} = m_{\mathsf{s}}$.

Coming back to the general case, the second step is to ensure that the function $m_\sim$ satisfies the properties stated in Definition 2.11 for graded membership functions. To this end, we require $\sim$ to be either equivalence closed or subsumption closed, which are properties often considered for CSiMs and CSuMs, respectively. More precisely, we say that the CM $\sim$ is *closed* if it is

- *equivalence closed:* for all $C, D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR})$ we have ($C \equiv D$ iff $C \sim D = 1$), or
- *subsumption closed:* for all $C, D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR})$ we have ($D \sqsubseteq C$ iff $C \sim D = 1$).

It is easy to see that $\sim_{\mathsf{su}}$ is subsumption closed.

Definition 3.5. A CM $\sim$ is called *standard* if it is equivalence invariant, role-depth reducing, signature reducing, and closed.

Standard CMs induce well-defined graded membership functions.

Proposition 3.6. *If $\sim$ is a standard CM, then $m_\sim$ is a graded membership function.*

Proof. As already shown in Proposition 3.4, the function $m_\sim$ is well-defined. Hence, it remains to show that $m_\sim$ satisfies properties *M1–M3* of Definition 2.11.

Property *M1* for $m_\sim$ follows from the assumption that $\sim$ is closed. In fact, if $d \in C^{\mathcal{I}}$ then $m_\sim^{\mathcal{I}}(d,C) \geq C \sim C$. Since $C \equiv C$ and $C \sqsubseteq C$, we have $C \sim C = 1$, and hence $m_\sim^{\mathcal{I}}(d,C) = 1$. Conversely, suppose that $m_\sim^{\mathcal{I}}(d,C) = 1$. This means that there exists a concept description D such that $C \sim D = 1$ and $d \in D^{\mathcal{I}}$. If $\sim$ is equivalence closed, then $C \equiv D$. Otherwise, it must be subsumption closed, which implies $D \sqsubseteq C$. In both cases we have that $d \in C^{\mathcal{I}}$.

Regarding property *M2*, as mentioned in Section 2.2, the direction from right to left is a consequence of *M1*. The other direction follows from equivalence invariance of $\sim$.

To show property *M3*, assume that $\phi : \mathcal{I} \to \mathcal{J}$ is a homomorphism and $d \in \Delta^{\mathcal{I}}$. Then $d \in D^{\mathcal{I}}$ implies $\phi(d) \in D^{\mathcal{J}}$ for all $D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR})$, which shows that $\{D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR}) \mid d \in D^{\mathcal{I}}\} \subseteq \{D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR}) \mid \phi(d) \in D^{\mathcal{J}}\}$, and thus $m^{\mathcal{I}}(d,C) = \max\{C \sim D \mid D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR}) \text{ and } d \in D^{\mathcal{I}}\} \leq \max\{C \sim D \mid D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR}) \text{ and } \phi(d) \in D^{\mathcal{J}}\} = m^{\mathcal{J}}(\phi(d),C)$. □

3.2. Reasoning in $\tau\mathcal{EL}(m_\sim)$

We start by proving that in general reasoning in logics of the form $\tau\mathcal{EL}(m_\sim)$ for standard CMs $\sim$ may be undecidable. To show this undecidability result, we first construct a class of standard CMs that contains uncountably many non-computable functions.

Definition 3.7. Let $N \subseteq \mathbb{N}$, $0 < a < 1$ a fixed rational number, $A \in \mathsf{NC}$ and $r \in \mathsf{NR}$. We define the concept measure $\sim_N^a$ as:

$$C \sim_N^a D := \begin{cases} 1 & \text{if } C \equiv D \\ \mu(C,D) & \text{otherwise,} \end{cases} \quad \text{where} \quad \mu(C,D) := \begin{cases} a & \text{if } \mathsf{rd}(C) = \mathsf{rd}(D) \in N \text{ and } \mathsf{sig}(D) \subseteq \{A, r\} \\ 0 & \text{otherwise.} \end{cases}$$

We now show that $\sim_N^a$ is a standard CM.

Lemma 3.8. *Let $N \subseteq \mathbb{N}$ and $\sim_N^a$ be defined as in Definition 3.7. Then, $\sim_N^a$ is a standard CM.*

Proof. *Equivalence closedness* of $\sim_N^a$ is clear from its definition. Thus, consider the other three properties required for a standard CM.

1. *Equivalence invariance*: let $C, C', D, D' \in \mathcal{C}_{\mathcal{EL}}$ such that $C \equiv C'$ and $D \equiv D'$. Then $C \sim_N^a D = C' \sim_N^a D'$ is an easy consequence of the following facts: $C \equiv D$ iff $C' \equiv D'$, $\mathsf{rd}(C) = \mathsf{rd}(C')$, $\mathsf{rd}(D) = \mathsf{rd}(D')$, and $\mathsf{sig}(D) = \mathsf{sig}(D')$.
2. *Role-depth reduction*: let $\mathfrak{r} : \mathcal{C}_{\mathcal{EL}} \to \mathbb{N}$ be such that $\mathfrak{r}(C) = \mathsf{rd}(C) + 1$ for all $C \in \mathcal{C}_{\mathcal{EL}}$. We show that $\mathfrak{r}$ is an r-reducing function for $\sim_N^a$ by setting $D' := D|_{\mathfrak{r}(C)}$. The case where $\mathsf{rd}(D) \leq \mathfrak{r}(C)$ is trivial since then $D|_{\mathfrak{r}(C)} = D$. Note that this also includes the case where $C \equiv D$. Otherwise, we have that $\mathsf{rd}(D) > \mathsf{rd}(D') > \mathsf{rd}(C)$. This means that $C \sim_N^a D = 0 = C \sim_N^a D'$. Hence, since $D \sqsubseteq D|_{\mathfrak{r}(C)}$, it follows that $\mathfrak{r}$ is r-reducing for $\sim_N^a$, and consequently $\sim_N^a$ is role-depth reducing.
3. *Signature reduction:* let $\mathfrak{s}$ be the function $\mathfrak{s}(C) := \mathsf{sig}(C) \cup \{A, r\}$, for all $C \in \mathcal{C}_{\mathcal{EL}}$. We prove that $\mathfrak{s}$ is an s-reducing function for $\sim_N^a$. Let $C, D \in \mathcal{C}_{\mathcal{EL}}$. If $C \sim_N^a D > 0$, we choose D' as D. Otherwise, D' is defined to be $\top$. In both cases it is easy to verify that $\mathsf{sig}(D') \subseteq \mathfrak{s}(C)$, $C \sim_N^a D \leq C \sim_N^a D'$ and $D \sqsubseteq D'$. Thus, since $\mathfrak{s}(C)$ is finite, this means that $\sim_N^a$ is signature reducing.

Summing up, we have shown that $\sim_N^a$ satisfies all the properties required for a standard CM. □

If we fix a, then each subset N of the natural numbers induces a standard CM $\sim_N^a$, and thus also a threshold logic $\tau\mathcal{EL}(m_{\sim_N^a})$. Since there are uncountably many such sets, but only countably many Turing machines, it is clear that uncountably many of these sets are undecidable. For undecidable sets N, the threshold logic $\tau\mathcal{EL}(m_{\sim_N^a})$ is undecidable as well.

Proposition 3.9. *If N is undecidable, then the function $\sim_N^a$ is not computable, and satisfiability in the logic $\tau\mathcal{EL}(m_{\sim_N^a})$ is undecidable.*

Proof. Assume that $\sim_N^a$ is computable. We claim that this yields a decision procedure for membership in N. In fact, given a natural number n, we can build the concept descriptions $C_n := \exists r.\cdots\exists r.\top$ and $D_n := \exists r.\cdots\exists r.A$ of role depth n. Obviously, $C_n \not\equiv D_n$, $\mathsf{rd}(C_n) = \mathsf{rd}(D_n) = n$ and $\mathsf{sig}(D_n) = \{r, A\}$. Thus, $C_n \sim_N^a D_n = a$ iff $n \in N$. Consequently, if we can compute the value $C_n \sim_N^a D_n$, then we can decide whether $n \in N$ or not.

Regarding satisfiability in $\tau\mathcal{EL}(m_{\sim_N^a})$, we consider the $\tau\mathcal{EL}(m_{\sim_N^a})$ concept description $\widehat{D}_n := (D_n)_{\geq a} \sqcap (D_n)_{\leq a}$, and claim that this concept description is satisfiable iff $n \in N$.

First, assume that $n \in N$. To show satisfiability of $\widehat{D}_n$, we consider the interpretation $\mathcal{I}$ that has domain $\Delta^{\mathcal{I}} := \{d_0, \ldots, d_n\}$, interprets A as $A^{\mathcal{I}} := \emptyset$ and r as $r^{\mathcal{I}} := \{(d_i, d_{i+1}) \mid 0 \leq i < n\}$. We have $d_0 \in C_n^{\mathcal{I}}$ and since $D_n \sim_N^a C_n = a$, we know that $m_{\sim_N^a}^{\mathcal{I}}(d_0, D_n) \geq a$. This shows $d_0 \in [(D_n)_{\geq a}]^{\mathcal{I}}$. To show that $d_0 \in [(D_n)_{\leq a}]^{\mathcal{I}}$, it is sufficient to show that, for any concept description $D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC},\mathsf{NR})$ with $d_0 \in D^{\mathcal{I}}$, we have $D_n \sim_N^a D \leq a$. This is an immediate consequence of the fact that $D_n \not\equiv D$ since $d_0 \in D^{\mathcal{I}} \setminus D_n^{\mathcal{I}}$. Consequently, we have shown that $d_0 \in \widehat{D}_n^{\mathcal{I}}$, which proves that $\widehat{D}_n$ is satisfiable.

Second, assume that $n \notin N$. Obviously this implies $D_n \sim_N^a D \in \{0,1\}$ for all concept descriptions $D \in \mathcal{C}_{\mathcal{EL}}$. Consequently, we have, for all interpretations $\mathcal{I}$ and elements $d \in \Delta^{\mathcal{I}}$, that $m^{\mathcal{I}}_{\sim_N^a}(d, D_n) \in \{0,1\}$. If $m^{\mathcal{I}}_{\sim_N^a}(d, D_n) = 0$, then $d \notin [(D_n)_{\geq a}]^{\mathcal{I}}$, and if $m^{\mathcal{I}}_{\sim_N^a}(d, D_n) = 1$, then $d \notin [(D_n)_{\leq a}]^{\mathcal{I}}$. This shows that $\widehat{D}_n$ is unsatisfiable. □

Since satisfiability can be reduced to the other inference problems in $\tau\mathcal{EL}(m_{\sim_N^a})$, this proposition also shows that these other problems are undecidable if N is undecidable.

Corollary 3.10. *If N is undecidable, then the subsumption, the consistency, and the instance problem are undecidable in $\tau\mathcal{EL}(m_{\sim_N^a})$.*

Proof. We have the following reductions:

- $\widehat{C}$ is unsatisfiable iff $\widehat{C} \sqsubseteq A \sqcap A_{<1}$;
- $\widehat{C}$ is satisfiable iff $\{\widehat{C}(a)\}$ is consistent;
- $\widehat{C} \sqsubseteq \widehat{D}$ iff $\{\widehat{C}(a)\} \models \widehat{D}(a)$.

Thus, satisfiability can be reduced to the subsumption, the consistency, and the instance problem. Since satisfiability in $\tau\mathcal{EL}(m_{\sim_N^a})$ is undecidable, this implies undecidability of the other three problems. □

Since undecidability of N implies that $\sim_N^a$ is not computable, all our examples of undecidable logics $\tau\mathcal{EL}(m_\sim)$ are such that $\sim$ is not computable. We will now show that computability of $\sim$ implies decidability of $\tau\mathcal{EL}(m_\sim)$ w.r.t. arbitrary $\tau\mathcal{EL}(m)$ TBoxes, provided that $\sim$ has computable r-reducing and s-reducing functions $\mathfrak{r}_\sim$ and $\mathfrak{s}_\sim$. Our proof of this fact employs an equivalence preserving and computable translation of $\tau\mathcal{EL}(m_\sim)$ concept descriptions into $\mathcal{ALC}$ concept descriptions. Since the standard reasoning problems are decidable in $\mathcal{ALC}$, such an effective translation obviously yields their decidability in $\tau\mathcal{EL}(m_\sim)$.

Recall that $\mathcal{ALC}$ [4] is obtained from $\mathcal{EL}$ by adding negation $\neg C$. Clearly, negation together with conjunction also yields disjunction $C \sqcup D$. Since $\mathcal{EL}$ is a fragment of $\mathcal{ALC}$, it suffices to show how to translate threshold concepts $C_{\bowtie t}$ into $\mathcal{ALC}$ concept descriptions. In addition, we can concentrate on the case where $\bowtie \in \{\geq, >\}$ since $C_{<t} \equiv \neg C_{\geq t}$ and $C_{\leq t} \equiv \neg C_{>t}$.

Lemma 3.11. *Let $\bowtie \in \{\geq, >\}$, $t \in [0,1] \cap \mathbb{Q}$, $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$ and $k = \mathfrak{r}_\sim(C)$. Then,*

$$C_{\bowtie t} \equiv \bigsqcup \{D \mid D \in \mathcal{R}^k(\mathfrak{s}^{\mathsf{C}}_\sim(C), \mathfrak{s}^{\mathsf{R}}_\sim(C)) \text{ and } C \sim D \bowtie t\}. \tag{12}$$

Proof. Let $\mathcal{I}$ be an interpretation and $d \in \Delta^{\mathcal{I}}$. By the semantics of threshold concepts and (10), we know that $d \in (C_{\bowtie t})^{\mathcal{I}}$ iff $\max\{C \sim D \mid D \in \mathcal{R}^k(\mathfrak{s}^{\mathsf{C}}_\sim(C), \mathfrak{s}^{\mathsf{R}}_\sim(C)) \text{ and } d \in D^{\mathcal{I}}\} \bowtie t$. Since $\bowtie \in \{\geq, >\}$, this is equivalent to saying that there is a $D \in \mathcal{R}^k(\mathfrak{s}^{\mathsf{C}}_\sim(C), \mathfrak{s}^{\mathsf{R}}_\sim(C))$ such that $C \sim D \bowtie t$ and $d \in D^{\mathcal{I}}$. This is in turn equivalent to $d \in \bigcup \{D^{\mathcal{I}} \mid D \in \mathcal{R}^k(\mathfrak{s}^{\mathsf{C}}_\sim(C), \mathfrak{s}^{\mathsf{R}}_\sim(C)) \text{ and } C \sim D \bowtie t\}$. □

Since $\mathcal{R}^k(\mathfrak{s}^{\mathsf{C}}_\sim(C), \mathfrak{s}^{\mathsf{R}}_\sim(C))$ is finite, the disjunction on the right-hand side of the equivalence in (12) is finite, and thus this right-hand side is an admissible $\mathcal{ALC}$ concept description. Based on this, every $\tau\mathcal{EL}(m)$ TBox $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$ can be translated into an equivalent $\mathcal{ALC}$ TBox.[9] More precisely, the equivalent TBox consists of the set of axioms $\widehat{\mathcal{T}} \cup \mathcal{T}$, where each occurrence of a threshold concept $C_{\bowtie t}$ in $\widehat{\mathcal{T}}$ is replaced with the $\mathcal{ALC}$ concept that results from translating $(u_{\mathcal{T}}(C))_{\bowtie t}$. The restrictions imposed on $\mathfrak{T}$ in Definitions 2.20 and 2.27 ensure that this transformation preserves acyclicity of $\mathfrak{T}$. The following example illustrates this transformation for the DL $\tau\mathcal{EL}(m_{\sim_{\mathsf{su}}})$ induced by the CM $\sim_{\mathsf{su}}$ of Example 3.1.

Example 3.12. Let $m = m_{\sim_{\mathsf{su}}}$. In addition, consider the $\tau\mathcal{EL}(m)$ TBox $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$, where:

$$\widehat{\mathcal{T}} := \left\{ \begin{array}{l} \alpha \doteq \exists s.E \sqcap \exists r.\beta \\ \beta \doteq E_{\geq 1/2} \sqcap B \\ A \sqsubseteq F_{<0.6} \end{array} \right\} \qquad \mathcal{T} := \left\{ \begin{array}{l} E \doteq \exists r.A \sqcap \exists r.B \\ F \doteq A \sqcap B \end{array} \right\}$$

To obtain an $\mathcal{ALC}$ TBox that is equivalent to $\mathfrak{T}$, we replace $E_{\geq 1/2}$ with the $\mathcal{ALC}$ concept that results from translating $u_{\mathcal{T}}(E)_{\geq 1/2}$, and $F_{<0.6}$ with the negation of the $\mathcal{ALC}$ concept that results from translating $u_{\mathcal{T}}(F)_{\geq 0.6}$. In order to apply the translation in (12) to $u_{\mathcal{T}}(E)_{\geq 1/2}$ and $u_{\mathcal{T}}(F)_{\geq 0.6}$ w.r.t. $\sim_{\mathsf{su}}$, we use the functions $\mathfrak{r}_{\sim_{\mathsf{su}}}(C) = \mathsf{rd}(C)$ and $\mathfrak{s}_{\sim_{\mathsf{su}}}(C) = \mathsf{sig}(C)$ introduced above for $\sim_{\mathsf{su}}$. For $u_{\mathcal{T}}(E)_{\geq 1/2}$ we then have:

$$u_{\mathcal{T}}(E)_{\geq 1/2} = (\exists r.A \sqcap \exists r.B)_{\geq 1/2}, \qquad \mathfrak{r}_{\sim_{\mathsf{su}}}(\exists r.A \sqcap \exists r.B) = 1, \qquad \mathfrak{s}_{\sim_{\mathsf{su}}}(\exists r.A \sqcap \exists r.B) = \{r, A, B\}.$$

Consequently, the translation of $u_{\mathcal{T}}(E)_{\geq 1/2}$ corresponds to the $\mathcal{ALC}$ concept consisting of the disjunction of all $\mathcal{EL}$ concepts D satisfying the following two properties:

[9] By equivalent we mean that they have the same models.

- $D \in \mathcal{R}^1(\{A, B\}, \{r\})$. This means that D is an $\mathcal{EL}$ concept defined over $\{A, B\}$ and $\{r\}$, which is in reduced form and has role-depth at most 1.
- $(\exists r.A \sqcap \exists r.B) \sim_{\mathsf{su}} D \geq 1/2$. By definition of $\sim_{\mathsf{su}}$, this means that D is subsumed by at least one top-level atom of $\exists r.A \sqcap \exists r.B$, i.e., $\exists r.A \sqsupseteq D$ or $\exists r.B \sqsupseteq D$.

The set G of all such $\mathcal{EL}$ concepts D can be expressed as follows:

$$G := \{X, X \sqcap Y \mid X \in \{\exists r.A, \exists r.B, \exists r.(A \sqcap B), \exists r.A \sqcap \exists r.B\} \text{ and } Y \in \{A, B, A \sqcap B\}\}.$$

Hence, $u_{\mathcal{T}}(E)_{\geq 1/2}$ is translated into the $\mathcal{ALC}$ concept $\bigsqcup_{D\in G} D$. Regarding $u_{\mathcal{T}}(F)_{\geq 0.6}$, we have:

$$u_{\mathcal{T}}(F)_{\geq 0.6} = (A \sqcap B)_{\geq 0.6}, \qquad \mathfrak{r}_{\sim_{\mathsf{su}}}(A \sqcap B) = 0, \qquad \mathfrak{s}_{\sim_{\mathsf{su}}}(A \sqcap B) = \{A, B\}.$$

In this case, the relevant disjuncts D are those such that $D \in \mathcal{R}^0(\{A, B\}, \emptyset)$ and $(A \sqcap B) \sim_{\mathsf{su}} D \geq 0.6$. It is easy to see that there is only one such $\mathcal{EL}$ concept (modulo $\equiv$), namely, $D = A \sqcap B$. Hence, $F_{<0.6}$ is translated into the $\mathcal{ALC}$ concept $\neg(A \sqcap B)$. Thus, $\mathfrak{T}$ is transformed into the following $\mathcal{ALC}$ TBox:

$$\mathcal{T}' := \left\{ \begin{array}{lll} \alpha \doteq \exists s.E \sqcap \exists r.\beta, & \beta \doteq \left(\bigsqcup_{D\in G} D\right) \sqcap B, & A \sqsubseteq \neg(A \sqcap B), \\ E \doteq \exists r.A \sqcap \exists r.B, & F \doteq A \sqcap B & \end{array} \right\}$$

Finally, note that removing $A \sqsubseteq F_{<0.6}$ from $\mathfrak{T}$ results in an acyclic $\tau\mathcal{EL}(m)$ TBox, whose translation yields $\mathcal{T}' \setminus \{A \sqsubseteq \neg(A \sqcap B)\}$ which is an acyclic $\mathcal{ALC}$ TBox. $\triangle$

Proposition 3.13. *For each acyclic (general) $\tau\mathcal{EL}(m)$ TBox $\mathfrak{T}$ there exists an acyclic (general) $\mathcal{ALC}$ TBox $\mathcal{T}'$ such that $\mathcal{I} \models \mathfrak{T}$ iff $\mathcal{I} \models \mathcal{T}'$ holds for all interpretations $\mathcal{I}$.*

Finally, if $\mathfrak{r}_\sim$ and $\mathfrak{s}_\sim$ are computable, then Lemma 2.9 implies that the set $\mathcal{R}^k(\mathfrak{s}^{\mathsf{C}}_\sim(C), \mathfrak{s}^{\mathsf{R}}_\sim(C))$ is computable. Hence, since $\sim$ is also computable by assumption, the $\mathcal{ALC}$ concept description on the right-hand side of (12) can effectively be computed.

Theorem 3.14. *Let $\sim$ be a computable standard CM that has computable r-reducing and s-reducing functions. Then, in $\tau\mathcal{EL}(m_\sim)$, satisfiability, subsumption, consistency and instance checking w.r.t. general $\tau\mathcal{EL}(m)$ TBoxes are decidable.*

The CM $\sim_{\mathsf{su}}$ of Example 3.1 is clearly computable, and so are the r-reducing and s-reducing functions for $\sim_{\mathsf{su}}$ we have introduced above. Consequently, reasoning in $\tau\mathcal{EL}(m_{\sim_{\mathsf{su}}})$ is decidable.

Since the cardinality of $\mathcal{R}^k(\mathfrak{s}^{\mathsf{C}}_\sim(C), \mathfrak{s}^{\mathsf{R}}_\sim(C))$ increases by one exponent with each increase of k, Theorem 3.14 yields only a *non-elementary* bound on the complexity of reasoning in $\tau\mathcal{EL}(m_\sim)$. The rest of this section is dedicated to introducing a restricted family of standard concept measures that, as we will show in Sections 4 and 5, induce threshold DLs with better complexity upper bounds for their reasoning problems.

3.3. A restricted class of concept measures

The restricted class of CMs we are about to introduce is obtained from the *simi framework* proposed in [14]. This framework can be used to define a variety of CSiMs and CSuMs for $\mathcal{EL}$ satisfying certain desirable properties. Here, we describe a fragment of *simi* that is sufficient for our purposes.

In this framework, one first defines a directional measure $\sim_d$, i.e., a CSuM that generalizes subsumption. If one is interested in obtaining a CSiM, which generalizes equivalence and must be symmetric, then one combines the values obtained by comparing the input concepts in both directions in an appropriate way. Intuitively, the definition of the directional measure $\sim_d$ is based on the following recursive characterization of subsumption in $\mathcal{EL}$ [28]: $C \sqsupseteq D$ iff

- every concept name $A \in top(C)$ also occurs in $top(D)$, and
- for every existential restriction $\exists r.E \in top(C)$ there is $\exists r.F \in top(D)$ such that $E \sqsupseteq F$.

When defining the value of $C \sim_d D$, we are looking, for every element $\mathsf{At} \in top(C)$, for an element $\mathsf{At}' \in top(D)$ that matches it best, and return the corresponding value. This value is obtained as follows. For concept names A, B, we do not require that they are equal, but use the value provided by a predefined primitive measure. For existential restrictions $\exists r.E, \exists s.F$, we also use the value provided by the primitive measure for the role names r, s and combine this value with the value $E \sim_d F$ obtained by a recursive application of $\sim_d$. In this combination, we can give more or less weight w to the comparison of the role names relative to the recursive comparison. Basically, to obtain the value $C \sim_d D$, we then take, for every element of $top(C)$, the maximal value obtained this way by comparing it with the elements of $top(D)$, sum up these maximal values over all elements of $top(C)$, and divide by the cardinality of $top(C)$. What is actually done in the definition of $\sim_d$ is a bit more complicated since we also assign a weight $g(\mathsf{At})$ to all atoms, which gives the value obtained for this atom more or less importance. Finally, to avoid counting redundant atoms twice, we use the reduced forms of C, D.

The formal definition of $C \sim_d D$ introduced below thus depends on the following parameters:

- A primitive measure $pm : (\mathsf{NC} \times \mathsf{NC}) \cup (\mathsf{NR} \times \mathsf{NR}) \to [0,1] \cap \mathbb{Q}$, satisfying the following property:
  - $pm(X,Y) = 1$ iff $X = Y$, for all $X, Y \in \mathsf{NC} \cup \mathsf{NR}$.

  In particular, the *default* primitive measure pm_{df} is defined as:

$$pm_{df}(X,Y) := \begin{cases} 1 & \text{if } X = Y \\ 0 & otherwise. \end{cases}$$

- A rational factor $w \in [0,1)$ and a function g that assigns to every concept and role name a weight in $\mathbb{Q}_{>0}$. This function is then extended to all $\mathcal{EL}$ atoms (i.e., concept name or existential restriction) by setting $g(\exists r.C) := g(r)$.[10]

Fixing these parameters yields a unique directional measure $\sim_d$, which in turn can be used to define a CSiM.

Definition 3.15. Let $C, D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$. If $C \equiv \top$, then $C \sim_d D := 1$; if $C \not\equiv \top$ and $D \equiv \top$, then $C \sim_d D := 0$; otherwise, we define

$$C \sim_d D := \frac{\sum\limits_{C' \in top(C^r)} \left[g(C') \cdot \max\limits_{D' \in top(D^r)} \left(simi_a(C', D') \right) \right]}{\sum\limits_{C' \in top(C^r)} g(C')}, \text{ where} \tag{13}$$

$simi_a(A,B) := pm(A,B)$ for all $A, B \in \mathsf{NC}$,

$simi_a(\exists r.E, \exists s.F) := pm(r,s)[w + (1-w)(E \sim_d F)]$, and

$simi_a(C', D') := 0$ in any other case.

The CSiM $\sim$ induced by $\sim_d$ is obtained by comparing the concepts in both directions with $\sim_d$:

$$C \sim D := (C \sim_d D) \otimes (D \sim_d C), \tag{14}$$

where $\otimes$ is a fuzzy connector, i.e., a *commutative* binary operator $\otimes : [0,1] \times [0,1] \to [0,1]$ satisfying some additional properties.[11]

A non-default primitive measure allows us to state similarity degrees that may exist between pairs of different concepts or roles in the application domain. For example, one may think that the concepts Picture and Movie are somewhat similar, and thus provide the pair of them with a similarity degree > 0. The function g can be used to express that some atoms are more important than others, and thus should contribute more to the computation of the similarity degree. As for the factor w, it weights the contribution of $pm(r,s)$ and $E \sim_d F$ in the computation of $simi_a(\exists r.E, \exists s.F)$. In particular, having $w > 0$ prevents the value $pm(r,s)$ from being ignored in case $E \sim_d F = 0$. Note that, in the main part (13) of the above definition, we consider the reduced forms C^r and D^r of the concepts C and D when collecting the top-level atoms, basically for the same reason we did this in Example 3.1. In the two special cases, this is not needed since the conditions formulated there ($C \equiv \top$ as well as $C \not\equiv \top$ and $D \equiv \top$) are invariant under equivalence. Also note that $C \not\equiv \top$ implies that the denominator in (13) cannot be zero since then $top(C^r) \neq \emptyset$.

The CM $\sim_{\mathsf{su}}$ introduced in Example 3.1 cannot be obtained as an instance of Definition 3.15 since it compares existential restrictions in $top(C^r)$ and $top(D^r)$ not recursively with the same measure, but rather w.r.t. subsumption.

Example 3.16. The probably simplest instance of Definition 3.15 is $\sim_d^*$, where w=0, g assigns 1 to all atoms, and pm is the default primitive measure pm_{df} assigning value 1 when $A = B$ ($r = s$), and 0 otherwise. The directional measure $\sim_d^*$ can be turned into a CSiM $\sim^*$ by using $\otimes$=min in (14). In [11] it was shown that $m_{\sim^*} = deg$, where deg is the original graded membership function introduced and investigated in [11,19]. Given the $\mathcal{EL}$ concept descriptions

$$C := A \sqcap B_1 \sqcap \exists r.(A \sqcap \exists r.B \sqcap \exists s.A),$$

$$D := A \sqcap B_2 \sqcap \exists r.(A \sqcap \exists r.A \sqcap \exists s.B) \sqcap \exists r.\exists s.A,$$

[10] The definition of w in [14] excludes the value 0. However, all the properties shown in [14] to be satisfied by $\sim_d$ that are relevant to obtain our results also hold for $w = 0$. Regarding the function g, the original definition of $simi$ in [14] allows for more general functions g, where $g(\exists r.C) \neq g(\exists r.D)$ may hold for different atoms $\exists r.C$ and $\exists r.D$.

[11] In addition to being commutative, such fuzzy connectors are also required to be *equivalence closed, weakly monotone, bounded and grounded* in [14], but for our purposes it is enough to know commutativity.

we want to calculate $C \sim_d^* D$. Since C has three top-level atoms and g assigns 1 to all atoms, the denominator of the fraction in (13) is 3. For the numerator, we need to find, for each top-level atom in C, the top-level atom in D that matches it best. For A in C this is clearly A in D, and this contributes 1 to the sum in the numerator. For B_1 in C, all the top-level atoms in D yield the value 0, and thus B_1 contributes 0 to the sum. For the existential restriction $\exists r.(A \sqcap \exists r.B \sqcap \exists s.A)$ in C, the only matches with a possible non-zero value are the two existential restrictions in D.

First, consider $\exists r.(A \sqcap \exists r.A \sqcap \exists s.B)$. Since these existential restrictions are for the same role and $w = 0$, the value we obtain from this match is $A \sqcap \exists r.B \sqcap \exists s.A \sim_d^* A \sqcap \exists r.A \sqcap \exists s.B$. Again, we have three top-level atoms, but now A matches with A with value 1 whereas the existential restrictions cannot be matched with a value greater than 0. Thus, we obtain the overall value of $1/3$ for this attempt to match the existential restriction in C.

Second, consider $\exists r.\exists s.A$. Here, we need to calculate $A \sqcap \exists r.B \sqcap \exists s.A \sim_d^* \exists s.A$. Since $w = 0$ and we use the default primitive measure, it is easy to see that we obtain the value $1/3$.

Summing up, the overall value we obtain is $1 + 1/3$ divided by 3, i.e., we have $C \sim_d^* D = 4/9$. △

We now proceed to introduce the restricted family of CMs that we are interested in. It consists of all standard CMs corresponding to directional instances $\sim_d$ of *simi*. To identify these measures, we analyze $\sim_d$ w.r.t. each of the properties required for a standard CM. *Subsumption closedness* and *equivalence invariance* are satisfied by all instances $\sim_d$. The former was already shown in [14], whereas the latter is an easy consequence of the facts that $C \sim_d D$ is computed using the reduced forms of C and D and that, up to associativity and commutativity of $\sqcap$, equivalent $\mathcal{EL}$ concepts have the same reduced form. Showing *role-depth reduction* for $\sim_d$ is more involved. The following lemma, whose proof can be found in the Appendix A.1, states that this property is also satisfied for all instances $\sim_d$.

Lemma 3.17. *Let $\sim_d$ be a directional instance of simi. Then, $\sim_d$ is role-depth reducing.*

Regarding *signature reduction*, not all directional instances of *simi* satisfy this property. The following is a simple example illustrating this.

Example 3.18. Consider a directional instance $\sim_d$ of *simi* for which there is an $A \in \mathsf{NC}$ such that $pm(A, X) > 0$ for infinitely many $X \in \mathsf{NC}$. Let $\mathfrak{s}$ be an arbitrary function from $\mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$ to $\mathcal{P}(\mathsf{NC} \cup \mathsf{NR})$ such that $\mathfrak{s}(C)$ is finite for all $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$. Then, there exists $B \notin \mathfrak{s}(A)$ such that $pm(A, B) > 0$, and thus $A \sim_d B > 0$. The only concept description D' over a signature not containing B and strictly subsuming B is $\top$. Since $A \sim_d \top = 0$, this means that there is no concept D' such that $B \sqsubseteq D'$ and D' satisfies all conditions required for $\mathfrak{s}$ to be s-reducing for $\sim_d$. △

It turns out that finiteness of the set $\{X \mid pm(Y, X) > 0\}$ for all $Y \in \mathsf{NC} \cup \mathsf{NR}$ is not only necessary, but also sufficient for $\sim_d$ to be signature reducing. To formalize this, let $\mathfrak{s}_{pm} : \mathsf{NC} \cup \mathsf{NR} \to \mathcal{P}(\mathsf{NC} \cup \mathsf{NR})$ be the function $\mathfrak{s}_{pm}(A) := \{B \in \mathsf{NC} \mid pm(A, B) \neq 0\}$, for all $A \in \mathsf{NC}$; and analogously for NR. We say that pm is *reduced* iff $\mathfrak{s}_{pm}(X)$ is finite for all $X \in \mathsf{NC} \cup \mathsf{NR}$. The next lemma shows that $\sim_d$ is signature reducing if pm is reduced.

Lemma 3.19. *Let $\sim_d$ be a directional instance of simi such that pm is reduced. Then, $\sim_d$ is signature reducing.*

Proof. For all $C \in \mathcal{C}_{\mathcal{EL}}$, we define the function $\mathfrak{s}_{\sim_d}(C)$ as follows:

$$\mathfrak{s}_{\sim_d}(C) := \bigcup_{A \in \mathsf{NC} \cap \mathsf{sig}(C)} \mathfrak{s}_{pm}(A) \cup \bigcup_{r \in \mathsf{NR} \cap \mathsf{sig}(C)} \mathfrak{s}_{pm}(r).$$

Since pm is reduced, the sets $\mathfrak{s}_{\sim_d}(C)$ are finite. Let $C, D \in \mathcal{C}_{\mathcal{EL}}$ be such that $\mathsf{sig}(D) \not\subseteq \mathfrak{s}_{\sim_d}(C)$. Then, there exists $X \in \mathsf{NC} \cup \mathsf{NR}$ such that $X \in \mathsf{sig}(D)$ and $X \notin \mathfrak{s}_{\sim_d}(C)$. By definition of $\mathfrak{s}_{\sim_d}$, it follows that $pm(Y, X) = 0$ for all $Y \in \mathsf{sig}(C)$. Let now D^* be the concept description obtained from D by replacing X (or $\exists X.D'$ in case $X \in \mathsf{NR}$) with $\top$. We have that $D \sqsubseteq D^*$ and $C \sim_d D = C \sim_d D^*$ (by definition of *simi*). Thus, repeating this process until $\mathsf{sig}(D^*) \subseteq \mathfrak{s}_{\sim_d}(C)$ yields a concept description D^* satisfying the conditions required for $\sim_d$ to be signature reducing. □

We are now ready to define the restricted class of CMs *simi-d* that will be investigated in the remainder of this paper.

Definition 3.20. The class *simi-d* is obtained from the directional measures of *simi* by restricting the admissible parameters as follows:

- pm is reduced;
- g, pm and the function $\mathfrak{s}_{pm}$ are computable in polynomial time.

Since all CMs $\sim_d$ in *simi-d* are computable and standard, Theorem 3.14 yields the following decidability result for $\tau\mathcal{EL}(m_{\sim_d})$. In later sections we will investigate the exact complexity of reasoning in $\tau\mathcal{EL}(m_{\sim_d})$.

Theorem 3.21. *Every CM $\sim_d \in$ simi-d induces a well-defined and computable membership function $m_{\sim_d}$, and consequently a decidable threshold DL $\tau\mathcal{EL}(m_{\sim_d})$.*

3.4. Why consider only directional instances of *simi*?

In our previous work [11,17], we have actually considered membership functions induced by undirectional instances of *simi*. One may wonder why we restrict the attention here to instances of the directional part of *simi*. The rest of this section is devoted to explaining this. Recall that an undirectional instance of *simi* is obtained from a directional instance $\sim_d$ using the identity (14).

To start with, we show that there are instances $\sim$ of *simi* obtained from directional measures $\sim_d \in$ *simi-d* such that the induced membership function $m_\sim$ is not well-defined. The following lemma proves this, and in addition rectifies a statement in [17] where we wrongly claimed that all instances $\sim$ of *simi* where g assigns 1 to atoms of the form $\exists r.C$ induce a well-defined membership function $m_\sim$. The instances of *simi* considered in the lemma are defined using a *strictly monotone* fuzzy connector $\otimes$, i.e.,

$$x < y \implies x \otimes z < y \otimes z \text{ holds for all } x, y, z \in [0, 1].$$

The more general class of *monotone* fuzzy connectors is obtained by using $\leq$ in the previous implication.

Lemma 3.22. *Let $\otimes$ be a strictly monotone fuzzy connector that has 1 as a unit. Then, there exists $\sim_d \in$ simi-d that combined with $\otimes$ according to (14) yields a CSiM $\sim$ such that $m_\sim$ is not well-defined.*

Proof. Let $\sim_d$ be a directional instance of *simi* such that:

- g assigns 1 to all atoms,
- pm modifies pm_{df} by setting $pm(A, A') > 0$, $pm(A', A) > 0$ and $pm(s, r) > 0$, where $r, s \in \mathsf{NR}$ are two fixed distinct role names and $A, A' \in \mathsf{NC}$ are two fixed distinct concept names, and
- the values $pm(A', A)$ and $pm(s, r)$ and the factor w are chosen such that they satisfy[12]:

$$pm(A', A) < pm(s, r) \cdot w \qquad \text{and} \qquad pm(A', A) + 3 \cdot pm(s, r) \leq 4 \cdot pm(s, r) \cdot w. \tag{15}$$

Notice that there are only finitely many pairs of distinct X and Y such that $pm(X, Y) > 0$. This implies that pm is reduced, and hence $\sim_d \in$ *simi-d*.

Let now $\sim$ be the instance of *simi* obtained from $\sim_d$ and $\otimes$ using (14). Further, let $\mathcal{I}$ be an interpretation with domain $\Delta^{\mathcal{I}} = \{d, e, f\}$ that interprets s as $s^{\mathcal{I}} = \{(d, e), (e, f), (f, f)\}$, the concept names A' and B as $A'^{\mathcal{I}} = \{d\}$ and $B^{\mathcal{I}} = \{f\}$, and all other concept and role names as empty sets. The concepts in reduced form that d belongs to are $\top, A', \exists s.\top, A' \sqcap \exists s.\top$ and conjunctions of the form:

$$A' \sqcap \bigsqcap_{i=1}^{k} \exists s.\exists s.C_i \quad \text{or} \quad \bigsqcap_{i=1}^{k} \exists s.\exists s.C_i, \tag{16}$$

where $k \geq 1$ and the C_i are k subsumption incomparable concepts in $\mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$ such that $f \in C_i^{\mathcal{I}}$ for all $i, 1 \leq i \leq k$. Since f is an instance of all concepts of the form $\exists s \ldots \exists s.B$ and these concepts are incomparable w.r.t. subsumption, reduced concepts of the form shown in (16) exist for all $k \geq 1$.

We prove that $m_\sim$ is not well-defined by showing that $m_\sim^{\mathcal{I}}(d, A \sqcap \exists r.\top)$ is not well-defined, i.e., the maximum in Definition 3.2 does not exist. The first step is to prove that $a_k := A \sqcap \exists r.\top \sim Z^k$ is a strictly increasing sequence, where Z^k is a concept of the form (16) with A' and k existential restrictions as its top-level atoms. The value a_k for a fixed $k \in \mathbb{N}$ is actually:

$$a_k = (A \sqcap \exists r.\top \sim_d Z^k) \otimes (Z^k \sim_d A \sqcap \exists r.\top) = \frac{pm(A, A')}{2} \otimes \frac{pm(A', A) + k \cdot pm(s, r) \cdot w}{k + 1}.$$

Regarding the left argument of $\otimes$ in the above term, note that, though we have required $pm(s, r) > 0$, we still have $pm(r, s) = 0$. As for the right argument, using the first inequality in (15), we obtain the following inequality:

$$\begin{aligned}
&(k + 2) \cdot (pm(A', A) + k \cdot pm(s, r) \cdot w) \ = \\
&(pm(A', A) + k \cdot pm(s, r) \cdot w) + (k + 1) \cdot (pm(A', A) + k \cdot pm(s, r) \cdot w) \ < \\
&(k + 1) \cdot pm(s, r) \cdot w + (k + 1) \cdot (pm(A', A) + k \cdot pm(s, r) \cdot w) \ = \\
&(k + 1) \cdot (pm(A', A) + (k + 1) \cdot pm(s, r) \cdot w).
\end{aligned}$$

Dividing the first and the last term in this sequence by $(k + 1) \cdot (k + 2)$ thus shows that

$$Z^k \sim_d A \sqcap \exists r.\top < Z^{k+1} \sim_d A \sqcap \exists r.\top.$$

Since the fuzzy connector $\otimes$ is assumed to be strictly monotone, this implies that $a_k < a_{k+1}$.

The second step is to compare this sequence with all other possible similarity values. The similarity values of $A \sqcap \exists r.\top$ with the first four reduced concepts mentioned above and concepts of the form (16) where A' does not occur, correspond to the following expressions (in order):

[12] For example, we can take $pm(A', A) = 1/5, pm(s, r) = 9/10, w = 7/8$.

$$v_1 := 0 \otimes 1, \qquad v_2 := \frac{pm(A, A')}{2} \otimes pm(A', A), \qquad v_3 := 0 \otimes pm(s, r),$$
$$v_4 := \frac{pm(A, A')}{2} \otimes \frac{pm(A', A) + pm(s, r)}{2}, \qquad v_5 := 0 \otimes (pm(s, r) \cdot w).$$

Hence, if the values $v_1, \dots, v_5$ are bounded by some a_k then $m_\sim^{\mathcal{I}}(d, A \sqcap \exists r.\top)$ is undefined, since $\max\{a_k \mid k \geq 1\}$ does not exist. To see that this is indeed the case, notice that monotonicity of $\otimes$ and $pm(A', A) < pm(s, r)$ imply that $v_3, v_5 \leq v_1$ and $v_2 \leq v_4$. Since $\otimes$ has 1 as a unit, it then follows that $v_3 = v_5 = v_1 = 0 \leq a_1$. As for v_4, by applying the cross-product between its right argument and the expression for $Z^2 \sim_d A \sqcap \exists r.\top$, we obtain the expressions:

$$3 \cdot (pm(A', A) + pm(s, r)) \qquad \text{and} \qquad 2 \cdot (pm(A', A) + 2 \cdot pm(s, r) \cdot w).$$

The second inequality in (15) yields that $3 \cdot (pm(A', A) + pm(s, r)) \leq 2 \cdot (pm(A', A) + 2 \cdot pm(s, r) \cdot w)$. This implies that $\frac{pm(A', A)+pm(s,r)}{2} \leq Z^2 \sim_d A \sqcap \exists r.\top$. Thus, monotonicity of $\otimes$ implies that $v_2 \leq v_4 \leq a_2$. $\square$

Functions $\otimes$ satisfying the hypothesis of the lemma are, for example, all strictly monotone *bounded t-norms*. In addition, by looking at the proof of the lemma, many other natural functions $\otimes$ can be identified that may cause $m_\sim$ to be not well-defined. For instance, while strict monotonicity is needed to ensure $a_k < a_{k+1}$, it is always applied w.r.t. the fix value $pm(A, A')/2 > 0$. Hence, monotone operators like *product* or *Dice's connector* [30] where strict monotonicity fails to hold only for 0 also yield $a_k < a_{k+1}$. Further, for *average* and min, it is not hard to see how to tune the values of pm to make $m_\sim^{\mathcal{I}}(d, A \sqcap \exists r.\top)$ undefined.

Overall, we have seen that many fuzzy connectors $\otimes$ in combination with a non-default primitive measure may cause $m_\sim$ to be not well-defined. Hence, the natural solution to this problem could be to restrict the attention to undirectional instances of *simi* where $pm = pm_{df}$. One class containing many such measures, called ***simi-mon***, was introduced in [17], albeit with a different motivation. The aim was to obtain a large family of graded membership functions containing *deg* (the graded membership function first introduced and investigated in [11]), such that the complexity of reasoning in the induced threshold DLs $\tau\mathcal{EL}(m_\sim)$ is the same as in $\tau\mathcal{EL}(deg)$.

Definition 3.23. The class *simi-mon* is obtained from the undirectional measures of *simi* by restricting the admissible parameters in the following way:

- $\otimes$ is computable in polynomial time and monotone w.r.t. $\geq$;
- g is computable in polynomial time;
- $pm = pm_{df}$.

Since the default measure clearly satisfies the conditions required by Definition 3.20, the directional measure $\sim_d$ associated with an element $\sim$ of *simi-mon* belongs to *simi-d*.

The definition of *simi-mon* given above is more general than the one provided in [17], where g was restricted to assign 1 to all atoms of the form $\exists r.C$. In fact, well-definedness of $m_\sim$ for all $\sim \in$ *simi-mon* was shown in [17] using the claim mentioned above (and disproved by Lemma 3.22) that all undirectional instances $\sim$ of *simi* where g assigns 1 to atoms of the form $\exists r.C$ induce a well-defined membership function $m_\sim$. In Appendix A.2 we provide a correct proof of well-definedness of $m_\sim$ for the extended version of *simi-mon* introduced in Definition 3.23.

In the present paper, we do not investigate the complexity of threshold logics $\tau\mathcal{EL}(m_\sim)$ for $\sim \in$ *simi-mon* explicitly. The reason is that for most instances $\sim$ of *simi-mon*, either $m_\sim$ is equal to some $m_{\sim_d}$ with $\sim_d \in$ *simi-d* or reasoning in $\tau\mathcal{EL}(m_\sim)$ can be reduced to reasoning in $\tau\mathcal{EL}(m_{\sim_d})$. This is a consequence of the following proposition, whose proof can be found in Appendix A.2.

Proposition 3.24. *Let $\sim \in$ simi-mon and $\sim_d$ be the associated directional measure. For all $\mathcal{EL}$ concept descriptions C, interpretations $\mathcal{I}$ and $e \in \Delta^{\mathcal{I}}$, there exists an $\mathcal{EL}$ concept description D such that $e \in D^{\mathcal{I}}$ and D satisfies the following:*

- $m_\sim^{\mathcal{I}}(e, C) = (C \sim_d D) \otimes 1$,
- $m_{\sim_d}^{\mathcal{I}}(e, C) = C \sim_d D$.

An easy consequence of this proposition is the following result (as special case, it follows that $m_{\sim_d^*} = deg$).

Corollary 3.25. *Let $\sim \in$ simi-mon and $\sim_d$ be the associated directional measure. If $\otimes$ is a t-norm, then $m_\sim = m_{\sim_d}$, and thus the logics $\tau\mathcal{EL}(m_\sim)$ and $\tau\mathcal{EL}(m_{\sim_d})$ coincide.*

Proof. Since $\otimes$ is a t-norm, the number 1 is a unit for it. Hence, Proposition 3.24 implies that $m_\sim = m_{\sim_d}$. $\square$

A second consequence of Proposition 3.24 is that, whenever $f(x) = x \otimes 1$ is strictly monotone, reasoning in $\tau\mathcal{EL}(m_\sim)$ can be reduced in polynomial time to reasoning in $\tau\mathcal{EL}(m_{\sim_d})$. This is shown in Lemma 3.29 with the help of an auxiliary result, namely, that $C \sim_d D$ is a rational number y/x_C, where the value x_C does not depend on D and can be computed in polynomial time in $s(C)$.

To show this, we prove the result for all $\sim_d \in$ *simi-d*. This more general result will be useful in later sections to obtain our complexity results.

Let us start by introducing some notation. In the computation of $C \sim_d D$, the denominator x_C is partly determined by the denominators of fractions p/q that correspond to the values of $pm(A, B)$ and $pm(r, s)$ in the computation of $simi_a$. To capture all such possible values q, given a primitive measure pm, we define a function d_{pm} that assigns to every $\mathcal{EL}$ atom At the following set:

$$d_{pm}(\mathsf{At}) := \begin{cases} \{q \mid pm(\mathsf{At}, B) = p/q \wedge B \in \mathsf{NC}\} & \text{if At} \in \mathsf{NC} \\ \{q \mid pm(r, s) = p/q \wedge s \in \mathsf{NR}\} & \text{if At is of the form } \exists r.C. \end{cases}$$

This function is extended to all $C \in \mathcal{C}_{\mathcal{EL}}$ by considering all the atoms occurring somewhere in C:

$$d_{pm}(C) := \{q \mid q \in d_{pm}(\mathsf{At}) \wedge \mathsf{At} \text{ is an atom occurring in } C\}.$$

We also define the set $\mathcal{S}_\sqcap(C)$ of maximal conjunctions in C as:

$$\mathcal{S}_\sqcap(C) := \{C\} \cup \bigcup_{\exists r.F \in top(C)} \mathcal{S}_\sqcap(F).$$

In Example 2.3, we have $\mathcal{S}_\sqcap(C) = \{C, A \sqcap B \sqcap \exists r.\top, A, \top\}$. Further, given a weighting function g and an $\mathcal{EL}$ atom At, we denote as $g_{\mathsf{At},1}/g_{\mathsf{At},2}$ the rational number $g(\mathsf{At})$. Based on this notation, the sum in the denominator of $C \sim_d D$ in (13) can be expressed as:

$$\sum_{\mathsf{At} \in top(C^r)} \frac{g_{\mathsf{At},1}}{g_{\mathsf{At},2}} = \frac{\sum\limits_{\mathsf{At} \in top(C^r)} g_{\mathsf{At},1} \cdot \prod\limits_{\mathsf{At}' \in top(C^r) \setminus \{\mathsf{At}\}} g_{\mathsf{At}',2}}{\prod\limits_{\mathsf{At} \in top(C^r)} g_{\mathsf{At},2}}. \tag{17}$$

In the rest of this section, we use g_C to denote the value of the numerator in (17). In addition, we always represent 0 with the fraction 0/1.

The following lemma states the form of x_C (see its proof in Appendix A.2).

Lemma 3.26. *Let $\sim_d \in$ simi-d and let the parameter w be of the form w_n/w_d. In addition, let $C \in \mathcal{C}_{\mathcal{EL}}$, let $k = \mathsf{rd}(C)$, and define*

$$x_C := w_d^k \cdot \prod_{q \in d_{pm}(C)} q^k \cdot \prod_{F \in \mathcal{S}_\sqcap(C)} g_F^k.$$

Then, for all $D \in \mathcal{C}_{\mathcal{EL}}$, there exists $y \in \mathbb{N}$ such that $C \sim_d D = y/x_C$.

Next, we analyze the complexity of computing x_C.

Lemma 3.27. *The number x_C defined in Lemma 3.26 can be computed in time polynomial in* $\mathsf{s}(C)$.

Proof. Let us start by looking at the sets $d_{pm}(C)$ and $\mathcal{S}_\sqcap(C)$.

- Since $\sim_d \in$ *simi-d*, the function $\mathfrak{s}_{pm}$ is computable in polynomial time. This implies that only polynomially many concept and role names can yield a denominator q different from 1. Thus, the cardinality of $d_{pm}(C)$ is polynomial in $\mathsf{s}(C)$.
- By definition of $\mathcal{S}_\sqcap(C)$, it is clear that this set contains at most $\mathsf{s}(C)$ conjunctions.

Now, x_C can be rewritten as z^k, where

$$z := w_d \cdot \prod_{q \in d_{pm}(C)} q \cdot \prod_{F \in \mathcal{S}_\sqcap(C)} g_F.$$

By definition of g_F in (17), for each $F \in \mathcal{S}_\sqcap(C)$ one can compute g_F by executing at most $\mathsf{s}(C) - 1$ sums and $\mathsf{s}(C)^2 - \mathsf{s}(C)$ multiplications. Therefore, since the cardinalities of $d_{pm}(C)$ and $\mathcal{S}_\sqcap(C)$ are both polynomial in $\mathsf{s}(C)$, to compute z, it suffices to perform a number m (polynomial in $\mathsf{s}(C)$) of operations (multiplication or addition). Hence, z^k can be computed with at most $m + \mathsf{s}(C) - 1$ operations, since $k = \mathsf{rd}(C) \leq \mathsf{s}(C)$.

The multiplication (sum) of two numbers n_1 and n_2 can be done in time polynomial in the number of bits b_1 and b_2 of their binary representations. In addition, the size in bits of $n_1 \cdot n_2$ $(n_1 + n_2)$ is at most $b_1 + b_2$ $(\max\{b_1, b_2\} + 1)$. Since g and pm are computable in polynomial time, the numbers $q, g_{\mathsf{At},1}$ and $g_{\mathsf{At},2}$ involved in the computation of z have size polynomial in $\mathsf{s}(C)$. Thus, since w_d is a constant, it follows that $x_C = z^k$ is computable in time polynomial in $\mathsf{s}(C)$. □

Coming back to measures $\sim \in$ *simi-mon*, all sets $d_{pm}(\mathsf{At})$ are equal to $\{1\}$ since $pm = pm_{df}$. This implies that $d_{pm}(C) = \{1\}$ as well, for all $C \in \mathcal{C}_{\mathcal{EL}}$. Hence, the following is a direct consequence of Lemma 3.26.

Lemma 3.28. *Let $\sim \in$ simi-mon. In addition, let $C \in \mathcal{C}_{\mathcal{EL}}$, let $k = \mathsf{rd}(C)$, and define*

$$x_C := w_d^k \cdot \prod_{F \in \mathcal{S}_\sqcap(C)} g_F^k.$$

Then, for all $D \in \mathcal{C}_{\mathcal{EL}}$, there exists $y \in \mathbb{N}$ such that $C \sim_d D = y/x_C$.

We are now ready to prove a second important consequence of Proposition 3.24.

Lemma 3.29. *Let $\sim \in$ simi-mon and $\sim_d$ be the associated directional measure. If $f(x) = x \otimes 1$ is strictly monotone, then reasoning in $\tau\mathcal{EL}(m_\sim)$ can be reduced in polynomial time to reasoning in $\tau\mathcal{EL}(m_{\sim_d})$.*

Proof. To prove our claim, we show that, given a threshold concept $C_{\bowtie t}$, one can compute in polynomial time (in the size of $C_{\bowtie t}$) an operator $\bowtie_c \in \{<, \leq, >, \geq\}$ and a value $t_c \in \mathbb{Q} \cap [0, 1]$ such that

$$\text{for all interpretations } \mathcal{I} \text{ and all } e \in \Delta^\mathcal{I}: \quad m_\sim^\mathcal{I}(e, C) \bowtie t \text{ iff } m_{\sim_d}^\mathcal{I}(e, C) \bowtie_c t_c. \tag{18}$$

This implies that $e \in [C_{\bowtie t}]^\mathcal{I}$ in $\tau\mathcal{EL}(m_\sim)$ iff $e \in [C_{\bowtie_c t_c}]^\mathcal{I}$ in $\tau\mathcal{EL}(m_{\sim_d})$. Thus, reasoning in $\tau\mathcal{EL}(m_\sim)$ can then be reduced to reasoning in $\tau\mathcal{EL}(m_{\sim_d})$ by replacing each threshold concept $C_{\bowtie t}$ in a $\tau\mathcal{EL}(m_\sim)$ concept description with $C_{\bowtie_c t_c}$.

By Proposition 3.24 and the definition of $m_\sim$ (see Definition 3.2), the equivalence (18) can be expressed as follows:

$$\text{for all interpretations } \mathcal{I} \text{ and } e \in \Delta^\mathcal{I}: \quad \mu_e \otimes 1 \bowtie t \text{ iff } \mu_e \bowtie_c t_c, \text{ where } \mu_e = \max\{C \sim_d X \mid X \in \mathcal{C}_{\mathcal{EL}} \wedge e \in X^\mathcal{I}\}. \tag{19}$$

We have shown in Lemma 3.28 that, for a given $\mathcal{EL}$ concept C, all possible values μ_e in (19) are numbers of the form y/x_C where $0 \leq y \leq x_C$ and $y, x_C \in \mathbb{N}$. Let $\mathcal{U}_C$ denote the set of all such numbers y/x_C. We now use strict monotonicity of f to identify a value $t_c \in \mathcal{U}_C$ and an operator $\bowtie_c$ satisfying (19). To this end, we make the following case distinction:

1. $x_0 \otimes 1 > t$ for all $x_0 \in \mathcal{U}_C$. We define $t_c := 0$, $\bowtie_c := <$ if $\bowtie \in \{<, \leq\}$, and $\bowtie_c := \geq$ if $\bowtie \in \{>, \geq\}$. In this case, $\mu_e \otimes 1 \bowtie t$ for $\bowtie \in \{<, \leq\}$ is false, independently of the element of $\mathcal{U}_C$ the number μ_e actually is, as is $\mu_e < 0$. For $\bowtie \in \{>, \geq\}$, $\mu_e \otimes 1 \bowtie t$ is true, as is $\mu_e \geq 0$. Thus, the equivalence (19) holds.
2. $x_0 \otimes 1 = t$ for some $x_0 \in \mathcal{U}_C$. In this case, we can define $t_c := x_0$ and $\bowtie_c := \bowtie$. By strict monotonicity of $f(x) = x \otimes 1$ the following holds for all $z \in \mathcal{U}_C$:

   $$z \otimes 1 < t \text{ iff } z < t_c \quad \text{and} \quad z \otimes 1 = t \text{ iff } z = t_c.$$

   Hence, since each μ_e from (19) belongs to $\mathcal{U}_C$, it is easy to see that (19) holds.
3. $x_0 \otimes 1 < t$ for some $x_0 \in \mathcal{U}_C$ and $z \otimes 1 \neq t$ for all $z \in \mathcal{U}_C$. We select t_c as the greatest number $x \in \mathcal{U}_C$ satisfying $x \otimes 1 < t$. This number exists since $\mathcal{U}_C$ is a finite set. As for $\bowtie_c$, we define $\bowtie_c := \leq$ if $\bowtie \in \{<, \leq\}$ and $\bowtie_c := >$ if $\bowtie \in \{>, \geq\}$.

   To see that this is a correct choice, consider first the case where $\bowtie = <$. For all $z \in \mathcal{U}_C$, if $z \otimes 1 < t$, then the selection of t_c and strict monotonicity of f imply that $z \leq t_c$. Conversely, $z \leq t_c$ implies $z \otimes 1 \leq t_c \otimes 1$ and the choice of t_c yields $t_c \otimes 1 < t$. Secondly, for $\bowtie = \leq$ we have that $z \otimes 1 \leq t$ implies that $z \otimes 1 < t$ since we assumed that $z \otimes 1 \neq t$ for all $z \in \mathcal{U}_C$. Hence, we obtain $z \leq t_c$ as in the previous case. Conversely, as in the previous case, $z \leq t_c$ implies $z \otimes 1 \leq t_c \otimes 1$, which in turn implies $z \otimes 1 \leq t$. Thirdly, if $\bowtie = >$, then $z \otimes 1 > t$ together with $t > t_c \otimes 1$ and strict monotonicity of f implies $z > t_c$. Conversely, $z > t_c$ yields $z \otimes 1 \geq t$ by the choice of t_c, and thus $z \otimes 1 > t$ due to the case we are in. Finally, for $\bowtie = \geq$, we know that $z \otimes 1 \geq t$ implies $z \otimes 1 > t$, and thus $z > t_c$ as in the previous case. The converse direction can also be handled as in the previous case. Summing up, since each μ_e from (19) belongs to $\mathcal{U}_C$, we have shown that also in this case (19) holds.

Based on these cases and using strict monotonicity of f, the values t_c and $\bowtie_c$ can be computed as follows:

- Check whether $0 \otimes 1 > t$ holds. If this is the case, then $x \otimes 1 > t$ holds for all $x \in \mathcal{U}_C$, and thus t_c and $\bowtie_c$ can be selected as described in the first case.
- If $0 \otimes 1 \leq t$, do a binary search over $\mathcal{U}_C$ to find the greatest $x \in \mathcal{U}_C$ such that $x \otimes 1 \leq t$. Select t_c and $\bowtie_c$ as defined in the second or third case above, depending on whether $x \otimes 1 = t$ holds or not.

A binary search over $\mathcal{U}_C$ takes $\log|\mathcal{U}_C|$ steps. Since $\mathcal{U}_C$ consists of x_C numbers and the value of the number x_C is at most exponential in $\mathsf{s}(C)$, the number of steps taken by the search is polynomial in $\mathsf{s}(C)$. In addition, each step of this search requires:

- To keep an interval in $\mathcal{U}_C$ with bounds y_1/x_C and y_2/x_C, where $y_1, y_2 \in \mathbb{N}$ and $y_1, y_2 \leq x_C$. By Lemma 3.27, we know that x_C can be computed in polynomial time in $\mathsf{s}(C)$. Hence, these intervals can be represented using polynomially many bits and computed in polynomial time.
- To compute $(y/x_C) \otimes 1$ and compare the result with t. Since y and x_C are of size polynomial in $\mathsf{s}(C)$ and $\otimes$ is computable in polynomial time, this means that computing $(y/x_C) \otimes 1$ and comparing the result with t requires polynomial time in $\mathsf{s}(C)$ and the size of t.

Overall, we have thus shown that a value t_c and an operator $\bowtie_c$ satisfying (18) can be computed in polynomial time. This shows that the desired polynomial time reduction function exists. $\square$

To summarize, we have introduced a class *simi-d* of standard CMs $\sim_d$ that can be obtained by instantiating the directional part of *simi*. These CMs induce the family of threshold DLs $\tau\mathcal{EL}(m_{\sim_d})$ that will be investigated in the next two sections. The restriction to using only directional instances of *simi* is justified since, on the one hand, it is not clear how to identify an interesting fragment of *simi* that uses a non-default primitive measure and induces well-defined membership functions. On the other hand, restricting to the default primitive measure produces (in many cases) threshold DLs that are equal or can be simulated by threshold DLs induced by measures in *simi-d*.

4. Reasoning with acyclic $\tau\mathcal{EL}(m_\sim)$ TBoxes

This section is devoted to investigating the complexity of reasoning in $\tau\mathcal{EL}(m_\sim)$ w.r.t. acyclic TBoxes, where $\sim\in$ *simi-d*. Since from now on we only consider directional measures from *simi-d*, we dispense with the index d and write $\sim$ rather than $\sim_d$.

Besides satisfiability and subsumption, we also consider consistency and instance checking for knowledge bases consisting of an (acyclic) TBox and an ABox. In the presence of a $\tau\mathcal{EL}(m)$ TBox, the concepts occurring in the ABox need to be correctly defined w.r.t. this TBox (see Definition 2.21). To formalize this, we introduce the notion of a $\tau\mathcal{EL}(m)$ *knowledge base* (KB). A $\tau\mathcal{EL}(m)$ KB is a pair $\mathcal{K}=(\mathfrak{T},\mathcal{A})$, where $\mathfrak{T}=(\widehat{\mathcal{T}},\mathcal{T})$ is a $\tau\mathcal{EL}(m)$ TBox and for all $\widehat{C}(a)\in\mathcal{A}$ the concept $\widehat{C}$ is correctly defined w.r.t. $\mathcal{T}$. This KB is acyclic if $\mathfrak{T}$ is an acyclic $\tau\mathcal{EL}(m)$ TBox.

The size of an acyclic KB $\mathcal{K}$ is defined as $\mathsf{s}(\mathcal{K}):=\mathsf{s}(\mathfrak{T})+\mathsf{s}(\mathcal{A})$, where $\mathsf{s}(\mathfrak{T})$ and $s(\mathcal{A})$ are defined as follows:

$$\mathsf{s}(\mathcal{A}):=\sum_{\widehat{C}(a)\in\mathcal{A}}\mathsf{s}(\widehat{C})+|\{r(a,b)\mid r(a,b)\in\mathcal{A}\}|,\qquad \mathsf{s}(\mathfrak{T}):=|\mathsf{NC}_d^{\mathfrak{T}}|+\sum_{\alpha\doteq\widehat{C}_\alpha\in\widehat{\mathcal{T}}\cup\mathcal{T}}\mathsf{s}(\widehat{C}_\alpha).$$

Together with the size of concepts, which has been defined before, this determines how we measure the size of the input in our complexity results.

For the instance checking problem, we consider the usual measures of data and combined complexity. The *combined complexity* is defined by considering $\mathfrak{T}$, $\mathcal{A}$ and the query concept $\widehat{C}$ as the inputs. The complexity results for instance checking given in Section 4.2.4 (upper bound) and Section 4.3.2 (lower bound) are concerned with combined complexity. In Section 4.4, we consider the *data complexity* of instance checking, which takes only the ABox $\mathcal{A}$ as input, and assumes that $\mathfrak{T}$ and the query concept are fixed parameters. In the presence of an acyclic TBox, we restrict the attention as usual (see, e.g., [31]) to *simple* ABoxes.[13] Without TBox, simple ABoxes are not very interesting. Thus, in this setting we allow for complex concept assertions in the ABox, similar to the setting considered in [32,33] for the DL $\mathcal{ALE}$. Our results on data complexity can be found in Section 4.4.

4.1. Contribution of this section and earlier results

The main contribution of this section is to investigate the computational complexity of reasoning in $\tau\mathcal{EL}(m_\sim)$ w.r.t. the empty TBox and w.r.t. non-empty acyclic $\tau\mathcal{EL}(m)$ TBoxes. In the following, we detail the corresponding results, which are shown in the subsequent subsections.

For the case of the empty TBox, we prove that satisfiability/non-subsumption and consistency/non-instance (w.r.t. *data complexity*) in $\tau\mathcal{EL}(m_\sim)$ are in NP for all CMs in *simi-d*, and that NP-hardness holds for such measures if they are defined using $pm=pm_{df}$. For the non-instance problem, we show that NP-hardness holds for all $\sim\in$ *simi-d*. In addition, we show that the non-instance problem in $\tau\mathcal{EL}(m_\sim)$ is PSpace-complete in combined complexity for all $\sim\in$ *simi-d*.

With respect to non-empty acyclic TBoxes, Proposition 2.29 together with the equality in Lemma 2.28 tells us that reasoning w.r.t. acyclic $\tau\mathcal{EL}(m)$ TBoxes can be reduced to reasoning w.r.t. the empty terminology, through unfolding. However, as shown by Nebel in [34] for the DL $\mathcal{FL}_0$, unfolding may produce concept descriptions of exponential size. The following is an adaptation of Nebel's example to $\mathcal{EL}$.

Example 4.1. The TBox $\mathcal{T}_n$ is inductively defined as follows ($n\geq 0$):

$$\mathcal{T}_0:=\{\alpha_0\doteq\top\}$$

$$\mathcal{T}_1:=\mathcal{T}_0\cup\{\alpha_1\doteq\exists r.\alpha_0\sqcap\exists s.\alpha_0\}$$

$$\ldots$$

$$\mathcal{T}_n:=\mathcal{T}_{n-1}\cup\{\alpha_n\doteq\exists r.\alpha_{n-1}\sqcap\exists s.\alpha_{n-1}\}$$

It is easy to see that $\mathsf{s}(\mathcal{T}_n)=\Theta(n)$, but $\mathsf{s}(u_{\mathcal{T}_n}(\alpha_n))\geq 2^n$. $\triangle$

[13] An ABox is called *simple* if its concept assertions are of the form $A(a)$ with $A\in\mathsf{NC}\cup\mathsf{ND}$.

Therefore, by applying unfolding and then using the NP decision procedures for satisfiability/non-subsumption and consistency/non-instance that we will provide for the empty TBox case, we obtain NExpTime algorithms to decide the same problems w.r.t. acyclic $\tau\mathcal{EL}(m_\sim)$ TBoxes, for all $\sim \in$ *simi-d*. The natural question to ask is thus: can we do better than NExpTime?

We positively answer this question, by showing that reasoning in $\tau\mathcal{EL}(m_\sim)$ w.r.t. acyclic TBoxes is in PSpace for all $\sim \in$ *simi-d*. Regarding lower bounds, we show PSpace-hardness for all $\sim \in$ *simi-d* defined using pm_{df}. Furthermore, we provide some particular results for the non-instance problem. First, we show that this problem is PSpace-complete in combined complexity for all $\sim \in$ *simi-d*, where PSpace-hardness already holds for the empty TBox. Second, regarding data complexity, we prove that the problem is NP-complete in the presence of acyclic TBoxes for all $\sim \in$ *simi-d*.

4.1.1. Earlier results

Preliminary results on the complexity of reasoning in extensions of $\mathcal{EL}$ by threshold concepts have already been obtained in [11,19,17]. For the case of the empty TBox, it is shown in [11] that satisfiability/non-subsumption and consistency/non-instance (w.r.t. *data complexity*) are NP-complete for $m_{\sim_d^*} = deg$. These results were later extended to all logics $\tau\mathcal{EL}(m_\sim)$ induced by CSiMs in *simi-mon*, where NP-hardness is shown only for a subclass of *simi-mon* [17]. With our new results for reasoning w.r.t. the empty TBox, we extend these earlier results into a much larger and more diverse family of threshold logics.

Regarding non-empty acyclic TBoxes, we showed in [19] that reasoning in $\tau\mathcal{EL}(deg)$ w.r.t. acyclic TBoxes is in PSpace. However, in [19] we were only able to prove Π_2^P/Σ_2^P lower bounds in this case. Our new contributions improve these results in two main ways. On the one hand, the PSpace upper bound carries over to our larger family of logics $\tau\mathcal{EL}(m_\sim)$ for all $\sim \in$ *simi-d*. Instead of adapting the existing algorithms for $\tau\mathcal{EL}(deg)$, we design new and simpler PSpace decision procedures. On the other hand, the PSpace-hardness results strengthen the Π_2^P/Σ_2^P-hardness result from [19], since deg corresponds to $m_{\sim_d^*}$ and $\sim_d^* \in$ *simi-d*. This closes the gap left open in [19] about the computational complexity of reasoning in $\tau\mathcal{EL}(deg)$ w.r.t. acyclic TBoxes. Furthermore, PSpace-hardness in combined complexity of non-instance for the empty TBox closes a gap left open in [11] for $\tau\mathcal{EL}(deg)$, where non-instance was only shown to be NP-hard.

4.2. Upper bounds

We start by devising algorithms that decide satisfiability of concepts of the form $\widehat{C} \sqcap \neg\widehat{D}$, first for the case without TBox and then w.r.t. an acyclic $\tau\mathcal{EL}(m_\sim)$ TBox. These algorithms can then be used to decide the non-subsumption and the satisfiability problem. In fact, we have $\widehat{C} \not\sqsubseteq \widehat{D}$ iff $\widehat{C} \sqcap \neg\widehat{D}$ is satisfiable, and $\widehat{C}$ is satisfiable iff $\widehat{C} \sqcap \neg(A \sqcap A_{<1})$ is satisfiable. Afterwards, these decision procedures are extended to tackle also reasoning w.r.t. acyclic KBs.

The main technical result on which these decision procedures are based is that satisfiable conjunctions $\widehat{C} \sqcap \neg\widehat{D}$ enjoy a particular bounded tree model property. We continue by showing this property in two main steps. First, we provide a *polynomially bounded tree model* property for the case where $\mathfrak{T} = \emptyset$, which will allow us to obtain a guess-and-check NP-decision procedure for satisfiability of $\widehat{C} \sqcap \neg\widehat{D}$. Besides the bound on the size of the model, we also give a polynomial bound for its depth. In a second step, we use unfolding to extend the bounded tree model property to arbitrary acyclic TBoxes. As we will see, unfolding does not incur an exponential blow-up in the additional bound. This is then exploited to come up with a PSpace decision procedure that covers arbitrary acyclic TBoxes.

4.2.1. Bounded tree model property

To state our tree model property, we first need to formalize the notion of the depth of a tree-shaped interpretation $\mathcal{I}$:

- The *depth* of $\mathcal{I}$ is the greatest $n \geq 0$ such that there is a sequence of the form $d_0 r_1 d_1 \ldots r_n d_n$ satisfying $(d_i, d_{i+1}) \in r_{i+1}^{\mathcal{I}}$ for all $i, 0 \leq i < n$.

The next lemma, whose proof is given in Appendix B, formally states our bounded tree model property.

Lemma 4.2. *Let $\sim \in$ simi-d. In addition, let $\widehat{C}$ and $\widehat{D}$ be two $\tau\mathcal{EL}(m_\sim)$ concept descriptions. If $\widehat{C} \sqcap \neg\widehat{D}$ is satisfiable in $\tau\mathcal{EL}(m_\sim)$, then there is a tree-shaped interpretation $\mathcal{I}$ with root element d_0 such that:*

1. $d_0 \in (\widehat{C} \sqcap \neg\widehat{D})^{\mathcal{I}}$, $|\Delta^{\mathcal{I}}| \leq \mathsf{s}(\widehat{C}) \cdot \mathsf{s}(\widehat{D})$*, and*
2. *$\mathcal{I}$ has depth at most* $\mathsf{rd}(\widehat{C} \sqcap \widehat{D})$.

The idea is now to test satisfiability of $\widehat{C} \sqcap \neg\widehat{D}$ by guessing a tree-shaped interpretation $\mathcal{I}$ of cardinality at most $\mathsf{s}(\widehat{C}) \cdot \mathsf{s}(\widehat{D})$. However, differently from the usual case in traditional DLs, we cannot assume that $\mathcal{I}$ interprets symbols outside of $\mathsf{sig}(\widehat{C}) \cup \mathsf{sig}(\widehat{D})$ as empty sets. Basically, since *simi-d* allows the use of non-default primitive measures, the interpretation of concept and role names not occurring in $\widehat{C}$ and $\widehat{D}$ may also be relevant when computing the interpretation of the concept $\widehat{C} \sqcap \neg\widehat{D}$. For instance, consider the concept $B_{\geq .8} \sqcap B_{\leq .8}$. As $pm(B, B) = 1$, an individual d can only belong to this concept if it belongs to another concept name, say A, such that $pm(B, A) = 0.8$.

However, we can use the fact that the elements of *simi-d* are signature reducing to show that we can restrict the attention to a finite part of the signature. More precisely, we have the following lemma, whose proof can be found in Appendix B.

Lemma 4.3. *Let $\sim \in$ simi-d and $\widehat{C}, \widehat{D}$ be $\tau\mathcal{EL}(m_\sim)$ concepts such that the conjunction $\widehat{C} \sqcap \neg\widehat{D}$ is satisfiable according to the semantics of $\tau\mathcal{EL}(m_\sim)$ extended with negation. Then, $\widehat{C} \sqcap \neg\widehat{D}$ has a model $\mathcal{I}$ such that $\mathcal{I}$ is as in Lemma 4.2 and only interprets concept names in $\mathfrak{C}(\widehat{C} \sqcap \neg\widehat{D}, \sim) := \{B \in \mathsf{NC} \mid B \in \mathfrak{s}_{pm}(A) \wedge A \in \mathsf{sig}(\widehat{C} \sqcap \neg\widehat{D})\}$ and role names in $\mathfrak{R}(\widehat{C} \sqcap \neg\widehat{D}, \sim) := \{s \in \mathsf{NR} \mid s \in \mathfrak{s}_{pm}(r) \wedge r \in \mathsf{sig}(\widehat{C} \sqcap \neg\widehat{D})\}$ as non-empty.*

These two lemmas yield a straightforward guess-and-check NP-algorithm to decide satisfiability of concepts of the form $\widehat{C} \sqcap \neg\widehat{D}$: first guess an interpretation $\mathcal{I}$ of size at most $\mathsf{s}(\widehat{C}) \cdot \mathsf{s}(\widehat{D})$ over the signature $\mathfrak{C}(\widehat{C} \sqcap \neg\widehat{D}, \sim) \cup \mathfrak{R}(\widehat{C} \sqcap \neg\widehat{D}, \sim)$, and then check using Algorithm 1 (in polynomial time) whether there exists $d \in \Delta^{\mathcal{I}}$ such that $d \in \widehat{C}^{\mathcal{I}}$ and $d \notin \widehat{D}^{\mathcal{I}}$. Notice that the sets $\mathfrak{C}(\widehat{C} \sqcap \neg\widehat{D}, \sim)$ and $\mathfrak{R}(\widehat{C} \sqcap \neg\widehat{D}, \sim)$ can be computed in polynomial time, since $\mathfrak{s}_{pm}$ is assumed to be computable in polynomial time. Hence, we respectively obtain an NP and a coNP upper bound for satisfiability and subsumption in $\tau\mathcal{EL}(m_\sim)$.

Theorem 4.4. *Let $\sim \in$ simi-d. In $\tau\mathcal{EL}(m_\sim)$, satisfiability is in NP and subsumption is in coNP.*

4.2.2. Extending the bounded tree model property to acyclic TBoxes

To simplify the presentation, we make two assumptions in the rest of the paper. First, we restrict our attention to satisfiability of concepts of the form $\alpha_1 \sqcap \neg\alpha_2$ w.r.t. an acyclic TBox $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$, where $\alpha_1, \alpha_2 \in \mathsf{NC}_d^{\mathfrak{T}}$. This is without loss of generality since a $\tau\mathcal{EL}(m_\sim)$ concept $\widehat{C}$ correctly defined w.r.t. $\mathcal{T}$ can be equivalently replaced by a fresh concept name $\alpha_{\widehat{C}} \in \mathsf{ND}$, if we add the definition $\alpha_{\widehat{C}} \doteq \widehat{C}$ to $\widehat{\mathcal{T}}$. Second, we assume that the TBox $\mathfrak{T}$ is in *normal* form. The normal form we employ here is an extension of the normal form introduced in [5] for $\mathcal{EL}$ TBoxes. An acyclic TBox $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$ is said to be *normalized* if $\alpha \doteq \widehat{C}_\alpha \in \widehat{\mathcal{T}} \cup \mathcal{T}$ implies that $\widehat{C}_\alpha$ is of the form

$$\widehat{P}_1 \sqcap \cdots \sqcap \widehat{P}_k \sqcap \exists r_1.\beta_1 \sqcap \cdots \sqcap \exists r_n.\beta_n,$$

where $k, n \geq 0$, $\widehat{P}_j$ is either a concept name $A \in \mathsf{NC}_{pr}^{\mathfrak{T}}$ or of the form $E_{\bowtie t}$ with $E \in \mathsf{NC}_d^{\mathcal{T}}$, and $\beta_1, \ldots, \beta_n \in \mathsf{NC}_d^{\mathfrak{T}}$. As shown in Appendix B.3, there is a polynomial translation of acyclic $\tau\mathcal{EL}(m_\sim)$ TBoxes into normalized ones that preserves inferences such as satisfiability and subsumption.

As mentioned earlier, by using unfolding, we can reduce satisfiability of $\alpha_1 \sqcap \neg\alpha_2$ w.r.t. $\mathfrak{T}$ to satisfiability of the concept $u_{\mathfrak{T}}(\alpha_1) \sqcap \neg u_{\mathfrak{T}}(\alpha_2)$. Therefore, Lemma 4.2 implies that $\alpha_1 \sqcap \neg\alpha_2$ is satisfiable w.r.t. $\mathfrak{T}$ iff there exists a tree-shaped interpretation $\mathcal{I}$ over $\mathsf{NC}_{pr}^{\mathfrak{T}} \cup \mathsf{NR}$ with root d_0 such that:

- $d_0 \in (u_{\mathfrak{T}}(\alpha_1) \sqcap \neg u_{\mathfrak{T}}(\alpha_2))^{\mathcal{I}}$, and
- $\mathcal{I}$ has depth at most $\mathsf{rd}(u_{\mathfrak{T}}(\alpha_1) \sqcap u_{\mathfrak{T}}(\alpha_2))$.

Moreover, since $\mathsf{sig}(u_{\mathfrak{T}}(\alpha_1)) \cup \mathsf{sig}(u_{\mathfrak{T}}(\alpha_2)) \subseteq \mathsf{sig}(\mathfrak{T})$, Lemma 4.3 tells us that we can restrict the concept and role names that $\mathcal{I}$ gives a non-empty interpretation to those in the sets

$$\begin{aligned} &\mathfrak{C}(\mathfrak{T}, \sim) := \{B \in \mathsf{NC} \mid B \in \mathfrak{s}_{pm}(A) \wedge A \in \mathsf{sig}(\mathfrak{T}) \cap \mathsf{NC}\} \cup \{B \in \mathsf{ND} \mid B \in \mathsf{NC}_{pr}^{\mathfrak{T}}\}, \text{ and} \\ &\mathfrak{R}(\mathfrak{T}, \sim) := \{s \in \mathsf{NR} \mid s \in \mathfrak{s}_{pm}(r) \wedge r \in \mathsf{sig}(\mathfrak{T})\}. \end{aligned} \tag{20}$$

Using the assumptions made above about $\mathfrak{T}$, we can show that the role depth of the unfolded conjunction is linear in the size of $\mathsf{s}(\mathfrak{T})$ (see Appendix B).

Proposition 4.5. *Let $\mathfrak{T}$ be an acyclic $\tau\mathcal{EL}(m_\sim)$ TBox and $\alpha \in \mathsf{NC}_d^{\mathfrak{T}}$. Then, $\mathsf{rd}(u_{\mathfrak{T}}(\alpha)) \leq \mathsf{s}(\mathfrak{T})$.*

All these properties together with Proposition 2.29 yield the following finite tree model property.

Proposition 4.6. *Let $\sim \in$ simi-d, $\mathfrak{T}$ an acyclic $\tau\mathcal{EL}(m_\sim)$ TBox and $\alpha_1, \alpha_2 \in \mathsf{NC}_d^{\mathfrak{T}}$. In $\tau\mathcal{EL}(m_\sim)$, if $\alpha_1 \sqcap \neg\alpha_2$ is satisfiable w.r.t. $\mathfrak{T}$, then there is a finite tree model $\mathcal{I}$ of $\mathfrak{T}$ with root element d_0 such that:*

- *$d_0 \in (\alpha_1 \sqcap \neg\alpha_2)^{\mathcal{I}}$, $\mathcal{I}$ has depth at most $\mathsf{s}(\mathfrak{T})$, and*
- *$\mathcal{I}$ only interprets symbols in $\mathfrak{C}(\mathfrak{T}, \sim)$, $\mathfrak{R}(\mathfrak{T}, \sim)$ and $\mathsf{NC}_d^{\mathfrak{T}}$ as non-empty.*

Using this property, we will now devise a PSpace procedure that decides satisfiability and subsumption w.r.t. acyclic $\tau\mathcal{EL}(m_\sim)$ TBoxes in $\tau\mathcal{EL}(m_\sim)$ for all $\sim \in$ *simi-d*.

4.2.3. A PSpace decision procedure for satisfiability w.r.t. acyclic $\tau\mathcal{EL}(m_\sim)$ TBoxes

In the following, we assume that the given acyclic TBox $\mathfrak{T}$ is normalized. Our algorithm is an adaptation of the so-called *witness algorithm*, which can, for example, be used to show that satisfiability in $\mathcal{ALC}$ w.r.t. acyclic TBoxes is in PSpace [2]. It non-deterministically tries to construct a tree model of $\mathfrak{T}$ whose root satisfies $\alpha_1 \sqcap \neg\alpha_2$. To represent individuals of this model, we use a simple notion of *type*. Let Γ^+ be a set of $\tau\mathcal{EL}(m_\sim)$ concept descriptions that is closed under building sub-descriptions, and $\Gamma = \Gamma^+ \cup \Gamma^-$, where $\Gamma^- = \{\neg\widehat{C} \mid \widehat{C} \in \Gamma^+\}$. A *type for* Γ is a set $\nu \subseteq \Gamma$ satisfying the following conditions:

t1) For all $\widehat{C}_1 \sqcap \widehat{C}_2 \in \Gamma^+$: $\widehat{C}_1 \sqcap \widehat{C}_2 \in \nu$ iff $\{\widehat{C}_1, \widehat{C}_2\} \subseteq \nu$,
t2) For all $\neg(\widehat{C}_1 \sqcap \widehat{C}_2) \in \Gamma^-$: $\neg(\widehat{C}_1 \sqcap \widehat{C}_2) \in \nu$ iff $\{\neg\widehat{C}_1, \neg\widehat{C}_2\} \cap \nu \neq \emptyset$,
t3) For all $\widehat{C}_1 \in \Gamma^+$: $\widehat{C}_1 \in \nu$ iff $\neg\widehat{C}_1 \notin \nu$.

Intuitively, a type ν identifies an individual d_ν of a model $\mathcal{I}$ of $\mathfrak{T}$ by stating which "relevant" concept descriptions d_ν belongs to or not. To define what is relevant in the presence of $\mathfrak{T}$, we need to take into account not only the descriptions occurring in $\mathfrak{T}$, but also the definition of $\sim$. The relevant concepts from $\mathfrak{T}$ correspond to the set $\mathsf{sub}(\mathfrak{T})$ of sub-descriptions occurring in $\mathfrak{T}$, which is formally defined as

$$\mathsf{sub}(\mathfrak{T}) := \bigcup_{\alpha \doteq \widehat{C}_\alpha \in \widehat{\mathcal{T}}} (\{\alpha\} \cup \mathsf{sub}(\widehat{C}_\alpha)) \cup \bigcup_{E \doteq C_E \in \mathcal{T}} (\{E\} \cup \mathsf{sub}(C_E)). \tag{21}$$

Regarding the influence of $\sim$, as explained before, concept names occurring in $\mathfrak{C}(\mathfrak{T},\sim)$ may also be important when evaluating a threshold concept. Using the set $\mathsf{sub}(\mathfrak{T}) \cup \mathfrak{C}(\mathfrak{T},\sim)$ as set Γ^+ of positive concepts, the overall set Γ of relevant concept descriptions, which we call the *closure* of $(\mathfrak{T},\sim)$, also denoted by $\mathsf{cl}(\mathfrak{T},\sim)$, is then defined as

$$\Gamma := \mathsf{cl}(\mathfrak{T},\sim) := \mathsf{sub}(\mathfrak{T}) \cup \{\neg\widehat{C} \mid \widehat{C} \in \mathsf{sub}(\mathfrak{T})\} \cup \{A, \neg A \mid A \in \mathfrak{C}(\mathfrak{T},\sim)\}.$$

Concepts $\neg E_{\bowtie t}$ are represented in $\mathsf{cl}(\mathfrak{T},\sim)$ with an equivalent non-negated threshold concept, e.g., $\neg E_{<t}$ corresponds to $E_{\geq t}$.

We then say that ν is a *type for* $\mathfrak{T}$ if it is a type for $\mathsf{cl}(\mathfrak{T},\sim)$ that additionally satisfies:

- for all $\alpha \doteq \widehat{C}_\alpha \in \widehat{\mathcal{T}}$: $\alpha \in \nu$ iff $\widehat{C}_\alpha \in \nu$.
- for all $E \doteq C_E \in \mathcal{T}$: $E \in \nu$ iff $C_E \in \nu$.

The next step is to define the notion of *successor candidate*. This has the purpose of capturing minimal requirements that the group of role successors of an individual d_ν must satisfy in an interpretation. We first need to introduce some notation. Given $\nu \subseteq \Gamma$, we define the following sets:

- $\mathsf{rol}(\nu) := \{r \mid \exists r.\widehat{C} \in \nu \text{ or } \neg\exists r.\widehat{C} \in \nu\}$,
- $\nu^+(r) := \{\widehat{C} \mid \exists r.\widehat{C} \in \nu\}$ for all $r \in \mathsf{rol}(\nu)$.

Then, given $r \in \mathsf{rol}(\nu)$ and a sequence $\varphi = \nu^1, \ldots, \nu^\ell$ of subsets of Γ ($\ell \geq 0$), we say that φ is an *r-successor candidate* of ν w.r.t. Γ if the following holds for all $\exists r.\widehat{C} \in \Gamma$:

c1) $\widehat{C} \in \nu^+(r)$ iff $\widehat{C} \in \nu^j$ for some $1 \leq j \leq \ell$.

Before moving into the presentation of our algorithm, let us first illustrate the previously introduced notions with the following example.

Example 4.7. Let $m = m_{\sim_d^*}$, where $\sim_d^*$ is the directional measure from Example 3.16. In addition, consider the acyclic $\tau\mathcal{EL}(m)$ TBox $\mathfrak{T} := (\widehat{\mathcal{T}}, \mathcal{T})$, where:

$$\widehat{\mathcal{T}} := \left\{ \begin{array}{l} \alpha \doteq \exists r.E_1 \sqcap \exists s.\beta \\ \beta \doteq (E_1)_{\geq 1/2} \sqcap B \end{array} \right\} \qquad \mathcal{T} := \left\{ \begin{array}{l} E_1 \doteq \exists r.E_2 \sqcap \exists r.E_3 \\ E_2 \doteq A,\ E_3 \doteq B \end{array} \right\}.$$

Types for $\mathfrak{T}$ are defined w.r.t. the closure of $(\mathfrak{T},\sim_d^*)$, i.e., $\mathsf{cl}(\mathfrak{T},\sim_d^*)$. To obtain this set, one first needs to compute the sets $\mathsf{sub}(\mathfrak{T})$ and $\mathfrak{C}(\mathfrak{T},\sim_d^*)$. As defined in (21), the set $\mathsf{sub}(\mathfrak{T})$ can be expressed as:

$$\mathsf{sub}(\mathfrak{T}) := \{\alpha, \beta, E_1, E_2, E_3\} \cup \mathsf{sub}(\exists r.E_1 \sqcap \exists s.\beta) \cup \mathsf{sub}((E_1)_{\geq 1/2} \sqcap B) \cup \mathsf{sub}(\exists r.E_2 \sqcap \exists r.E_3) \cup \{A, B\},$$

where the sets of the form $\mathsf{sub}(X)$ on the right-hand side of this expression are easily obtained by applying the definition of sub-description. To obtain the set $\mathfrak{C}(\mathfrak{T},\sim_d^*)$, we apply (20) w.r.t. $\mathfrak{T}$ and $\sim_d^*$. Note that $\mathsf{NC}_{pr}^{\mathfrak{T}} = \{A, B\}$ since $\mathsf{NC}_d^{\mathfrak{T}} = \{\alpha, \beta, E_1, E_2, E_3\}$. Moreover, since $\sim_d^*$ is defined using the default primitive measures pm_{df}, we know that $pm_{df}(A, B) \neq 0$ iff $A = B$ (for all $A, B \in \mathsf{NC}$). Hence, by definition of $\mathfrak{s}_{pm}$, we have $\mathfrak{s}_{pm_{df}}(A) = \{A\}$ and $\mathfrak{s}_{pm_{df}}(B) = \{B\}$. From this, one can derive that $\mathfrak{C}(\mathfrak{T},\sim_d^*) = \{A, B\}$. Thus, since $\mathfrak{C}(\mathfrak{T},\sim_d^*) \subseteq \mathsf{sub}(\mathfrak{T})$, the closure of $(\mathfrak{T},\sim_d^*)$ consists of the following set:

$$\mathsf{cl}(\mathfrak{T},\sim_d^*) = \mathsf{sub}(\mathfrak{T}) \cup \{\neg\widehat{C} \mid \widehat{C} \in \mathsf{sub}(\mathfrak{T})\}.$$

Finally, let us now briefly look at the notions of type and role successor candidates.

- By definition, a type for $\mathfrak{T}$ is a type for $\mathsf{cl}(\mathfrak{T},\sim_d^*)$ satisfying the definitions and GCIs in $\mathfrak{T}$. Suppose, for instance, that ν is a type for $\mathfrak{T}$ and $\{\alpha, \neg\beta\} \subseteq \nu$. Then, satisfying the definition for α in $\widehat{\mathcal{T}}$ requires that $\exists r.E_1 \sqcap \exists s.\beta \in \nu$. The latter, together with condition t1) in the definition of type, implies that $\{\exists r.E_1, \exists s.\beta\} \subseteq \nu$. Furthermore, the application of condition t3) to $\neg\beta \in \nu$ yields

Algorithm 2 Satisfiability of $\alpha_1 \sqcap \neg\alpha_2$ w.r.t. acyclic $\tau\mathcal{EL}(m_\sim)$ TBoxes in $\tau\mathcal{EL}(m_\sim)$, where $\sim \in$ *simi-d*.

Input: An acyclic $\tau\mathcal{EL}(m_\sim)$ TBox $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$ with $\mathsf{NC}_d^{\mathcal{T}} = \{E_1, \dots, E_m\}$ and $\alpha_1, \alpha_2 \in \mathsf{NC}_d^{\mathfrak{T}}$.
Output: "yes", if $\alpha_1 \sqcap \neg\alpha_2$ is satisfiable in $\tau\mathcal{EL}(m_\sim)$ w.r.t. $\mathfrak{T}$; "no" otherwise.

1: **return** "yes" iff WITNESS$(\nu, 0)$ does not fail for some $\nu \subseteq \mathsf{cl}(\mathfrak{T}, \sim)$ s.t. $\{\alpha_1, \neg\alpha_2\} \subseteq \nu$

1: **procedure** WITNESS(ν, d)
2: **if** $(\mathsf{d} > \mathfrak{d})$ **or** (ν is not a type for $\mathfrak{T}$) **then** *fail*
3: **for all** $r \in \mathfrak{R}(\mathfrak{T}, \sim)$ **do**
4: non-deterministically choose a value $0 \leq \ell_r \leq \mathfrak{n}$ and sets $\nu_r^1, \dots, \nu_r^{\ell_r} \subseteq \mathsf{cl}(\mathfrak{T}, \sim)$
5: **if** $r \in \mathsf{rol}(\nu)$ and $\nu_r^1, \dots, \nu_r^{\ell_r}$ is not an r-successor candidate of ν w.r.t. $\mathsf{cl}(\mathfrak{T}, \sim)$ **then** *fail*
6: **for all** $j \in \{1 \dots \ell_r\}$ **do**
7: $q^{(\nu,r,j)} :=$ WITNESS$(\nu_r^j, \mathsf{d} + 1)$
8: **for all** $1 \leq i \leq m$ **do**
9: $q_i^\nu :=$ DEGREES$(\nu \cap \mathsf{NC}, \{(r, q^{(\nu,r,j)}) \mid r \in \mathfrak{R}(\mathfrak{T}, \sim) \wedge 1 \leq j \leq \ell_r\}, E_i)$
10: **if** $q_i^\nu \bowtie t$ does not hold for some $(E_i)_{\bowtie t} \in \nu$ **then** *fail*
11: **return** $q^\nu = (q_1^\nu, \dots, q_m^\nu)$

1: **procedure** DEGREES(S, Q, E_i)
2: $aux := 0$
3: **for all** $A \in top(C_{E_i}) \cap N_C$ **do**
4: $aux := aux + g(A) \cdot \max\{simi_a(A, B) \mid B \in S\}$
5: **for all** $\exists s.E_p \in top(C_{E_i})$ **do**
6: $aux := aux + g(s) \cdot \max\limits_{(r,q) \in Q} \{pm(s, r)[w + (1 - w)q_p]\}$
7: **return** $aux / \sum_{C' \in top(C_{E_i})} g(C')$

that $\beta \notin \nu$. Consequently, to satisfy the definition for β in $\widehat{\mathcal{T}}$ it must be that $(E_1)_{\geq 1/2} \sqcap B \notin \nu$. This implies that $\{(E_1)_{\geq 1/2}, B\} \nsubseteq \nu$, by condition t1). It is not hard to extend this partial definition of ν into a concrete subset of $\mathsf{cl}(\mathfrak{T}, \sim_d^*)$ that is indeed a type for $\mathfrak{T}$.
- Since $\{\exists r.E_1, \exists s.\beta\} \subseteq \nu$, this means that $\{r, s\} \subseteq \mathsf{rol}(\nu)$, $E_1 \in \nu^+(r)$ and $\beta \in \nu^+(s)$. Hence, by condition c1), an r-successor candidate of ν w.r.t. $\mathsf{cl}(\mathfrak{T}, \sim_d^*)$ must contain a type ν_r' such that $E_1 \in \nu_r'$, whereas an s-successor candidate must contain a type ν_s' such that $\beta \in \nu_s'$. Obviously, the composition of an r-successor candidate of ν also depends on whether $\exists r.E_2$ and $\exists r.E_3$ belong to ν or not. For example, if $\exists r.E_2 \notin \nu$, then $E_2 \notin \nu^+(r)$. Consequently, condition c1) implies that an r-successor candidate of ν cannot contain a type ν_r' such that $E_2 \in \nu_r'$. $\triangle$

Based on the previously defined notions, Algorithm 2 invokes the procedure WITNESS to guess a tree-shaped model $\mathcal{I}$ of $\mathfrak{T}$ satisfying $\alpha_1 \sqcap \neg\alpha_2$. The values $\mathfrak{d} := \mathsf{s}(\mathfrak{T})$ and $\mathfrak{n} := 2 \cdot \mathsf{s}(\mathfrak{T})$ respectively bound the depth of the model and the number of successors considered for each individual of it. The distinctive feature of the algorithm is that it needs to verify whether an individual d_ν represented by a type ν is an instance of each threshold concept $E_{\bowtie t} \in \nu$. To do this, it computes a tuple $q^\nu = (q_1^\nu, \dots, q_m^\nu)$ such that

$$q_i^\nu = \widehat{m}_\sim^{\mathcal{I}}(d_\nu, E_i, \mathcal{T}), \tag{22}$$

where $E_1, \dots, E_m$ are the defined concepts in $\mathcal{T}$. This tuple is then used in Line 10 to verify whether $d_\nu \in [(E_i)_{\bowtie t}]^{\mathcal{I}}$. To compute q^ν, the procedure DEGREES is invoked for each E_i, and receives as input the concept names in ν and the set of all (already computed) tuples $q^{(\nu,r,j)}$ corresponding to the successors of d_ν represented by each type ν_r^j. In this procedure, the q in $(r, q) \in Q$ is an m-tuple and q_p is its p-th component, corresponding to the defined concept E_p.

The following example describes a successful run of Algorithm 2 for $\tau\mathcal{EL}(m_{\sim_d^*})$ on input $\mathfrak{T}$ and $\alpha \sqcap \neg\beta$, where $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$ is the $\tau\mathcal{EL}(m)$ TBox introduced in Example 4.7.

Example 4.8. A run of Algorithm 2 for the logic $\tau\mathcal{EL}(m_{\sim_d^*})$ requires the sets $\mathsf{cl}(\mathfrak{T}, \sim_d^*)$, $\mathsf{NC}_d^{\mathcal{T}}$ and $\mathfrak{R}(\mathfrak{T}, \sim_d^*)$. We already know from Example 4.7 how $\mathsf{cl}(\mathfrak{T}, \sim_d^*)$ looks like. In addition, it is easy to see that $\mathsf{NC}_d^{\mathcal{T}} = \{E_1, E_2, E_3\}$. As for $\mathfrak{R}(\mathfrak{T}, \sim_d^*)$, since $\sim_d^*$ is defined using pm_{df}, we know that:

$$pm_{df}(r, s) \neq 0 \quad \text{iff} \quad r = s \quad (\text{for all } r, s \in \mathsf{NR}).$$

Hence, by definition of $\mathfrak{s}_{pm}$, we have that $\mathfrak{s}_{pm_{df}}(r) = \{r\}$ and $\mathfrak{s}_{pm_{df}}(s) = \{s\}$. Thus, by (20) we obtain that $\mathfrak{R}(\mathfrak{T}, \sim_d^*) = \{r, s\}$.

We now use Fig. 4 to describe a run of the algorithm. The left-hand side of this figure sketches the recursive tree of a successful run of WITNESS that leads to a successful run of Algorithm 2 on input $\mathfrak{T}$ and $\alpha \sqcap \neg\beta$. The nodes of the tree contain the pair (ν, d) and the values ℓ_r, ℓ_s and q^ν of the recursive call of WITNESS they represent. The types at each node are partially described, but can easily be completed into a type for $\mathfrak{T}$ such that the test at Line 5 never fails. They only contain the necessary information to illustrate the relevant aspects of the algorithm. Let us start by looking at the types and the values ℓ_r, ℓ_s.

- Since $\{\alpha, \neg\beta\} \subseteq \nu$, we have already seen in Example 4.7 that ν satisfies that:

$$\{\exists r.E_1, \exists s.\beta\} \subseteq \nu \quad \text{and} \quad \{(E_1)_{\geq 1/2}, B\} \nsubseteq \nu.$$

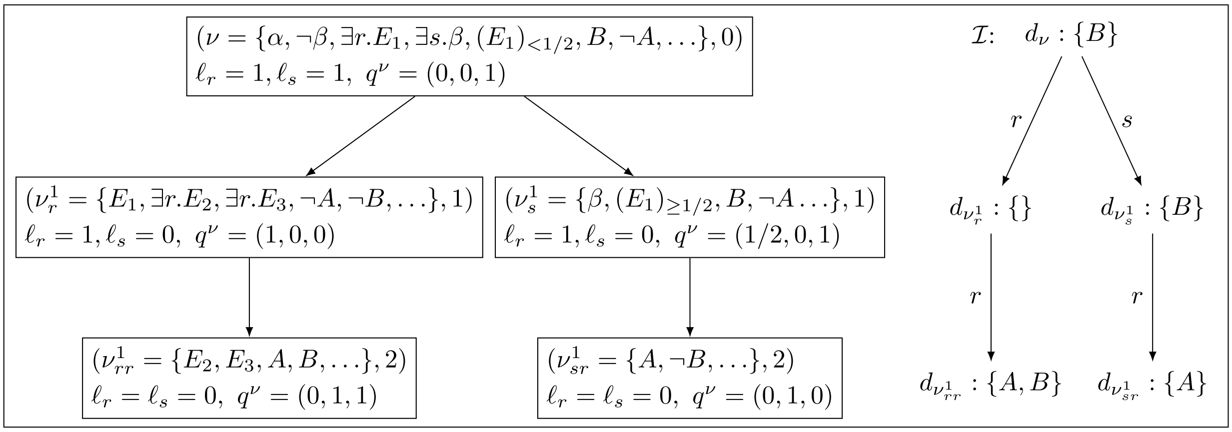


Fig. 4. Successful run of Algorithm 2 and its induced interpretation for Example 4.8.

To satisfy the non-inclusion, the run guesses $\neg(E_1)_{\geq 1/2}$ in ν, i.e., $(E_1)_{<1/2}$.

- To satisfy the test at Line 5 for ν, the values of ℓ_r and ℓ_s must be at least 1, since $\{\exists r.E_1, \exists s.\beta\} \subseteq \nu$. Then, the types ν_r^1 and ν_s^1 containing E_1 and β, respectively, are guessed to provide the necessary r-successor and s-successor candidates of ν. Note that $\{\exists r.E_2, \exists r.E_3\} \subseteq \nu_r^1$ is needed since ν_r^1 is a type for $\mathfrak{T}$ and $E_1 \in \nu_r^1$. Similarly for $\{(E_1)_{\geq 1/2}, B\} \subseteq \nu_s^1$ w.r.t. $\beta \in \nu_s^1$.
- The call for ν_r^1 guesses $\ell_r = 1$ and ν_{rr}^1 with $\{E_2, E_3, A, B\} \subseteq \nu_{rr}^1$. In this way, the sequence ν_{rr}^1 serves as an r-successor candidate of ν_r^1, which is required since $\{\exists r.E_2, \exists r.E_3\} \subseteq \nu_r^1$. As for ν_s^1, its recursive call guesses $\ell_r = 1$ and ν_{sr}^1 to ensure that the test at Line 10 will not fail for $(E_1)_{\geq 1/2} \in \nu_s^1$.
- Lastly, the recursive calls for ν_{rr}^1 and ν_{sr}^1 guess both $\ell_r = \ell_s = 0$, which means that no additional nested recursive calls are executed.

Let us now look at the computation of q^ν.

- From the parameters used to define $\sim_d^*$ and the definitions of E_2, E_3 in $\mathcal{T}$, if follows that q_2^ν and q_3^ν are always in $\{0, 1\}$. In particular, $q_2^\nu = 1$ iff $A \in S$ for the parameter S of DEGREES, and similarly for q_3^ν and B. This implies, for instance, $q^\nu = (\_, 1, 0)$ for ν_{sr}^1 and $q^\nu = (\_, 0, 0)$ for ν_r^1.
- Regarding E_1 and q_1^ν, DEGREES iterates over the top-level atoms $\exists r.E_2$ and $\exists r.E_3$ of C_{E_1}. In this case $q_1^\nu \in \{0, 1/2, 1\}$. The exact value depends on the tuples of the form $(r, (\_, q_2, q_3)) \in Q$. For instance, to compute q_1^ν for ν_s^1, DEGREES is invoked with parameters $(\{B\}, \{(r, (0, 1, 0))\}, E_1)$, where $(0, 1, 0)$ is the tuple returned for ν_{sr}^1. This yields $q_1^\nu = 1/2$, which ensures that the test at Line 10 does not fail for $(E_1)_{\geq 1/2} \in \nu_s^1$. Analogously, for the initial type ν, DEGREES is called with $Q = \{(r, (1, 0, 0)), (s, (1/2, 0, 1))\}$. This yields $q_1^\nu = 0$ and a successful test at Line 10 for $(E_1)_{<1/2} \in \nu$.

Finally, the described run induces the tree-shaped primitive interpretation $\mathcal{I}$ depicted on the right-hand side of Fig. 4. More precisely, each recursive call WITNESS(ν, d) yields an element $d_\nu \in \Delta^\mathcal{I}$. The primitive concepts satisfied by d_ν are determined by ν, whereas its successors are determined by the subsequent nested recursive calls. In our example, one can verify that (22) holds for all guessed types and their corresponding elements of $\Delta^\mathcal{I}$ w.r.t. $\widehat{m}^\mathcal{I}_{\sim_d^*}$. This yields $d_\nu \in [(E_1)_{<1/2}]^\mathcal{I}$ and $d_{\nu_s^1} \in [(E_1)_{\geq 1/2}]^\mathcal{I}$. It is then not hard to see that the unique extension of $\mathcal{I}$ into a model of $\mathfrak{T}$ satisfies $\alpha \sqcap \neg\beta$. △

For the correctness of the procedure, one further assumption is needed regarding $\mathcal{T}$. Recall that, by Definition 2.23, the right-hand side of equation (22) is expanded to $m_\sim^\mathcal{I}(d_\nu, u_\mathcal{T}(E_i))$. In addition, the definition of $\sim \in$ *simi-d* employs the reduced form of $u_\mathcal{T}(E_i)$ in the computation of the value $m_\sim^\mathcal{I}(d_\nu, u_\mathcal{T}(E_i))$. Since $u_\mathcal{T}(E_i)$ may have size exponential in the size of $\mathcal{T}$, it is not a good idea to compute $u_\mathcal{T}(E_i)$ first, and then reduce it. The subroutine DEGREES is designed to avoid unfolding, but to guarantee that it computes the right value for q_i^ν, we need to ensure that $u_\mathcal{T}(E_i)$ is already reduced. To achieve this, we introduce the notion of a reduced form for acyclic $\mathcal{EL}$ TBoxes. Its definition is based on the reduced form for $\mathcal{EL}$ concepts introduced by Küsters [27] (see Section 2).

Definition 4.9. Let $\mathcal{T}$ be an acyclic $\mathcal{EL}$ TBox and C an $\mathcal{EL}$ concept description. We say that C is *reduced* w.r.t. $\mathcal{T}$ if, C is *reduced* according to Küsters' definition modulo $\sqsubseteq_\mathcal{T}$ (i.e., $\sqsubseteq_\mathcal{T}$ is used to identify redundancies instead of $\sqsubseteq$). Then, we say that $\mathcal{T}$ is in *reduced form* if for all $E \doteq C_E \in \mathcal{T}$ the concept C_E is reduced w.r.t. $\mathcal{T}$. Finally, a $\tau\mathcal{EL}(m_\sim)$ TBox $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$ is in reduced form if $\mathcal{T}$ is in reduced form.

As shown in Appendix B.3, one can restrict the attention without loss of generality to normalized $\tau\mathcal{EL}(m_\sim)$ TBoxes in reduced form. Moreover, it is also proved there that unfolding always yields concepts in reduced form.

Lemma 4.10. *Let $\mathcal{T}$ be a normalized acyclic $\mathcal{EL}$ TBox in reduced form. Then, for all concept names $E \in \mathsf{NC}_d^{\mathcal{T}}$, the unfolded $\mathcal{EL}$ concept description $u_{\mathcal{T}}(E)$ is reduced.*

We now prove that Algorithm 2 is correct, i.e., it is *sound* and *complete*. The proof is divided into the three lemmas that follow. The first one shows that DEGREES computes the value of the right-hand side of (22), when invoked with appropriate parameters. This result is then used in the proofs of Lemma 4.12 and Lemma 4.13, which show soundness and completeness of the algorithm, respectively.

Lemma 4.11. *Let $\sim \in$ simi-d, $\mathcal{T} \in \mathfrak{S}_{\mathsf{D}}$ normalized and reduced, $\mathcal{I}$ a model of $\mathcal{T}$, and $d \in \Delta^{\mathcal{I}}$ an element with finitely many role successors. Define, for all $f \in \Delta^{\mathcal{I}}$, the set S_f and the rational number q_i^f $(1 \leq i \leq m)$ as*

$$S_f := \{A \mid f \in A^{\mathcal{I}} \wedge A \in \mathsf{NC}\} \quad \text{and} \quad q_i^f := \widehat{m}_{\sim}^{\mathcal{I}}(f, E_i, \mathcal{T}).$$

In addition, let Q_i $(1 \leq i \leq m)$ be sets such that

- *$Q_i \subseteq \{(r, (q_1^f, \ldots, q_m^f)) \mid f \in \Delta^{\mathcal{I}} \wedge (d, f) \in r^{\mathcal{I}} \wedge r \in \mathsf{NR}\}$, and*
- *for all $\exists s.E_p \in top(C_{E_i})$ and $r \in \mathsf{NR}$, if d has r-successors then Q_i contains a tuple $(r, (q_1^e, \ldots, q_m^e))$ such that $(d, e) \in r^{\mathcal{I}}$ and*

$$pm(s,r)[w + (1-w)q_p^e] = \max\{pm(s,r)[w + (1-w)q_p^f] \mid (d, f) \in r^{\mathcal{I}}\}.$$

Then, with input parameters (S_d, Q_i, E_i), DEGREES *computes the value* $\text{DEGREES}(S_d, Q_i, E_i) = q_i^d$.

Proof. By Definitions 2.23 and 3.2 we have:

$$q_i^d = \widehat{m}_{\sim}^{\mathcal{I}}(d, E_i, \mathcal{T}) = m_{\sim}^{\mathcal{I}}(d, u_{\mathcal{T}}(E_i)) = \max\{u_{\mathcal{T}}(E_i) \sim D \mid D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR}) \text{ and } d \in D^{\mathcal{I}}\}. \tag{23}$$

Let D^* be an $\mathcal{EL}$ concept description such that $d \in D^{*\mathcal{I}}$ and $u_{\mathcal{T}}(E_i) \sim D^* = q_i^d$. Since $\mathcal{T}$ is normalized and in reduced form, we know that every top-level atom of $u_{\mathcal{T}}(E_i)$ is of the form $A \in \mathsf{NC}$ or $\exists s.u_{\mathcal{T}}(E_p)$, and $u_{\mathcal{T}}(E_i)$ is in reduced form. Then, by the directional definition of *simi*, the computation of $u_{\mathcal{T}}(E_i) \sim D^*$ consists of two main operations:
i) adding up the values $v(F) := g(F) \cdot \max_{D' \in top(D^*)} (simi_a(F, D'))$ for all $F \in top(u_{\mathcal{T}}(E_i))$, and
ii) dividing the obtained value by the sum of all $g(F)$.
Therefore, according to Line 7 of the procedure DEGREES, to show that $\text{DEGREES}(S_d, Q_i, E_i) = q_i^d$, it is enough to verify that the sum in i) is equal to the final value of *aux*. To this end, we show that for all $F \in top(u_{\mathcal{T}}(E_i))$ the value $v(F)$ is the value added to *aux* at Line 4 or 6. We consider two cases:

a) $F = A \in \mathsf{NC}$. If D' is of the form $\exists r.X$, we know that $simi_a(A, D') = 0$. Otherwise, $D' \in \mathsf{NC}$, and $d \in D'^{\mathcal{I}}$ implies $D' \in S_d$. Hence, it follows that $v(F) \leq g(A) \cdot \max\{simi_a(A, B) \mid B \in S_d\}$. We now assume that $v(F) < g(A) \cdot \max\{simi_a(A, B) \mid B \in S_d\}$, and lead this to a contradiction. In fact, in this case there must be a $B \in S_d$ such that $simi_a(A, B) > simi_a(A, D')$ for all $D' \in top(D^*)$. But then $d \in (D^* \sqcap B)^{\mathcal{I}}$ and $u_{\mathcal{T}}(E_i) \sim D^* < u_{\mathcal{T}}(E_i) \sim (D^* \sqcap B)$, which contradicts our assumption that D^* yields the maximal value q_i^d.

b) $F = \exists s.u_{\mathcal{T}}(E_p)$. Let $D' \in top(D^*)$ be such that $v(F) = g(s) \cdot simi_a(F, D')$ (recall that in this case $g(F) = g(s)$). If $D' \in \mathsf{NC}$ then $v(F) = 0$. Otherwise, $D' = \exists r.G$ and

$$v(F) = g(s) \cdot simi_a(F, \exists r.G) = g(s) \cdot pm(s,r)[w + (1-w)(u_{\mathcal{T}}(E_p) \sim G)].$$

Since $d \in D'^{\mathcal{I}}$, there exists $f \in \Delta^{\mathcal{I}}$ such that $(d, f) \in r^{\mathcal{I}} \wedge f \in G^{\mathcal{I}}$. By maximality of q_p^f in (23), the following inequality holds:

$$v(F) \leq g(s) \cdot pm(s,r)[w + (1-w)q_p^f].$$

Moreover, since $\exists s.E_p \in top(C_{E_i})$ and d has r-successors in $\mathcal{I}$, there exists $(r, q^e) \in Q_i$ such that:

$$pm(s,r)[w + (1-w)q_p^f] \leq pm(s,r)[w + (1-w)q_p^e].$$

Hence, it follows that $v(F)$ is not greater than the value added to *aux* at Line 6 for $\exists s.E_p$.

Conversely, let $(r, q) \in Q_i$ be such that q adds the maximal value to *aux* for $\exists s.E_p$. By definition of Q_i, there is $f \in \Delta^{\mathcal{I}}$ such that $(d, f) \in r^{\mathcal{I}}$ and $q = q^f$. Let D_f be such that $f \in D_f^{\mathcal{I}}$ and $q_p^f = u_{\mathcal{T}}(E_p) \sim D_f$. Then, the value added to *aux* for $\exists s.E_p$ corresponds to

$$g(s) \cdot pm(s,r)[w + (1-w)(u_{\mathcal{T}}(E_p) \sim D_f)] = g(s) \cdot simi_a(F, \exists r.D_f).$$

As in case a), we can show that assuming $simi_a(F, \exists r.D_f) > simi_a(F, \exists r.D')$ leads to a contradiction since then $D^* \sqcap \exists r.D_f$ would yield a larger value than D^*. Consequently, $v(F) = simi_a(F, \exists r.D_f)$, which is the value added to *aux* for $\exists s.E_p$.

This completes the proof of the lemma. $\square$

We are now ready to show soundness of the algorithm. Suppose that it answers “yes” on input $\mathfrak{T}, \alpha_1, \alpha_2$. This means that there exists $v_0 \subseteq \mathsf{cl}(\mathfrak{T}, \sim)$ such that $\{\alpha_1, \neg\alpha_2\} \subseteq v_0$ and WITNESS has a successful run on $(v_0, 0)$. Then, Lemma 4.12 below implies that $\alpha_1 \sqcap \neg\alpha_2$ is satisfiable w.r.t. $\mathfrak{T}$. Note that, given the successful run on $(v_0, 0)$, the lemma implies the existence of a model $\mathcal{I}$ of $\mathfrak{T}$ satisfying property 1 in the statement of the lemma. Hence, since $\{\alpha_1, \neg\alpha_2\} \subseteq v_0$, the model $\mathcal{I}$ satisfies $\alpha_1 \sqcap \neg\alpha_2$. The idea underlying the proof of Lemma 4.12 is to use the recursion tree of a successful run on a pair (v, d), to inductively build a model $\mathcal{I}$ of $\mathfrak{T}$ satisfying properties 1 and 2 in the statement of the lemma. The first step of the proof consists of applying this inductive construction to obtain the interpretation $\mathcal{I}$. Afterwards, we show that $\mathcal{I}$ is a model of $\mathfrak{T}$ satisfying properties 1 and 2, by using Lemma 4.11 and the properties inductively satisfied by the other types guessed along the run.

Lemma 4.12. *Let $v \subseteq \mathsf{cl}(\mathfrak{T}, \sim)$ and $0 \leq \mathsf{d} \leq \mathfrak{d}$. If* WITNESS *has a successful run on (v, d) with output $q^v = (q_1^v, \ldots, q_m^v)$, then there exists a model $\mathcal{I}$ of $\mathfrak{T}$ and $d_v \in \Delta^{\mathcal{I}}$ such that:*

1. $d_v \in \bigcap_{X \in v} X^{\mathcal{I}}$,
2. $q_i^v = \widehat{m}_{\sim}^{\mathcal{I}}(d_v, E_i, \mathcal{T})$ *for all* $i, 1 \leq i \leq m$.

Proof. The proof is by induction on $\mathfrak{d} - \mathsf{d}$. Since WITNESS has a successful run on (v, d), this means that v is a type for $\mathfrak{T}$. Let $r \in \mathfrak{R}(\mathfrak{T}, \sim)$ and ℓ_r be the corresponding value guessed at Line 4. If $\ell_r > 0$, then later at Line 7 there must occur non-failing calls to WITNESS on $(v_r^j, \mathsf{d}+1)$ returning tuples $q^{(v,r,j)}$ for all $j, 1 \leq j \leq \ell_r$. The application of induction to $(v_r^j, \mathsf{d}+1)$ yields an interpretation $\mathcal{I}_{r,j}$ and $d_{r,j} \in \Delta^{\mathcal{I}_{r,j}}$ such that $\mathcal{I}_{r,j} \models \mathfrak{T}$, and $d_{r,j}$ and $q^{(v,r,j)}$ satisfy 1 and 2 w.r.t. v_r^j and $\mathcal{I}_{r,j}$. We assume without loss of generality that the domains of these interpretations are pairwise disjoint. We build an interpretation $\mathcal{I}$ as follows:

- $\Delta^{\mathcal{I}} := \{d_v\} \cup \{e \in \Delta^{\mathcal{I}_{r,j}} \mid r \in \mathfrak{R}(\mathfrak{T}, \sim) \wedge 1 \leq j \leq \ell_r\}$,
- $A^{\mathcal{I}} := \{d_v \mid \text{if } A \in v\} \cup \{e \in A^{\mathcal{I}_{r,j}} \mid r \in \mathfrak{R}(\mathfrak{T}, \sim) \wedge 1 \leq j \leq \ell_r\}$ for all $A \in \mathsf{NC} \cup \mathsf{ND}$,
- $r^{\mathcal{I}} := \{(d_v, d_{r,j}) \mid r \in \mathfrak{R}(\mathfrak{T}, \sim) \wedge 1 \leq j \leq \ell_r\} \cup \bigcup_{s \in \mathfrak{R}(\mathfrak{T}, \sim)} \bigcup_{j=1}^{\ell_s} r^{\mathcal{I}_{s,j}}$ for all $r \in \mathsf{NR}$.

Let us first show that $\mathcal{I}$ is a model of $\mathcal{T}$. By construction of $\mathcal{I}$, we have that $e \in \Delta^{\mathcal{I}_{r,j}}$ implies $(e, f) \notin s^{\mathcal{I}}$ for all $s \in \mathsf{NR}$ and $f \notin \Delta^{\mathcal{I}_{r,j}}$. From this, we obtain the following for all interpretations $\mathcal{I}_{r,j}$, $e \in \Delta^{\mathcal{I}_{r,j}}$ and $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC} \cup \mathsf{ND}, \mathsf{NR})$:

$$e \in C^{\mathcal{I}_{r,j}} \text{ iff } e \in C^{\mathcal{I}} \tag{24}$$

Since $\mathcal{I}_{r,j} \models \mathcal{T}$, this means that e satisfies the definitions in $\mathcal{T}$ w.r.t. $\mathcal{I}$. Hence, we only need to show that d_v also satisfies the definitions in $\mathcal{T}$. To this end, we show that the following holds for all $\mathcal{EL}$ concepts X in $\mathsf{cl}(\mathfrak{T}, \sim)$:

$$X \in v \text{ iff } d_v \in X^{\mathcal{I}} \tag{25}$$

We distinguish the following cases:

- $X = A \in \mathsf{NC} \cup \mathsf{ND}$. In this case, the equivalence (25) is a direct consequence of the definition of $\mathcal{I}$.
- $X = \exists r.C$. Suppose $X \in v$. For the test at Line 5 not to fail, WITNESS must have guessed $\ell_r > 0$ and an r-successor candidate $v_r^1, \ldots, v_r^{\ell_r}$ of v w.r.t. $\mathsf{cl}(\mathfrak{T}, \sim)$ such that $C \in v_r^j$ for some $1 \leq j \leq \ell_r$ (see c1)). This yields $d_{r,j} \in C^{\mathcal{I}_{r,j}}$ by induction, and thus $d_{r,j} \in C^{\mathcal{I}}$ by (24). Hence, since $(d_v, d_{r,j}) \in r^{\mathcal{I}}$, we obtain $d_v \in (\exists r.C)^{\mathcal{I}}$.
  Conversely, assume that $d_v \in (\exists r.C)^{\mathcal{I}}$. Then, there exists $d_{r,j} \in \Delta^{\mathcal{I}}$ such that $d_{r,j} \in C^{\mathcal{I}}$. By (24) we have that $d_{r,j} \in C^{\mathcal{I}_{r,j}}$. Further, by induction we know that $d_{r,j}$ satisfies 1 w.r.t. v_r^j and $\mathcal{I}_{r,j}$. Assume that $C \notin v_r^j$. Then t3) would yield $\neg C \in v_r^j$, and thus $d_{r,j} \notin C^{\mathcal{I}_{r,j}}$ by 1, which contradicts $d_{r,j} \in C^{\mathcal{I}}$. Thus, we must have $C \in v_r^j$. Since the guessed r-successor candidate $v_r^1, \ldots, v_r^{\ell_r}$ of v w.r.t. $\mathsf{cl}(\mathfrak{T}, \sim)$ satisfies property c1), this implies that $\exists r.C \in v$.
- Otherwise, X is a conjunction of concept names and existential restrictions. The equivalence (25) is then an easy consequence of the previous cases and the fact that v satisfies t1).

Summing up, we have show (25) for all $\mathcal{EL}$ concepts in $\mathsf{cl}(\mathfrak{T}, \sim)$. Hence, for all $E \doteq C_E \in \mathcal{T}$ we have that $d_v \in E^{\mathcal{I}}$ iff $E \in v$ iff $C_E \in v$ iff $d_v \in C_E^{\mathcal{I}}$. The second equivalence is a consequence of the fact that v is a type for $\mathfrak{T}$. This shows that $\mathcal{I}$ is a model of $\mathcal{T}$.

Next, we consider arbitrary (non-negated) $\tau\mathcal{EL}(m)$ concept descriptions. Using the same argument as for $\mathcal{EL}$, we obtain the analogon of (24) for all $\tau\mathcal{EL}(m)$ concepts $\widehat{C}$. Moreover, the inductive property satisfied by $q^{(v,r,j)}$ w.r.t. $\mathcal{I}_{r,j}$ also holds w.r.t. $\mathcal{I}$, i.e.:

$$q_i^{(v,r,j)} := \widehat{m}_{\sim}^{\mathcal{I}}(d_{r,j}, E_i, \mathcal{T}), \text{ for all } 1 \leq i \leq m.$$

This, together with the definition of $\mathcal{I}$ and the fact that $\mathcal{I} \models \mathcal{T}$, allows us to apply Lemma 4.11 to obtain that q^v satisfies property 2. In fact, it is easy to see that the calls of DEGREES within the procedure WITNESS are of the form considered in Lemma 4.11.

To finish the proof of the lemma, it remains to show that d_v satisfies 1 and $\mathcal{I}$ is a model of $\widehat{\mathcal{T}}$. To this end, we first extend (25) to all $\tau\mathcal{EL}(m)$ concepts $\widehat{X}$ in $\mathsf{cl}(\mathfrak{T}, \sim)$. We distinguish cases corresponding to the possible forms of $\widehat{X}$:

- $\widehat{X} = (E_i)_{\bowtie t}$. As q^v satisfies 2, we have that $q_i^v = \widehat{m}_{\sim}^{\mathcal{I}}(d_v, E_i, \mathcal{T})$. Assume that $(E_i)_{\bowtie t} \in v$. Then, $q_i^v \bowtie t$ holds because WITNESS has a non-failing run on input (v, d) with output q^v. Hence, $\widehat{m}_{\sim}^{\mathcal{I}}(d_v, E_i, \mathcal{T}) \bowtie t$ and thus $d_v \in [(E_i)_{\bowtie t}]^{\mathcal{I}}$. Conversely, suppose that $(E_i)_{\bowtie t} \notin v$. By t3), there is $(E_i)_{\bowtie' t} \in v$ such that $(E_i)_{\bowtie' t} \equiv \neg(E_i)_{\bowtie t}$. The previous arguments applied to $(E_i)_{\bowtie' t}$ yields $d \in [(E_i)_{\bowtie' t}]$, and thus $d_v \notin [(E_i)_{\bowtie t}]^{\mathcal{I}}$.
- The remaining cases are the ones considered for $\mathcal{EL}$, and can be treated similarly as for $\mathcal{EL}$.

Hence, (25) also holds for all $\tau\mathcal{EL}(m)$ concepts in $\mathsf{cl}(\mathfrak{T}, \sim)$. This, together with v satisfying t3), can be used to show that d_v satisfies 1.

Finally, as shown for $\mathcal{T}$, we also have that all $e \in \Delta^{\mathcal{I}} \setminus \{d_v\}$ satisfy the definitions in $\widehat{\mathcal{T}}$. Regarding d_v, the same argument used for $\mathcal{T}$ yields $d_v \in \alpha^{\mathcal{I}}$ iff $d_v \in \widehat{C}_\alpha^{\mathcal{I}}$ for all $\alpha \doteq \widehat{C}_\alpha \in \widehat{\mathcal{T}}$. Thus, $\mathcal{I} \models \mathfrak{T}$. □

It remains to show that Algorithm 2 is complete. The idea of the proof is to show, given a conjunction $\alpha_1 \sqcap \neg\alpha_2$ satisfiable w.r.t. $\mathfrak{T}$, how to use the finite tree model property obtained in Proposition 4.6 to guide a successful run of Algorithm 2 on input $\mathfrak{T}, \alpha_1, \alpha_2$.

Lemma 4.13. *If $\alpha_1 \sqcap \neg\alpha_2$ is satisfiable w.r.t. $\mathfrak{T}$, then Algorithm 2 answers "yes".*

Proof. Assume that $\alpha_1 \sqcap \neg\alpha_2$ is satisfiable w.r.t. $\mathfrak{T}$. By Proposition 4.6, there exists a finite tree model $\mathcal{I}$ of $\mathfrak{T}$ of depth at most $\mathfrak{d}$ whose root d_0 satisfies $\alpha_1 \sqcap \neg\alpha_2$.

Let us define for each $d \in \Delta^{\mathcal{I}}$ the set $v_d := \{X \mid d \in X^{\mathcal{I}} \text{ and } X \in \mathsf{cl}(\mathfrak{T}, \sim)\}$. In addition, let $q_i^d := \widehat{m}_{\sim}^{\mathcal{I}}(d, E_i, \mathcal{T})$ for all $1 \le i \le m$. Since $\mathcal{I}$ is a model of $\mathfrak{T}$, it is easy to see that v_d is a type for $\mathfrak{T}$ and that $\{\alpha_1, \neg\alpha_2\} \subseteq v_{d_0}$. Using these types we describe how to guide a successful run of WITNESS on $(v_{d_0}, 0)$.

Let $d \in \Delta^{\mathcal{I}}$ and l_d be the length of the path in $\mathcal{I}$ from d_0 to d. We show the following claim:

WITNESS has a successful run on (v_d, l_d) such that

$$q_i^{v_d} = \widehat{m}_{\sim}^{\mathcal{I}}(d, E_i, \mathcal{T}) \quad \text{for all } 1 \le i \le m. \tag{26}$$

The proof is by induction on $\mathfrak{d} - l_d$. Since $l_d \le \mathfrak{d}$ and v_d is a type for $\mathfrak{T}$, there is no failure at Line 2. For each $r \in \mathfrak{R}(\mathfrak{T}, \sim)$, we define a set Δ_r^d of r-successors of d as follows:

- If $r \in \mathsf{rol}(v_d)$, then for each $\widehat{C} \in v_d^+(r)$ we add to Δ_r^d an element e such that: $e \in \Delta^{\mathcal{I}}$, $(d, e) \in r^{\mathcal{I}}$ and $e \in \widehat{C}^{\mathcal{I}}$. Note that the definition of $v_d^+(r)$ yields $\exists r.\widehat{C} \in v_d$, which implies $d \in (\exists r.\widehat{C})^{\mathcal{I}}$ by definition of v_d. Hence, such an element e exists in $\mathcal{I}$.
- For each $\exists s.E_p$ occurring in $\mathcal{T}$, we select an r-successor e of d in $\mathcal{I}$ (if there is any) such that:

  $$pm(s, r)[w + (1 - w)q_p^e] = \max\{pm(s, r)[w + (1 - w)q_p^f] \mid (d, f) \in r^{\mathcal{I}}\}.$$

  We add e to Δ_r^d. Since $\mathcal{I}$ is finite, the maximum above always exists.
- Nothing else is added to Δ_r^d.

It is easy to see that $|v_d^+(r)| \le \mathsf{s}(\mathfrak{T})$. Hence, by the definition of Δ_r^d, we know that $|\Delta_r^d| \le \mathfrak{n}$. Using these sets, from the loop at Line 3 on, the run behaves as follows: it guesses ℓ_r as $|\Delta_r^d|$ and the subsequent ℓ_r subsets of $\mathsf{cl}(\mathfrak{T}, \sim)$ as the types associated to the individuals in Δ_r^d. Let us now show why this selection leads to a successful run satisfying (26).

- Let $r \in \mathsf{rol}(v_d)$ and $\exists r.\widehat{C} \in \mathsf{cl}(\mathfrak{T}, \sim)$. To see that there is no failure at Line 5, we show that $(v_d)_r^1, \ldots, (v_d)_r^{\ell_r}$ is an r-successor candidate of v_d w.r.t. $\mathsf{cl}(\mathfrak{T}, \sim)$. This is equivalent to showing that $\widehat{C} \in v_d^+(r)$ iff $\widehat{C} \in v_e$ for some $e \in \Delta_r^d$. Assume that $\widehat{C} \in v_d^+(r)$. By definition of Δ_r^d, there is $e \in \Delta_r^d$ such that $e \in \widehat{C}^{\mathcal{I}}$. Since $\widehat{C} \in \mathsf{cl}(\mathfrak{T}, \sim)$, this means that $\widehat{C} \in v_e$. Conversely, let $e \in \Delta_r^d$ be such that $\widehat{C} \in v_e$. By definition of Δ_r^d and v_e, we know that $(d, e) \in r^{\mathcal{I}}$ and $e \in \widehat{C}^{\mathcal{I}}$. This implies $d \in (\exists r.\widehat{C})^{\mathcal{I}}$, and thus $\exists r.\widehat{C} \in v_d$, which yields $\widehat{C} \in v_d^+(r)$.
- Recursive call at Line 7. Since $v_r^j = v_e$ for some $e \in \Delta_r^d$ and $l_e = l_d + 1$, the application of induction yields a successful run of WITNESS on $(v_r^j, l_d + 1)$ such that the computed tuple $q^{(v,r,j)}$ satisfies the following for all $1 \le i \le m$:

  $$q_i^{(v,r,j)} = \widehat{m}_{\sim}^{\mathcal{I}}(e, E_i, \mathcal{T}) \quad \text{for all } 1 \le i \le m. \tag{27}$$

- Execution of DEGREES. The procedure is invoked with parameters $(v_d \cap \mathsf{NC}, Q, E_i)$, where $Q = \{(r, q^{(v,r,j)}) \mid r \in \mathfrak{R}(\mathfrak{T}, \sim) \wedge 1 \le j \le \ell_r\}$. By applying Lemma 4.11 w.r.t. d and E_i, we obtain that the computed value $q_i^{v_d}$ satisfies (26). The following observations show that the lemma is applicable:

  - $\mathcal{I}$ is a model of $\mathcal{T}$, and d has finitely many role successors since $\mathcal{I}$ is finite.
  - $\mathcal{I}$ only interprets concept names in $\mathfrak{C}(\mathfrak{T}, \sim) \cup \mathsf{NC}_d^{\mathfrak{T}}$ as non-empty (see Proposition 4.6). Hence, since $\mathfrak{C}(\mathfrak{T}, \sim) \subseteq \mathsf{cl}(\mathfrak{T}, \sim)$ and $\mathsf{NC}_d^{\mathfrak{T}} \subseteq \mathsf{ND}$, the definition of v_d yields that $v_d \cap \mathsf{NC} = S_d$.
  - Let $r \in \mathsf{NR}$ be such that d has r-successors. By Proposition 4.6, $\mathcal{I}$ only interprets role names in $\mathfrak{R}(\mathfrak{T}, \sim)$ as non-empty, and thus $r \in \mathfrak{R}(\mathfrak{T}, \sim)$. Then, (27) and the definition of Δ_r^d ensure that Q has the form required for Q_i in Lemma 4.11.

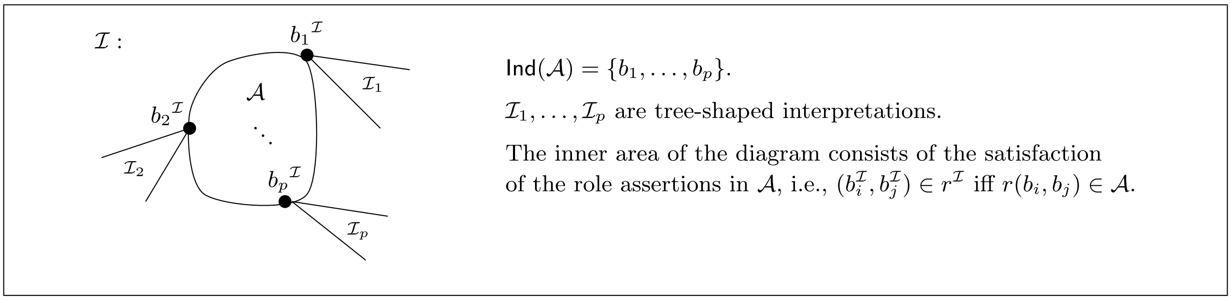


Fig. 5. Shape of a tree-like model of a satisfiable $\tau\mathcal{EL}(m)$ ABox $\mathcal{A}$.

- Finally, in Line 10, the definition of v_d ensures that $(E_i)_{\bowtie t} \in v_d$ implies $d \in [(E_i)_{\bowtie t}]^{\mathcal{I}}$. This can only be the case if $\hat{m}^{\mathcal{I}}_{\sim}(d, E_i, \mathcal{T}) \bowtie t$. Since we already know that $q_i^{v_d}$ satisfies (26), the procedure cannot fail at Line 10.

Hence, we have proved our claim. Since $l_{d_0} = 0$, we have thus shown that WITNESS has a successful run on $(v_{d_0}, 0)$. $\square$

Let us conclude by looking at the complexity of Algorithm 2. Since $\sim \in$ *simi-d*, the function $\mathfrak{s}_{pm}$ is polynomial time computable. This means that $\mathfrak{C}(\mathfrak{T}, \sim)$ and $\mathfrak{R}(\mathfrak{T}, \sim)$ can be computed in time polynomial in $\mathsf{s}(\mathfrak{T})$. As a direct consequence, $\mathsf{cl}(\mathfrak{T}, \sim)$ enjoys the same property. Testing conditions t1)-t3) on a set $v \subseteq \mathsf{cl}(\mathfrak{T}, \sim)$ and c1) on a sequence of $\ell_r \leq \mathfrak{n}$ such sets, can also be done in polynomial time. Further, each recursive call of WITNESS stores at most $|\mathfrak{R}(\mathfrak{T}, \sim)| \cdot \mathfrak{n}$ types v and their tuples q^v. Regarding the values q_i^v $(1 \leq i \leq m)$, they are of the form $u_{\mathcal{T}}(E_i) \sim D$ for some $D \in \mathcal{C}_{\mathcal{EL}}$. We can reuse the proof of Lemma 3.27 to show that q_i^v can be represented using space polynomial in $\mathsf{s}(\mathcal{T})$. In fact, Proposition 4.5 tells us that $\mathsf{rd}(u_{\mathcal{T}}(E_i)) \leq \mathsf{s}(\mathcal{T})$, and it is not hard to show (by induction on $\mathsf{rd}(u_{\mathcal{T}}(E_i))$) that the cardinalities of the sets $d_{pm}(u_{\mathcal{T}}(E_i))$ and $\mathcal{S}_{\sqcap}(u_{\mathcal{T}}(E_i))$ are polynomial in $\mathsf{s}(\mathcal{T})$. Hence, the total space required by a recursive call is polynomial in $\mathsf{s}(\mathfrak{T})$, and it is clear that DEGREES also runs in polynomial space. Thus, since WITNESS performs at most $\mathfrak{d} = \mathsf{s}(\mathfrak{T})$ nested recursive calls, Algorithm 2 is a non-deterministic polynomial space decision procedure for satisfiability of concepts of the form $\alpha_1 \sqcap \neg\alpha_2$ w.r.t. acyclic TBoxes in $\tau\mathcal{EL}(m_{\sim})$. This means that satisfiability and non-subsumption are in NPSpace. Then, by Savitch's theorem [35] we obtain the following results.

Theorem 4.14. *Let $\sim \in$ simi-d. In $\tau\mathcal{EL}(m_{\sim})$, satisfiability and subsumption w.r.t. acyclic $\tau\mathcal{EL}(m_{\sim})$ TBoxes are in PSpace.*

4.2.4. Reasoning with acyclic knowledge bases

We now turn to *reasoning w.r.t. acyclic $\tau\mathcal{EL}(m)$ knowledge bases*. In this setting, the interesting reasoning tasks are *consistency* and *instance checking*. These two problems can be reduced to consistency of knowledge bases of the form $(\mathfrak{T}, \mathcal{A} \cup \{\neg\widehat{D}(a)\})$ where $a \in \mathsf{NI}$ and $\widehat{D}$ is correctly defined w.r.t. $\mathcal{T}$. We can always assume that $a \in \mathsf{Ind}(\mathcal{A})$ by including the assertion $\top(a)$ in $\mathcal{A}$.

In the following, we describe how to adapt the previous technical results and algorithms to solve this problem. Like for satisfiable acyclic $\tau\mathcal{EL}(m_{\sim})$ TBoxes, consistent acyclic $\tau\mathcal{EL}(m_{\sim})$ knowledge bases enjoy a particular bounded model property. The difference is, however, that now the structure of the ABox in a knowledge base may require such bounded models to contain cycles, i.e., they are not guaranteed to be tree-shaped. To establish this new form of bounded model property, we introduce the notion of a tree-like model. Tree-like models have the form sketched in Fig. 5 and are formally defined as follows.

Definition 4.15. Let $\mathcal{A}$ be a $\tau\mathcal{EL}(m)$ ABox and $\mathsf{Ind}(\mathcal{A}) = \{b_1, \ldots, b_p\}$. A model $\mathcal{I}$ of $\mathcal{A}$ is called *tree-like* if there exist p mutually disjoint tree-shaped interpretations $\mathcal{I}_1, \ldots, \mathcal{I}_p$, with respective root elements $d_1, \ldots, d_p$, such that:

- $\Delta^{\mathcal{I}} := \Delta^{\mathcal{I}_1} \cup \ldots \cup \Delta^{\mathcal{I}_p}$,
- $A^{\mathcal{I}} := \bigcup_{i=1}^{p} A^{\mathcal{I}_i}$, for all $A \in \mathsf{NC}$,
- $r^{\mathcal{I}} := \{(d_i, d_j) \mid r(b_i, b_j) \in \mathcal{A} \wedge 1 \leq i, j \leq p\} \cup \bigcup_{i=1}^{p} r^{\mathcal{I}_i}$, for all $r \in \mathsf{NR}$, and
- $b_i^{\mathcal{I}} := d_i$ $(1 \leq i \leq p)$.

Given a $\tau\mathcal{EL}(m)$ ABox $\mathcal{A}$, for each $b \in \mathsf{Ind}(\mathcal{A})$ we denote as $\widehat{C}_b$ the $\tau\mathcal{EL}(m)$ concept description $\bigsqcap_{\widehat{C}(b) \in \mathcal{A}} \widehat{C}$. As a first result, we obtain a bounded tree-like model property for consistent KBs of the form $(\emptyset, \mathcal{A} \cup \{\neg\widehat{D}(a)\})$. This is stated in the following lemma, whose proof can be found in Appendix B.2.

Lemma 4.16. *Let $\sim \in$ simi-d, $\widehat{D}$ a $\tau\mathcal{EL}(m)$ concept description, $\mathcal{A}$ a $\tau\mathcal{EL}(m)$ ABox, and $a \in \mathsf{Ind}(\mathcal{A})$. If the ABox $\mathcal{A} \cup \{\neg\widehat{D}(a)\}$ is consistent in $\tau\mathcal{EL}(m_{\sim})$, then it has a tree-like model $\mathcal{I}$ such that*

- $|\Delta^{\mathcal{I}_b}| \leq \mathsf{s}(\widehat{C}_b) \cdot \mathsf{s}(\widehat{D})^u$ *for all $b \in \mathsf{Ind}(\mathcal{A})$, where $u = |\mathsf{sub}(\widehat{D})|$,*

Algorithm 3 Consistency of $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$ in $\tau\mathcal{EL}(m_\sim)$, where $\sim \in$ *simi-d*.

Input: An acyclic KB $(\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T}), \mathcal{A})$ with $\mathsf{NC}_d^{\mathcal{T}} = \{E_1, \dots, E_m\}$, $\alpha \in \mathsf{NC}_d^{\mathfrak{T}}$ and $a \in \mathsf{Ind}(\mathcal{A})$.
Output: "yes", if $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is consistent in $\tau\mathcal{EL}(m_\sim)$; "no" otherwise.

1: guess $v_b \subseteq \mathsf{cl}(\mathfrak{T}, \sim)$ for each $b \in \mathsf{Ind}(\mathcal{A})$
2: **for all** $b \in \mathsf{Ind}(\mathcal{A})$ **do**
3:     **if** (v_b is not a type for $\mathfrak{T}$) $\vee (\beta(b) \in \mathcal{A} \wedge \beta \notin v_b) \vee (b = a \wedge \neg\alpha \notin v_b)$ **then** *fail*
4:     **for all** $\exists r.\beta \in \mathsf{cl}(\mathfrak{T}, \sim)$ and $r(b, c) \in \mathcal{A}$ **do**
5:         **if** $(\neg\exists r.\beta \in v_b \wedge \beta \in v_c)$ **then** *fail*
6:     $Q_b := \text{WITNESS}^*(v_b, 0)$
7: $q^{v_b} = (q_1^{v_b}, \dots, q_m^{v_b}) := (0, \dots, 0)$ for all $b \in \mathsf{Ind}(\mathcal{A})$
8: **for all** $1 \le i \le m$ **do**
9:     **for all** $b \in \mathsf{Ind}(\mathcal{A})$ **do**
10:         $q_i^{v_b} := \text{DEGREES}(v_b \cap \mathsf{NC}, Q_b \cup \{(r, q^{v_c}) \mid r(b, c) \in \mathcal{A}\}, E_i)$
11: **return** *success* iff $q_i^{v_b} \bowtie t$ holds for all $(E_i)_{\bowtie t} \in v_b$ and $b \in \mathsf{Ind}(\mathcal{A})$

- *$\mathcal{I}$ only interprets concept names in $\mathfrak{C}(\mathcal{A}, \widehat{D}, \sim) := \{B \in \mathsf{NC} \mid B \in \mathfrak{s}_{pm}(A) \wedge A \in \mathsf{sig}(\mathcal{A}) \cup \mathsf{sig}(\widehat{D})\}$ and role names in $\mathfrak{R}(\mathcal{A}, \widehat{D}, \sim) := \{s \in \mathsf{NR} \mid s \in \mathfrak{s}_{pm}(r) \wedge r \in \mathsf{sig}(\mathcal{A}) \cup \mathsf{sig}(\widehat{D})\}$ as non-empty.*

Since $\sum_{b \in \mathsf{Ind}(\mathcal{A})} \mathsf{s}(\widehat{C}_b) \le \mathsf{s}(\mathcal{A})$, the lemma shows that every consistent ABox $\mathcal{A} \cup \{\neg\widehat{D}(a)\}$ has a model of size at most $\mathsf{s}(\mathcal{A}) \cdot \mathsf{s}(\widehat{D})^u$. This observation can be used to obtain a non-deterministic procedure for deciding consistency and non-instance in $\tau\mathcal{EL}(m_\sim)$: first guess an interpretation $\mathcal{I}$ of size at most $\mathsf{s}(\mathcal{A}) \cdot \mathsf{s}(\widehat{D})^u$ over the signature $\mathfrak{C}(\mathcal{A}, \widehat{D}, \sim) \cup \mathfrak{R}(\mathcal{A}, \widehat{D}, \sim)$, and then test whether $\mathcal{I} \models \mathcal{A}$ and $a^{\mathcal{I}} \notin \widehat{D}^{\mathcal{I}}$. Both checks can be done in polynomial time in the size of $\mathcal{A}, \widehat{D}$ and $\mathcal{I}$ (see Corollary 2.19 and Theorem 2.18). Since consistency corresponds to the particular case where $\widehat{D} = A \sqcap A_{<1}$ (and thus u is a constant), this yields an NP-algorithm for deciding consistency. Regarding non-instance, this algorithm runs in non-deterministic exponential time w.r.t. *combined complexity*. A better (PSpace) upper bound is obtained from the decision procedure described next for arbitrary acyclic TBoxes.

Theorem 4.17. *Let $\sim \in$ simi-d. Without a TBox, the consistency problem is in NP in $\tau\mathcal{EL}(m_\sim)$.*

In the presence of a non-empty TBox, we can restrict the attention to KBs of the form $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$, where all concept assertions in $\mathcal{A}$ are of the form $\beta(b)$ and α and the βs are defined concepts in $\mathsf{NC}_d^{\mathfrak{T}}$. We use unfolding to extend Lemma 4.16 to the case where $\mathfrak{T} \neq \emptyset$. The proof of the next lemma can be found in Appendix B.2.

Lemma 4.18. *Let $\sim \in$ simi-d, $(\mathfrak{T}, \mathcal{A})$ an acyclic $\tau\mathcal{EL}(m)$ KB, $\alpha \in \mathsf{NC}_d^{\mathfrak{T}}$ and $a \in \mathsf{Ind}(\mathcal{A})$. In $\tau\mathcal{EL}(m_\sim)$, if $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is consistent, then it has a finite tree-like model $\mathcal{I}$ such that for all $b \in \mathsf{Ind}(\mathcal{A})$:*

1. *$\mathcal{I}_b$ is a model of $\mathfrak{T}$, has depth at most $\mathsf{s}(\mathfrak{T})$, and it only interprets symbols in $\mathfrak{C}(\mathfrak{T}, \sim)$, $\mathfrak{R}(\mathfrak{T}, \sim)$ and $\mathsf{NC}_d^{\mathfrak{T}}$ as non-empty.*
2. *For all $X \in \mathsf{cl}(\mathfrak{T}, \sim)$: $b^{\mathcal{I}} \in X^{\mathcal{I}}$ iff $b^{\mathcal{I}} \in X^{\mathcal{I}_b}$.*

Based on this lemma, Algorithm 3 uses a slight adaptation (called WITNESS*) of the procedure WITNESS introduced in Section 4.2.3 to search for a tree-like model $\mathcal{I}$ of $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$. Basically, WITNESS is modified as follows:

- when invoked with $(v_b, 0)$, a non-failing run returns the set of all pairs $(r, q^{(v_b, r, j)})$ instead of the tuple q^{v_b}, i.e., in Algorithm 3, Q_b gets assigned the set $\{(r, q^{(v_b, r, j)}) \mid r \in \mathfrak{R}(\mathfrak{T}, \sim) \wedge 1 \le j \le \ell_r\}$.

This modification is made for the following reason. A tuple q^{v_b} computed by WITNESS contains the values $q_i^{v_b} = \widehat{m}_\sim^{\mathcal{I}_b}(b^{\mathcal{I}}, E_i, \mathcal{T})$, where $\mathcal{I}_b$ is the interpretation built in Lemma 4.12. However, this computation does not take into account the possible additional successors of $b^{\mathcal{I}}$ in the combined tree-like interpretation $\mathcal{I}$ (as required by assertions $r(b, c) \in \mathcal{A}$). Hence, it need not be the case that $q_i^{v_b} = \widehat{m}_\sim^{\mathcal{I}}(b^{\mathcal{I}}, E_i, \mathcal{T})$. To rectify this, we assume that $E_1, \dots, E_m$ is a post-order enumeration of $\mathsf{NC}_d^{\mathcal{T}}$ w.r.t. the dependency relation $\rightarrow$ (see Definition 2.1). The loop at Line 8 then follows this enumeration to compute the values $q_i^{v_b}$. The following lemma shows that, with the right parameters, this computation produces the intended values.

Lemma 4.19. *Let $\sim \in$ simi-d, $\mathcal{T} \in \mathfrak{S}_{\mathsf{D}}$ normalized and reduced, $\mathcal{A}$ an $\mathcal{EL}$ ABox, and $\mathcal{I}$ a finite tree-like model of $(\mathcal{T}, \mathcal{A})$. Define, for all $b \in \mathsf{Ind}(\mathcal{A})$ and all $f \in \Delta^{\mathcal{I}}$, the set S_b and the rational number q_i^f as:*

$$S_b := \{A \mid b^{\mathcal{I}} \in A^{\mathcal{I}} \wedge A \in \mathsf{NC}\} \quad \text{and} \quad q_i^f := \widehat{m}_\sim^{\mathcal{I}}(f, E_i, \mathcal{T}).$$

Further, for each $b \in \mathsf{Ind}(\mathcal{A})$, let D_b be a set such that

- *$D_b \subseteq \{(r, (q_1^f, \dots, q_m^f)) \mid f \in \Delta^{\mathcal{I}_b} \wedge (b^{\mathcal{I}}, f) \in r^{\mathcal{I}_b} \wedge r \in \mathsf{NR}\}$, and*
- *for all $1 \le i \le m$, $\exists s.E_p \in top(C_{E_i})$ and $r \in \mathsf{NR}$, if $b^{\mathcal{I}}$ has r-successors in $\mathcal{I}_b$ then D_b contains a tuple $(r, (q_1^e, \dots, q_m^e))$ such that $(b^{\mathcal{I}}, e) \in r^{\mathcal{I}_b}$ and*

$$pm(s,r)[w+(1-w)q_p^e] = \max\{pm(s,r)[w+(1-w)q_p^f] \mid (d,f) \in r^{\mathcal{I}_b}\}.$$

If $v_b \cap \mathsf{NC} = S_b$ and $Q_b = D_b$ for all $b \in \mathsf{Ind}(\mathcal{A})$, then the loop at Line 8 of Algorithm 3 computes tuples q^{v_b} such that $q_i^{v_b} = \hat{m}_{\sim}^{\mathcal{I}}(b^{\mathcal{I}}, E_i, \mathcal{T})$ for all $1 \le i \le m$.

Proof. We show, by induction on $i, 1 \le i \le m$, that for all $b \in \mathsf{Ind}(\mathcal{A})$ the value $q_i^{v_b}$ computed at iteration i is equal to $\hat{m}_{\sim}^{\mathcal{I}}(b^{\mathcal{I}}, E_i, \mathcal{T})$. This proves the lemma, since $q_i^{v_b}$ is assigned only once and $\mathsf{NC}_d^{\mathcal{T}} = \{E_1, \dots, E_m\}$.

Consider the set $Q_{b,\mathcal{A}} := \{(r,(q_1^e,\dots,q_m^e)) \mid r(b,c) \in \mathcal{A} \wedge e = c^{\mathcal{I}}\}$. Further, let $Q_{b,\mathcal{A}}^i$ be a variant of $Q_{b,\mathcal{A}}$ such that $|Q_{b,\mathcal{A}}^i| = |Q_{b,\mathcal{A}}|$, and each tuple in $Q_{b,\mathcal{A}}$ has a corresponding one $(r,(x_1^e,\dots,x_m^e))$ in $Q_{b,\mathcal{A}}^i$ satisfying that $x_j^e = q_j^e$ for all $1 \le j < i$. The following observations hold for these two sets:

- $Q_{b,\mathcal{A}}$ covers all role successors of $b^{\mathcal{I}}$ not in $\Delta^{\mathcal{I}_b}$. This, together with the definition of D_b and the form of $\mathcal{I}$, implies that $Q_b \cup Q_{b,\mathcal{A}}$ has the form required in Lemma 4.11 for a set Q_i. Hence, $\textsc{degrees}(v_b \cap \mathsf{NC}, Q_b \cup Q_{b,\mathcal{A}}, E_i)$ computes the value $\hat{m}_{\sim}^{\mathcal{I}}(b^{\mathcal{I}}, E_i, \mathcal{T})$.
- A computation of $\textsc{degrees}(\_,\_,E_i)$ only uses values q_p such that $\exists s.E_p \in top(C_{E_i})$. Note that $p < i$, since $\exists s.E_p \in top(C_{E_i})$ implies $E_i \to E_p$. Hence, by definition of $Q_{b,\mathcal{A}}^i$ and the observation about $Q_{b,\mathcal{A}}$, it follows that $\textsc{degrees}(v_b \cap \mathsf{NC}, Q_b \cup Q_{b,\mathcal{A}}^i, E_i)$ also computes the value $\hat{m}_{\sim}^{\mathcal{I}}(b^{\mathcal{I}}, E_i, \mathcal{T})$.

Hence, to complete the proof, it is enough to show that, at iteration i, the set $\{(r, q^{v_c}) \mid r(b,c) \in \mathcal{A}\}$ (see Line 10) is of the form $Q_{b,\mathcal{A}}^i$. The induction hypothesis yields that, at iteration i, the values $q_j^{v_c}$ are equal to $\hat{m}_{\sim}^{\mathcal{I}}(c^{\mathcal{I}}, E_j, \mathcal{T})$ for all $j, 1 \le j < i$. Thus, the set $\{(r, q^{v_c}) \mid r(b,c) \in \mathcal{A}\}$ is indeed as required. □

We are now ready to show that Algorithm 3 is correct. To show soundness, given a successful run of the algorithm on input $(\mathfrak{T}, \mathcal{A}), \alpha$ and a, we first apply Lemma 4.12 to the corresponding successful runs of $\textsc{witness}^*$ on each pair $(v_b, 0)$ to obtain interpretations $\mathcal{I}_b$ satisfying all concept assertions $\beta(b) \in \mathcal{A} \cup \{\neg\alpha(a)\}$. These interpretations are then combined into a model $\mathcal{J}$ of $\mathfrak{T}$ satisfying the role assertions in $\mathcal{A}$. Finally, we show that this combination preserves the satisfaction of the concept assertions in $\mathcal{A}$, and hence, $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is consistent. As for the completeness of the algorithm, the idea of the proof is to apply Lemma 4.18 to obtain a tree-like model $\mathcal{I}$ of a consistent knowledge base $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$, and then use $\mathcal{I}$ to guide a successful run of the algorithm on input $(\mathfrak{T}, \mathcal{A}), \alpha, a$.

Proposition 4.20. *Algorithm 3 answers "yes" iff $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is consistent.*

Proof. ($\Rightarrow$) Suppose Algorithm 3 has a successful run. Then, for all $b \in \mathsf{Ind}(\mathcal{A})$, $\textsc{witness}^*$ has a successful run on $(v_b, 0)$. The proof of Lemma 4.12 shows the existence of finite and tree-shaped models $\mathcal{I}_b$ of $\mathfrak{T}$ such that:

1. $d_b \in \bigcap_{X \in v_b} X^{\mathcal{I}_b}$, and
2. the set Q_b returned by $\textsc{witness}^*$ is of the form

$$Q_b := \{(r,(q_1^e,\dots,q_m^e)) \mid e \in \Delta^{\mathcal{I}_b} \wedge (d_b, e) \in r^{\mathcal{I}_b} \wedge r \in \mathsf{NR}\},$$

where $q_i^e = \hat{m}_{\sim}^{\mathcal{I}_b}(e, E_i, \mathcal{T})$ for all $1 \le i \le m$.

We define the interpretation $\mathcal{I}$ as the disjoint union of the primitive interpretations underlying each $\mathcal{I}_b$, where $\cdot^{\mathcal{I}}$ extends $r^{\mathcal{I}}$ with $\{(d_b, d_c) \mid r(b,c) \in \mathcal{A}\}$. By Proposition 2.29, $\mathcal{I}$ can be uniquely extended into a model $\mathcal{J}$ of $\mathfrak{T}$. Let us define $b^{\mathcal{J}} := d_b$ for all $b \in \mathsf{Ind}(\mathcal{A})$. To conclude the proof, it remains to show that $\mathcal{J}$ satisfies $\mathcal{A} \cup \{\neg\alpha(a)\}$. Clearly, $\mathcal{J}$ satisfies all role assertions in $\mathcal{A}$. To see that it also satisfies the concept assertions, we first need to show some properties satisfied by $\mathcal{J}$.

By definition, we have that $\cdot^{\mathcal{I}}$ and $\cdot^{\mathcal{J}}$ coincide on $\mathsf{NC}_{pr}^{\mathfrak{T}}$ and NR. Hence, since each $\mathcal{I}_b$ is tree-shaped, the following holds for all $e \in \Delta^{\mathcal{I}_b} \setminus \{d_b\}$ and $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$:

$$e \in C^{\mathcal{I}_b} \text{ iff } e \in C^{\mathcal{J}}.$$

A consequence of this and the definition of $\hat{m}$ is that for all $i, 1 \le i \le m$, the following holds:

$$\hat{m}_{\sim}^{\mathcal{I}_b}(e, E_i, \mathcal{T}) = \hat{m}_{\sim}^{\mathcal{J}}(e, E_i, \mathcal{T}).$$

Two consequences can be derived from this equality. First, one can see that w.r.t. $\mathcal{J}$, the sets Q_b satisfy the properties imposed on D_b in Lemma 4.19. Hence, since $\mathcal{J}$ is a finite, tree-like model of $(\mathfrak{T}, \{r(b,c) \in \mathcal{A}\})$, we can apply Lemma 4.19 to obtain:

$$q_i^{v_b} = \hat{m}_{\sim}^{\mathcal{J}}(b^{\mathcal{J}}, E_i, \mathcal{T}). \tag{28}$$

Secondly, since $\mathcal{I}_b$ is tree-shaped and $\cdot^{\mathcal{I}}$ and $\cdot^{\mathcal{J}}$ coincide on $\mathsf{NC}_{pr}^{\mathfrak{T}}$ and NR, we obtain the following for all $e \in \Delta^{\mathcal{I}_b} \setminus \{d_b\}$ and $\tau\mathcal{EL}(m)$ concepts $\widehat{C}$ defined over the signature $\mathsf{NC}_{pr}^{\mathfrak{T}} \cup \mathsf{NR}$:

$$e \in \widehat{C}^{\mathcal{I}_b} \text{ iff } e \in \widehat{C}^{\mathcal{J}}. \tag{29}$$

We can now show that $\mathcal{J}$ satisfies all concept assertions in $\mathcal{A}$. Note that $\neg\alpha \in \nu_a$ and $\beta \in \nu_b$ for all $\beta(b) \in \mathcal{A}$, because a successful run does not fail at Line 3. In addition, by our assumption on the shape of the ABox, $a \in \mathsf{Ind}(\mathcal{A})$, $\alpha \in \mathsf{NC}_d^{\mathfrak{T}}$ and $\beta \in \mathsf{NC}_d^{\mathfrak{T}}$ for all $\beta(b) \in \mathcal{A}$. Hence, it is enough to show that the following holds for all $b \in \mathsf{Ind}(\mathcal{A})$ and $\beta \in \mathsf{NC}_d^{\mathfrak{T}}$:

$$\beta \in \nu_b \text{ iff } b^{\mathcal{J}} \in \beta^{\mathcal{J}}. \tag{30}$$

Let $\beta \doteq \widehat{C}_\beta \in \widehat{\mathcal{T}} \cup \mathcal{T}$. We have that $\beta \in \nu_b$ iff $\widehat{C}_\beta \in \nu_b$ iff $\mathsf{At} \in \nu_b$ for all $\mathsf{At} \in top(\widehat{C}_\beta)$, since ν_b is a type for $\mathfrak{T}$ and it satisfies t1). In addition, $b^{\mathcal{J}} \in \beta^{\mathcal{J}}$ iff $b^{\mathcal{J}} \in \mathsf{At}^{\mathcal{J}}$ for all $\mathsf{At} \in top(\widehat{C}_\beta)$. Hence, (30) holds if the following holds for all $b \in \mathsf{Ind}(\mathcal{A})$, $\beta \in \mathsf{NC}_d^{\mathfrak{T}}$ and $\mathsf{At} \in top(\widehat{C}_\beta)$:

$$\mathsf{At} \in \nu_b \text{ iff } b^{\mathcal{J}} \in \mathsf{At}^{\mathcal{J}}. \tag{31}$$

We prove (31) by induction on the partial order induced by $\rightarrow$ over $\mathsf{NC}_d^{\mathfrak{T}}$. Based on the assumption that $\mathfrak{T}$ is normalized, we distinguish the cases corresponding to the possible forms of At:

- $\mathsf{At} = A \in \mathsf{NC}_{pr}^{\mathfrak{T}}$. Since ν_b satisfies t3), by 1. (see beginning of this proof) we have that $A \in \nu_b$ iff $d_b \in A^{\mathcal{I}_b}$. Moreover, the construction of $\mathcal{J}$ yields $d_b \in A^{\mathcal{I}_b}$ iff $b^{\mathcal{J}} \in A^{\mathcal{J}}$. Hence, $A \in \nu_b$ iff $b^{\mathcal{J}} \in A^{\mathcal{J}}$.
- $\mathsf{At} = (E_i)_{\bowtie t}$. If $(E_i)_{\bowtie t} \in \nu_b$, non-failure at Line 11 implies that $q_i^{\nu_b} \bowtie t$ holds. Hence, since $q_i^{\nu_b} = \widehat{m}_{\sim}^{\mathcal{J}}(b^{\mathcal{J}}, E_i, \mathcal{T})$ by (28), Definition 2.25 tells us that $b^{\mathcal{J}} \in [(E_i)_{\bowtie t}]^{\mathcal{J}}$. Conversely, assume that $(E_i)_{\bowtie t} \notin \nu_b$. Since ν_b satisfies t3), there is $(E_i)_{\bowtie' t} \in \nu_b$ such that $(E_i)_{\bowtie' t} \equiv \neg(E_i)_{\bowtie t}$. Again, non-failure at Line 11, (28) and Definition 2.25 yield $b^{\mathcal{J}} \in [(E_i)_{\bowtie' t}]^{\mathcal{J}}$. Hence, $b^{\mathcal{J}} \notin [(E_i)_{\bowtie t}]^{\mathcal{J}}$.
- $\mathsf{At} = \exists r.\gamma$, where $\gamma \in \mathsf{NC}_d^{\mathfrak{T}}$. As above, we have that $\exists r.\gamma \in \nu_b$ iff $d_b \in (\exists r.\gamma)^{\mathcal{I}_b}$. Suppose $\exists r.\gamma \in \nu_b$. Since $d_b \in (\exists r.\gamma)^{\mathcal{I}_b}$ and $\mathcal{I}_b$ is tree-shaped, there is $e \in \Delta^{\mathcal{I}_b} \setminus \{d_b\}$ such that $(d_b, e) \in r^{\mathcal{I}_b}$ and $e \in \gamma^{\mathcal{I}_b}$. By applying Lemma 2.28 and (29), we obtain $e \in u_{\mathfrak{T}}(\gamma)^{\mathcal{I}_b}$, $e \in u_{\mathfrak{T}}(\gamma)^{\mathcal{J}}$ and $e \in \gamma^{\mathcal{J}}$. Hence, $b^{\mathcal{J}} \in (\exists r.\gamma)^{\mathcal{J}}$.
  Conversely, if $b^{\mathcal{J}} \in (\exists r.\gamma)^{\mathcal{J}}$, then there is $e \in \Delta^{\mathcal{J}}$ such that $(b^{\mathcal{J}}, e) \in r^{\mathcal{J}}$ and $e \in \gamma^{\mathcal{J}}$. If $e \in \Delta^{\mathcal{I}_b} \setminus \{d_b\}$, then Lemma 2.28 and (29) again yield $e \in \gamma^{\mathcal{I}_b}$. Hence, $d_b \in (\exists r.\gamma)^{\mathcal{I}_b}$ and $\exists r.\gamma \in \nu_b$. Otherwise, $e = c^{\mathcal{I}}$ for $c \in \mathsf{Ind}(\mathcal{A})$. Since $\beta \rightarrow \gamma$, induction yields $\gamma \in \nu_c$. Hence, non-failure at Line 5 implies that $\neg\exists r.\gamma \notin \nu_b$. The latter and t3) yield that $\exists r.\gamma \in \nu_b$.

This completes the proof of the *only-if* direction of the lemma.

($\Leftarrow$) Assume that $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is consistent. We use the tree-like model $\mathcal{I}$ from Lemma 4.18 to guide a successful run of Algorithm 3. To start with, the sets ν_b are guessed at Line 1 as $\{X \mid b^{\mathcal{I}} \in X^{\mathcal{I}} \wedge X \in \mathsf{cl}(\mathfrak{T}, \sim)\}$. This, together with $\mathcal{I} \models (\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$, ensures that ν_b is a type for $\mathfrak{T}$, and:

- $\neg\alpha \in \nu_a$, $\beta \in \nu_b$ for all $\beta(b) \in \mathcal{A}$, and
- $\beta \in \nu_c$ implies $\exists r.\beta \in \nu_b$, for all $\exists r.\beta \in \mathsf{cl}(\mathfrak{T}, \sim)$ and $r(b, c) \in \mathcal{A}$.

Hence, there is no failure at Lines 3 and 5.

To continue, for each $b \in \mathsf{Ind}(\mathcal{A})$, Lemma 4.18 tells us that $\mathcal{I}_b$ is a finite and tree-shaped model of $\mathfrak{T}$, has depth at most $\mathsf{s}(\mathfrak{T})$ and interprets only symbols in $\mathfrak{C}(\mathfrak{T}, \sim)$, $\mathfrak{R}(\mathfrak{T}, \sim)$ and $\mathsf{NC}_d^{\mathfrak{T}}$ as non-empty. The tree shape of $\mathcal{I}_b$ and the definition of $\widehat{m}$ yield that:

$$\widehat{m}_{\sim}^{\mathcal{I}}(e, E_i, \mathcal{T}) = \widehat{m}_{\sim}^{\mathcal{I}_b}(e, E_i, \mathcal{T}) \quad \text{ for all } e \in \Delta^{\mathcal{I}_b} \setminus \{b^{\mathcal{I}}\} \text{ and } 1 \leq i \leq m. \tag{32}$$

Further, by 2. of Lemma 4.18 we obtain that $\nu_b = \{X \mid b^{\mathcal{I}} \in X^{\mathcal{I}_b} \wedge X \in \mathsf{cl}(\mathfrak{T}, \sim)\}$. Hence, we can apply the proof of Lemma 4.13 to ν_b and $\mathcal{I}_b$ to obtain a successful run of WITNESS* on $(\nu_b, 0)$ such that Q_b satisfies:

- $Q_b \subseteq \{(r, (q_1^e, \ldots, q_m^e)) \mid (b^{\mathcal{I}}, e) \in r^{\mathcal{I}_b} \wedge r \in \mathsf{NR}\}$, where $q_i^e = \widehat{m}_{\sim}^{\mathcal{I}_b}(e, E_i, \mathcal{T})$ $(1 \leq i \leq m)$.
- For each $\exists s.E_p$ occurring in $\mathcal{T}$ and $r \in \mathsf{NR}$, if $b^{\mathcal{I}}$ has an r-successor in $\mathcal{I}_b$ then Q_b contains a tuple $(r, (q_1^e, \ldots, q_m^e))$ such that $(b^{\mathcal{I}}, e) \in r^{\mathcal{I}_b}$ and

$$pm(s, r)[w + (1 - w)q_p^e] = \max\{pm(s, r)[w + (1 - w)q_p^f] \mid (b^{\mathcal{I}}, f) \in r^{\mathcal{I}_b}\}.$$

The properties satisfied by Q_b and (32) ensure that Q_b has the form required for D_b in Lemma 4.19. Therefore, the application of Lemma 4.19 yields $q_i^{\nu_b} = \widehat{m}_{\sim}^{\mathcal{I}}(b^{\mathcal{I}}, E_i, \mathcal{T})$ for all $i, 1 \leq i \leq m$. Since $(E_i)_{\bowtie t} \in \nu_b$ implies that $b^{\mathcal{I}} \in [(E_i)_{\bowtie t}]^{\mathcal{I}}$ (by definition of ν_b), it follows from Definition 2.25 that $q_i^{\nu_b} \bowtie t$ holds for all $(E_i)_{\bowtie t} \in \nu_b$. Thus, no failure occurs at Line 11, and Algorithm 3 answers "yes". □

Let us now look at the complexity of Algorithm 3. In the same way as for WITNESS, we can show that a run of WITNESS* uses space polynomial in $\mathsf{s}(\mathfrak{T})$. This implies that the returned sets Q_b are of polynomial size. In addition, we also know that the tuples in Q_b and each tuple q^{ν_b} can be represented using space polynomial in $\mathsf{s}(\mathfrak{T})$. Hence, since only one set ν_b, one set Q_b and one tuple q^{ν_b} is stored for each $b \in \mathsf{Ind}(\mathcal{A})$, it is not hard to see that Algorithm 3 runs in non-deterministic polynomial space in the size of the input. This, together with Proposition 4.20, yields the following results.

Theorem 4.21. *Let $\sim \in$ simi-d. In $\tau\mathcal{EL}(m_\sim)$, consistency and instance checking w.r.t. acyclic $\tau\mathcal{EL}(m)$ knowledge bases are in PSpace.*

4.3. Lower bounds

The PSpace lower bounds are obtained by reductions from the satisfiability problem of concept descriptions in the DL $\mathcal{ALC}$. Concept satisfiability in $\mathcal{ALC}$ is PSpace-hard even for the set of $\mathcal{ALC}$ concept descriptions defined over a single role name [36]. Our reductions are based on an alternative proof of this result, given in [2], that uses a PSpace-complete decision problem about *finite* Boolean *games* [37]. The $\mathcal{ALC}$ concept description C_G constructed in [2] consists of the conjunction of the four concepts:

- $C_1 = \bigsqcap_{i\in\{1,3,\dots,n-1\}} \forall r^i.(\exists r.\neg P_{i+1} \sqcap \exists r.P_{i+1})$.
- $C_2 = \bigsqcap_{i\in\{0,2,\dots,n-2\}} \forall r^i.\exists r.\top$.
- $C_3 = \bigsqcap_{1\le j\le i<n} \forall r^i.((P_j \to \forall r.P_j) \sqcap (\neg P_j \to \forall r.\neg P_j))$.
- $C_4 = \forall r^n.\phi^*$,

where n is even, $P_1,\dots,P_n$ are concept names, ϕ is a propositional formula in CNF with variables $p_1,\dots,p_n$ and ϕ^* is the $\mathcal{ALC}$ concept description that results from replacing in ϕ each variable p_i with P_i, $\wedge$ with $\sqcap$ and $\vee$ with $\sqcup$. In other words, ϕ^* is of the form $\phi_1^* \sqcap \cdots \sqcap \phi_m^*$ where each ϕ_j^* is a disjunction $X_1 \sqcup \cdots \sqcup X_{\ell_j}$ of concepts in $\{P_i,\neg P_i \mid 1\le i\le n\}$.

Theorem 4.22 *([2]). Satisfiability of concept descriptions of the form C_G is PSpace-hard.*

We first show PSpace-hardness of concept satisfiability w.r.t. acyclic TBoxes in $\tau\mathcal{EL}(m_\sim)$. The reduction also yields NP/coNP-hardness of reasoning w.r.t. $\mathfrak{T}=\emptyset$. Afterwards, we show that the instance problem is PSpace-hard even w.r.t. $\mathfrak{T}=\emptyset$.

4.3.1. PSpace-hardness of concept satisfiability

To reduce satisfiability of $C_G = C_1 \sqcap C_2 \sqcap C_3 \sqcap C_4$ to satisfiability w.r.t. acyclic TBoxes in $\tau\mathcal{EL}(m_\sim)$, we first define an acyclic $\mathcal{ALC}$ TBox $\mathcal{T}_{C_G}$ with defined concept names $\alpha_0,\dots,\alpha_n$ such that C_G is satisfiable iff α_0 is satisfiable w.r.t. $\mathcal{T}_{C_G}$. This simplifies the use of non-$\mathcal{EL}$ operators in $C_1,\dots,C_4$.

To each concept name P_i we associate a concept name $\bar{P}_i \in \mathsf{NC}$ not occurring in C_G, which is meant to represent $\neg P_i$. The complementary semantics of $P_i,\neg P_i$ can be captured with the concept $\neg(P_i \sqcap \bar{P}_i) \sqcap (P_i \sqcup \bar{P}_i)$. This is encoded in $\mathcal{T}_{C_G}$, as required by C_1, C_2, C_3, C_4, with the help of the following abbreviations:

- $N_1 \doteq \neg(P_1 \sqcap \bar{P}_1)$ and $N_i \doteq N_{i-1} \sqcap \neg(P_i \sqcap \bar{P}_i)$, $1 < i \le n$.
- $D_1 \doteq P_1 \sqcup \bar{P}_1$ and $D_i \doteq D_{i-1} \sqcap (P_i \sqcup \bar{P}_i)$, $1 < i \le n$.

Once we have this correspondence between $\neg P_i$ and $\bar{P}_i$, and taking into account that $P_j \to \forall r.P_j$ is equivalent to $\neg(P_j \sqcap \exists r.\neg P_j)$, we use the following abbreviations to simulate the conjuncts in C_3:

- $G_1 \doteq \neg(P_1 \sqcap \exists r.\bar{P}_1) \sqcap \neg(\bar{P}_1 \sqcap \exists r.P_1)$ and $G_i \doteq G_{i-1} \sqcap \neg(P_i \sqcap \exists r.\bar{P}_i) \sqcap \neg(\bar{P}_i \sqcap \exists r.P_i)$, $1 < i < n$.

Finally, to appropriately translate ϕ^*, we define the transformation $\xi(C)$ for $\mathcal{ALC}$ concept descriptions C as the result of replacing in C every occurrence of $\neg P_i$ by $\bar{P}_i$. Let H_i denote $N_i \sqcap D_i \sqcap G_i$ $(0 < i < n)$ and $\top$ for $i = 0$. Then, $\mathcal{T}_{C_G}$ consists of all the definitions for N_i, D_i, G_i together with:

$$\alpha_i \doteq \begin{cases} \exists r.\alpha_{i+1} \sqcap H_i & i \in \{0,2,4,\dots,n-2\},\\ \exists r.(\alpha_{i+1} \sqcap P_{i+1}) \sqcap \exists r.(\alpha_{i+1} \sqcap \bar{P}_{i+1}) \sqcap H_i & i \in \{1,3,\dots,n-1\},\\ \xi(\phi^*) \sqcap N_n \sqcap D_n & i = n.\end{cases}$$

Lemma 4.23. *C_G is satisfiable iff α_0 is satisfiable w.r.t. $\mathcal{T}_{C_G}$.*

Proof. $(\Rightarrow)$ Assume that C_G is satisfiable. This means that there is an interpretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$ such that $d \in C_G^{\mathcal{I}}$. We transform $\mathcal{I}$ into an interpretation $\mathcal{I}'$ by defining $\bar{P}_i^{\mathcal{I}'} := \{d \in \Delta^{\mathcal{I}} \mid d \notin P_i^{\mathcal{I}}\}$ for all $1 \le i \le n$. This means that:

$$e \in P_i^{\mathcal{I}} \text{ iff } e \in P_i^{\mathcal{I}'} \text{ iff } e \notin \bar{P}_i^{\mathcal{I}'} \quad \text{for all } e \in \Delta^{\mathcal{I}}. \tag{33}$$

Moreover, since $\mathcal{T}_{C_G}$ is acyclic, $\mathcal{I}'$ can be extended to interpret all α_i, N_i, D_i, G_i into a model $\mathcal{J}$ of $\mathcal{T}_{C_G}$. Our goal now is to show that $d \in \alpha_0^{\mathcal{J}}$.

For all $0 \le i \le n$, let $\mathcal{I}(d,i)$ denote the set of elements in $\Delta^{\mathcal{I}}$ that are reachable from d through a path of length i labeled with the role name r. We prove the following claim by induction on $i, 0 \le i \le n$:

$$e \in \mathcal{I}(d, n-i) \text{ implies that } e \in \alpha_{n-i}^{\mathcal{J}}.$$

Induction Base. For $i = 0$, since $d \in C_4^{\mathcal{I}}$, we have that $e \in \mathcal{I}(d,n)$ implies $e \in \phi^{*\mathcal{I}}$. Consequently, by the definition of ξ and (33), it follows that $e \in \alpha_n^{\mathcal{J}}$.

In the *Induction Step*, we consider two cases:

- $n - i$ is even. Since $d \in (C_2 \sqcap C_3)^{\mathcal{I}}$, every $e \in \mathcal{I}(d, n-i)$ satisfies:
  - $e \in (\exists r.\top)^{\mathcal{I}}$, and
  - $e \in [(P_j \to \forall r.P_j) \sqcap (\neg P_j \to \forall r.\neg P_j)]^{\mathcal{I}}$ for all $j, 1 \le j \le n - i$, if $i \ne n$.

  Let f be such that $(e,f) \in r^{\mathcal{I}}$. This implies that $f \in \mathcal{I}(d, n-(i-1))$. Hence, induction yields $f \in \alpha_{n-i+1}^{\mathcal{J}}$, and thus $e \in (\exists r.\alpha_{n-i+1})^{\mathcal{J}}$. By (33), e also belongs to N_{n-i} and D_{n-i} in $\mathcal{J}$. Further, for $i \ne n$, we also have $e \in G_{n-i}^{\mathcal{J}}$. Thus, $e \in (\exists r.\alpha_{n-i+1} \sqcap H_{n-i})^{\mathcal{J}} = \alpha_{n-i}^{\mathcal{J}}$.
- $n - i$ is odd. Since $d \in C_1^{\mathcal{I}}$, there exist two elements e_1 and e_2 such that $(e,e_1) \in r^{\mathcal{I}}$, $(e,e_2) \in r^{\mathcal{I}}$, $e_1 \in P_{n-i+1}^{\mathcal{I}}$ and $e_2 \notin P_{n-i+1}^{\mathcal{I}}$. Since $\{e_1, e_2\} \subseteq \mathcal{I}(d, n-(i-1))$, the application of induction yields that $\{e_1, e_2\} \subseteq \alpha_{n-i+1}^{\mathcal{J}}$. Moreover, (33) implies that $e_1 \in P_{n-i+1}^{\mathcal{J}}$, $e_2 \in \bar{P}_{n-i+1}^{\mathcal{J}}$ and $e \in (N_{n-i} \sqcap D_{n-i})^{\mathcal{J}}$. Finally, $e \in G_{n-i}^{\mathcal{J}}$ follows again from the fact that $d \in C_3^{\mathcal{I}}$ and $e \in \mathcal{I}(d, n-i)$. Hence, $e \in \alpha_{n-i}^{\mathcal{J}}$.

Thus, since $d \in \mathcal{I}(d, 0)$, we have shown that $d \in \alpha_0^{\mathcal{J}}$.

($\Leftarrow$) Assume that α_0 is satisfiable w.r.t. $\mathcal{T}_{C_G}$. Then, there is a tree model $\mathcal{I}$ of $\mathcal{T}_{C_G}$ with root element $d \in \Delta^{\mathcal{I}}$ such that $d \in \alpha_0^{\mathcal{I}}$ [2]. Let $\mathcal{J}$ be the interpretation obtained from $\mathcal{I}$ by restricting the domain $\Delta^{\mathcal{I}}$ to the following subset $\Delta^{\mathcal{J}}$:

$$\Delta^{\mathcal{J}} := \{d\} \cup \bigcup_{i=1}^{n} \mathcal{I}(d,i) \cap \alpha_i^{\mathcal{I}}.$$

Note that, since $\mathcal{I}$ is tree shaped, for all $e \in \Delta^{\mathcal{I}}$ there is at most one $1 \le i \le n$ such that $e \in \mathcal{I}(d,i)$. Hence, by definition of $\mathcal{J}$, the following holds for all $e \in \Delta^{\mathcal{J}}$ and $1 \le i \le n$:

$$e \in \mathcal{J}(d,i) \quad \Rightarrow \quad e \in \mathcal{I}(d,i) \text{ and } e \in \alpha_i^{\mathcal{I}}. \tag{34}$$

We now use (34) and $\mathcal{I}$ to show that $d \in C_G^{\mathcal{J}}$. We make a case distinction for the concepts C_1, C_2, C_3, C_4.

(C_1) Let $i \in \{1, 3, \dots, n-1\}$ and let $e \in \Delta^{\mathcal{J}}$ be such that $e \in \mathcal{J}(d,i)$. By (34), we have $e \in \mathcal{I}(d,i)$ and $e \in \alpha_i^{\mathcal{I}}$. Hence, since $\mathcal{I}$ is a model of $\mathcal{T}_{C_G}$ there are $e_1, e_2 \in \Delta^{\mathcal{I}}$ such that:

$$(e,e_j) \in r^{\mathcal{I}}, \qquad e_j \in \alpha_{i+1}^{\mathcal{I}}, \qquad e_1 \in P_{i+1}^{\mathcal{I}}, \qquad e_2 \in \bar{P}_{i+1}^{\mathcal{I}} \qquad (j = 1, 2).$$

From $(e,e_j) \in r^{\mathcal{I}}$, $e_j \in \alpha_{i+1}^{\mathcal{I}}$ and $e \in \mathcal{I}(d,i)$, we obtain that $e_1, e_2 \in \mathcal{I}(d, i+1) \cap \alpha_{i+1}^{\mathcal{I}}$. Therefore, $e_1, e_2 \in \Delta^{\mathcal{J}}$. In addition, $e_2 \in \alpha_{i+1}^{\mathcal{I}}$ implies that $e_2 \in N_{i+1}^{\mathcal{I}}$. The definition of N_{i+1} yields $e_2 \notin P_{i+1}^{\mathcal{I}}$ because $e_2 \in \bar{P}_{i+1}^{\mathcal{I}}$. Hence, since $\mathcal{J}$ is the restriction of $\mathcal{I}$ to $\Delta^{\mathcal{J}}$, we have $(e,e_1) \in r^{\mathcal{J}}$, $(e,e_2) \in r^{\mathcal{J}}$, $e_1 \in P_{i+1}^{\mathcal{J}}$ and $e_2 \notin P_{i+1}^{\mathcal{J}}$. Thus, we can conclude that $d \in C_1^{\mathcal{J}}$.

(C_2) Let $i \in \{0, 2, \dots, n-2\}$ and $e \in \Delta^{\mathcal{J}}$ be such that $e \in \mathcal{J}(d,i)$. Again, (34) yields $e \in \mathcal{I}(d,i)$ and $e \in \alpha_i^{\mathcal{I}}$. By definition of α_i, there is $f \in \Delta^{\mathcal{I}}$ such that $(e,f) \in r^{\mathcal{I}}$ and $f \in \alpha_{i+1}^{\mathcal{I}}$. This implies that $f \in \mathcal{I}(d, i+1) \cap \alpha_{i+1}^{\mathcal{I}}$. Hence, it follows that $f \in \Delta^{\mathcal{J}}$ and $e \in (\exists r.\top)^{\mathcal{J}}$. Thus, $d \in C_2^{\mathcal{J}}$.

(C_3) Let $1 \le j \le i < n$ and let $e \in \Delta^{\mathcal{J}}$ be such that $e \in \mathcal{J}(d,i)$. To see that $d \in C_3^{\mathcal{J}}$, we need to show that $e \in (P_j \to \forall r.P_j)^{\mathcal{J}}$ and $e \in (\neg P_j \to \forall r.\neg P_j)^{\mathcal{J}}$. Suppose that $e \in P_j^{\mathcal{J}}$ and let $f \in \Delta^{\mathcal{J}}$ be such that $(e,f) \in r^{\mathcal{J}}$. We have that $f \in \mathcal{J}(d, i+1)$ because $e \in \mathcal{J}(d,i)$. Hence, by (34) and the definition of $\mathcal{J}$ we obtain:

$$e \in \alpha_i^{\mathcal{I}}, \qquad f \in \alpha_{i+1}^{\mathcal{I}}, \qquad (e,f) \in r^{\mathcal{I}}, \qquad e \in P_j^{\mathcal{I}}.$$

From $e \in \alpha_i^{\mathcal{I}}$ it follows that $e \in G_i^{\mathcal{I}}$. Hence, the definition of G_i, $e \in P_j^{\mathcal{I}}$ and $(e,f) \in r^{\mathcal{I}}$ yield that $f \notin \bar{P}_j^{\mathcal{I}}$. Further, $f \in \alpha_{i+1}^{\mathcal{I}}$ implies that $f \in D_{i+1}^{\mathcal{I}}$. Hence, by definition of D_{i+1} and $f \notin \bar{P}_j^{\mathcal{I}}$, it follows that $f \in P_j^{\mathcal{I}}$. Finally, since $\mathcal{J}$ is the restriction of $\mathcal{I}$ to $\Delta^{\mathcal{J}}$ and $f \in \Delta^{\mathcal{J}}$, we obtain $f \in P_j^{\mathcal{J}}$. Therefore, the arbitrary selection of f implies that $e \in (P_j \to \forall r.P_j)^{\mathcal{J}}$. The case for $\neg P_j \to \forall r.\neg P_j$ can be shown using similar arguments. Thus, $d \in C_3^{\mathcal{J}}$.

(C_4) Let $e \in \Delta^{\mathcal{J}}$ be such that $e \in \mathcal{J}(d,n)$. Then (34) yields $e \in \alpha_n^{\mathcal{I}}$. Hence, $e \in [\xi(\phi^*)]^{\mathcal{I}}$ and $e \in (N_n \sqcap D_n)^{\mathcal{I}}$. The latter, together with the definition of $\mathcal{J}$, implies that $e \in \bar{P}_j^{\mathcal{I}}$ iff $e \notin P_j^{\mathcal{I}}$ iff $e \notin P_j^{\mathcal{J}}$ iff $e \in (\neg P_j)^{\mathcal{J}}$ for all $j, 1 \le j \le n$. Hence, $e \in [\xi(\phi^*)]^{\mathcal{I}}$ implies $e \in \phi^{*\mathcal{J}}$. Thus, we have shown that $d \in C_4^{\mathcal{J}}$.

Summing up, we have proved that $d \in C_G^{\mathcal{J}}$, which finishes the proof of ($\Leftarrow$). □

The second and final step, is to translate $\mathcal{T}_{C_G}$ into a $\tau\mathcal{EL}(m_\sim)$ TBox $\mathfrak{T}_{C_G} = (\widehat{\mathcal{T}}, \mathcal{T})$. Notice that, the conjuncts in N_i and G_i are either defined concepts or negated $\mathcal{EL}$ concepts of the form $\neg(P_i \sqcap \bar{P}_i)$, $\neg(P_i \sqcap \exists r.\bar{P}_i)$ or $\neg(\bar{P}_i \sqcap \exists r.P_i)$, which can be expressed in $\tau\mathcal{EL}(m_\sim)$ as (see Proposition 2.13):

$$(P_i \sqcap \bar{P}_i)_{<1} \qquad (P_i \sqcap \exists r.\bar{P}_i)_{<1} \qquad (\bar{P}_i \sqcap \exists r.P_i)_{<1}.$$

To deal with the concepts of the form $P_i \sqcup \bar{P}_i$ in D_i and the disjunctions in $\xi(\phi^*)$, we use the following $\tau\mathcal{EL}(m)$ concept descriptions:

$$(P_i \sqcap \bar{P}_i)_{>0} \quad \text{and} \quad \hat{\xi}(\phi^*) := \prod_{i=1}^{m}(X_{i_1} \sqcap \cdots \sqcap X_{i_{n_i}})_{>0},$$

where $X_{i_1}, \dots, X_{i_{n_i}}$ are the disjuncts in $\xi(\phi_i^*)$. Hence, we define the TBox $\hat{\mathcal{T}}$ as the result of replacing the non-$\mathcal{EL}$ concepts occurring in $\mathcal{T}_{C_G}$ with their corresponding threshold concepts. It is clear that the pair $\mathfrak{T}_{C_G} = (\hat{\mathcal{T}}, \emptyset)$ is an acyclic $\tau\mathcal{EL}(m)$ TBox. However, for the simulation of disjunction to work correctly, we need to restrict the concept measure.

Lemma 4.24. *Let $\sim \in$ simi-d be such that $pm = pm_{df}$. Then, α_0 is satisfiable w.r.t. $\mathcal{T}_{C_G}$ iff α_0 is satisfiable w.r.t. $\mathfrak{T}_{C_G}$ in $\tau\mathcal{EL}(m_\sim)$.*

Proof. Let $\mathcal{I}$ be any interpretation. Since $pm = pm_{df}$, it is the case that $e \in (P_j \sqcup \bar{P}_j)^\mathcal{I}$ iff $e \in [(P_j \sqcap \bar{P}_j)_{>0}]^\mathcal{I}$. The same applies for the conjuncts in $\xi(\phi^*)$ and their translation in $\hat{\xi}(\phi^*)$. In addition, we know that, in $\tau\mathcal{EL}(m_\sim)$, the equivalence $\neg C \equiv C_{<1}$ holds for all $C \in \mathcal{C}_{\mathcal{EL}}$. This shows that $\mathcal{I} \models \mathcal{T}_{C_G}$ iff $\mathcal{I} \models \mathfrak{T}_{C_G}$ holds for all interpretations $\mathcal{I}$, which completes the proof of the lemma. □

It is not hard to see that $\mathcal{T}_{C_G}$ is of size polynomial in the size of C_G, and so is $\mathfrak{T}_{C_G}$. Therefore, satisfiability of concepts of the form C_G is polynomial-time reducible to concept satisfiability w.r.t. acyclic $\tau\mathcal{EL}(m_\sim)$ TBoxes. Regarding the empty TBox, it is not hard to see from the reduction above, that satisfiability of ϕ can be reduced to satisfiability of the $\tau\mathcal{EL}(m_\sim)$ concept obtained from $\xi(\phi^*) \sqcap N_n \sqcap D_n$. This yields NP-hardness of concept satisfiability in $\tau\mathcal{EL}(m_\sim)$. Moreover, we know that satisfiability can be reduced to the consistency and non-subsumption problem. Finally, since PSpace is closed under complement, using the upper bounds from the previous section we obtain the following results.

Theorem 4.25. *Let $\sim \in$ simi-d such that $pm = pm_{df}$. In $\tau\mathcal{EL}(m_\sim)$, satisfiability and consistency are NP-complete, whereas subsumption is coNP-complete. These problems become PSpace-complete w.r.t. acyclic $\tau\mathcal{EL}(m_\sim)$ TBoxes.*

From these results one can also infer that instance checking is PSpace-hard in combined complexity w.r.t. acyclic $\tau\mathcal{EL}(m)$ TBoxes. We next show that this is even the case w.r.t. $\mathfrak{T} = \emptyset$ and any $\sim \in$ *simi-d*.

4.3.2. PSpace hardness of instance checking w.r.t. the empty TBox

We reuse the reduction provided in [33] to show that instance checking in the DL $\mathcal{ALE}$ is PSpace-hard w.r.t. combined complexity. This reduction is from satisfiability of $\mathcal{ALC}$ concepts. Basically, given an $\mathcal{ALC}$ concept C, the reduction relies on being able to express C as a conjunction $D \sqcap \neg D_1 \sqcap \cdots \sqcap \neg D_k$, where D and each D_j are $\mathcal{ALE}$ concepts. Then, satisfiability of C is reduced to non-instance of the individual a in the concept $E_0(a)$ w.r.t. the following ABox:

$$\mathcal{A} := \{s(a,a), D_1(a), \dots, D_k(a), s(a,b), E_1(b), \dots, E_k(b), D(b)\},$$

where $a, b \in \mathsf{NI}$, $s \in \mathsf{NR}$, $E_k \in \mathsf{NC}$, s and E_k do not occur in C, and $E_i := \exists s.(D_{i+1} \sqcap E_{i+1})$ for all $i, 0 \le i < k$.

To adapt this idea to our case, we show how to express concepts of the form $C_G = C_1 \sqcap C_2 \sqcap C_3 \sqcap C_4$, as introduced above, as a conjunction $\hat{D} \sqcap \neg\hat{D}_1 \sqcap \cdots \sqcap \neg\hat{D}_k$ where $\hat{D}, \hat{D}_1, \dots, \hat{D}_k$ are $\tau\mathcal{EL}(m)$ concept descriptions. First, using the equivalences $\forall r.C \equiv \neg\exists r.\neg C$, $\forall r.(C \sqcap D) \equiv \forall r.C \sqcap \forall r.D$, $C \to D \equiv \neg C \sqcup D$, and *De Morgan's* laws, we obtain the following equivalences:

$$C_1 \equiv \prod_{i \in \{1,3,\dots,n-1\}} \neg\exists r^i.\neg(\exists r.\neg P_{i+1} \sqcap \exists r.P_{i+1}), \qquad C_2 \equiv \exists r.\top \sqcap \prod_{i \in \{2,\dots,n-2\}} \neg\exists r^i.\neg\exists r.\top,$$

$$C_3 \equiv \prod_{1 \le i \le j < n} \neg\exists r^i.(P_j \sqcap \exists r.\neg P_j) \sqcap \neg\exists r^i.(\neg P_j \sqcap \exists r.P_j), \qquad C_4 \equiv \prod_{j=1}^{m} \neg\exists r^n.\neg\phi_j^*.$$

From this, it is easy to see that C_2 and C_3 are equivalent to the concepts

$$C_2' := \exists r.\top \sqcap \prod_{i \in \{2,\dots,n-2\}} \neg\exists r^i.(\exists r.\top)_{<1}, \quad C_3' := \prod_{1 \le i \le j < n} \neg\exists r^i.(P_j \sqcap \exists r.(P_j)_{<1}) \sqcap \neg\exists r^i.((P_j)_{<1} \sqcap \exists r.P_j),$$

and C_4 is equivalent to the concept:

$$C_4' := \prod_{j=1}^{m} \neg\exists r^n.(\hat{\neg}X_1 \sqcap \cdots \sqcap \hat{\neg}X_{\ell_j}),$$

where $\hat{\neg}X_p = P_i$ if $X_p = \neg P_i$ and $\hat{\neg}X_p = (P_i)_{<1}$ if $X_p = P_i$.

As for C_1, due to the nested negation inside $\neg\exists r^i$, it is not immediate how to simulate/express it as a conjunction of negated $\tau\mathcal{EL}(m)$ concepts. To deal with this issue, we again introduce concept names $\bar{P}_i$ representing $\neg P_i$. The intention is then to simulate C_1 with the following concept:

$$C_1' := \bigsqcap_{i\in\{1,3,\ldots,n-1\}} \neg\exists r^i.(\exists r.\bar{P}_{i+1} \sqcap \exists r.P_{i+1})_{<1}.$$

However, for this to be correct, the complementary semantics of P_i and $\neg P_i$ needs to be enforced also for the pair P_i and $\bar{P}_i$. This can be expressed by the $\mathcal{ALC}$ concept $C_5 := \bigsqcap_{i=2,4,\ldots,n} \forall r^i.(\neg(P_i \sqcap \bar{P}_i) \sqcap (P_i \sqcup \bar{P}_i))$, which is equivalent to:

$$C_5' := \bigsqcap_{i=2,4,\ldots,n} \neg\exists r^i.(P_i \sqcap \bar{P}_i) \sqcap \neg\exists r^i.((P_i)_{<1} \sqcap (\bar{P}_i)_{<1}).$$

Given this, it is now easy to prove the following proposition.

Proposition 4.26. *Let $\mathcal{I}$ be an interpretation. If $d \in (C_1' \sqcap \cdots \sqcap C_5')^{\mathcal{I}}$ then $d \in (C_1 \sqcap \cdots \sqcap C_4)^{\mathcal{I}}$. Conversely, if $\bar{P}_i^{\mathcal{I}} = \neg P_i^{\mathcal{I}}$ for all $i, 1 \le i \le n$, then $d \in (C_1 \sqcap \cdots \sqcap C_4)^{\mathcal{I}}$ implies $d \in (C_1' \sqcap \cdots \sqcap C_5')^{\mathcal{I}}$.*

Let $\widehat{D} = \exists r.\top$ and $\neg\widehat{D}_1, \ldots, \neg\widehat{D}_k$ be an enumeration of the remaining conjuncts in $C_1' \sqcap \cdots \sqcap C_5'$. Let $\widehat{E}_k \in \mathsf{NC}$. We define the concepts $\widehat{E}_0, \ldots, \widehat{E}_{k-1}$ like $E_0, \ldots, E_{k-1}$, but using $\widehat{D}_1, \ldots, \widehat{D}_k, \widehat{E}_k$ instead of $D_1, \ldots, D_k, E_k$. The definition of the ABox $\mathcal{A}$ is adapted accordingly, i.e., we define:

$$\widehat{\mathcal{A}} := \{s(a,a), \widehat{D}_1(a), \ldots, \widehat{D}_k(a), s(a,b), \widehat{E}_1(b), \ldots, \widehat{E}_k(b), \widehat{D}(b)\}.$$

Clearly, $\widehat{E}_0$ and $\widehat{\mathcal{A}}$ are a concept and an ABox in $\tau\mathcal{EL}(m)$. Hence, the following lemma shows that satisfiability of $\mathcal{ALC}$ concepts can be reduced to non-instance in $\tau\mathcal{EL}(m)$.

Lemma 4.27. *Let $\sim \in$ simi-d. The $\mathcal{ALC}$ concept C_G is satisfiable iff $\mathcal{A} \not\models \widehat{E}_0(a)$ in $\tau\mathcal{EL}(m_\sim)$.*

Proof. ($\Rightarrow$) Assume that C_G is satisfiable. Since the concept names $\bar{P}_i$ $(1 \le i \le n)$ do not occur in C_G, there is an interpretation $\mathcal{I}_b$ such that $\bar{P}_i^{\mathcal{I}_b} = \neg P_i^{\mathcal{I}_b}$ for all $i, 1 \le i \le n$, and $d_b \in C_G^{\mathcal{I}_b}$ for some $d_b \in \Delta^{\mathcal{I}_b}$. Hence, Proposition 4.26 yields $d_b \in (C_1' \sqcap \cdots \sqcap C_5')^{\mathcal{I}_b}$. In addition, it is easy to see that each $\widehat{D}_j$ $(1 \le j \le k)$ is satisfiable,[14] and since $\widehat{D}_j = \exists r^i \ldots$ with $i > 0$, there is an interpretation $\mathcal{I}_a$ and $d_a \in \Delta^{\mathcal{I}_a}$ such that $d_a \in \widehat{D}_j^{\mathcal{I}_a}$ for all $j, 1 \le j \le k$. Since $\widehat{E}_k$ does not occur in C_G, we can assume that $d_b \in \widehat{E}_k^{\mathcal{I}_b}$ and $d_a \notin \widehat{E}_k^{\mathcal{I}_a}$.

We define an interpretation $\mathcal{J}$ with domain $\Delta^{\mathcal{J}} := \Delta^{\mathcal{I}_a} \cup \Delta^{\mathcal{I}_b}$, where $\Delta^{\mathcal{I}_a}$ and $\Delta^{\mathcal{I}_b}$ are assumed to be mutually disjoint sets. Concept names and role names distinct from s are interpreted in $\mathcal{J}$ as the union of their interpretations in $\mathcal{I}_a$ and $\mathcal{I}_b$, whereas s is interpreted as $s^{\mathcal{J}} := \{(d_a, d_a), (d_a, d_b)\}$. Finally, we define $a^{\mathcal{J}} = d_a$ and $b^{\mathcal{J}} = d_b$. Note that the following holds for all $e, f \in \Delta^{\mathcal{J}}$: $(e,f) \in r^{\mathcal{J}}$ implies that $e, f \in \Delta^{\mathcal{I}_a}$ or $e, f \in \Delta^{\mathcal{I}_b}$. Hence, since s does not occur in $C_1' \sqcap \ldots \sqcap C_5'$ and $\widehat{D}_j$ $(1 \le j \le k)$, we obtain

$$b^{\mathcal{J}} \in (C_1' \sqcap \ldots \sqcap C_5')^{\mathcal{J}} \qquad \text{and} \qquad a^{\mathcal{J}} \in \widehat{D}_j^{\mathcal{J}} \quad (1 \le j \le k). \tag{35}$$

It is clear that $\mathcal{J}$ satisfies all assertions in $\widehat{\mathcal{A}} \setminus \{\widehat{E}_1(b), \ldots, \widehat{E}_{k-1}(b)\}$. To satisfy the remaining assertions as well, we need to extend $\mathcal{J}$ further. Recall that $\widehat{D}_j$ $(1 \le j \le k)$ are satisfiable. Hence, since $\widehat{E}_k$ and s do not occur in $\widehat{D}_j$ and either $\widehat{E}_j = \widehat{E}_k$ or $\widehat{E}_j = \exists s.(\ldots)$, the following holds:

$$\text{If } \widehat{E}_j \text{ is satisfiable then } \widehat{E}_j \sqcap \widehat{D}_j \text{ is also satisfiable.} \tag{36}$$

The right-hand side of (36) implies that $\widehat{E}_{j-1}$ is also satisfiable since $\widehat{E}_{j-1} = \exists s.(\widehat{E}_j \sqcap \widehat{D}_j)$. Therefore, starting with $\widehat{E}_k \in \mathsf{NC}$, we can apply this observation to obtain that $\widehat{E}_j \sqcap \widehat{D}_j$ $(1 \le j \le k)$ are satisfiable concepts. Hence, there are interpretations $\mathcal{J}_\ell$ and $e_\ell \in \Delta^{\mathcal{J}_\ell}$ $(1 \le \ell < k)$ such that $e_\ell \in (\widehat{E}_{\ell+1} \sqcap \widehat{D}_{\ell+1})^{\mathcal{J}_\ell}$. We extend $\mathcal{J}$ by making each e_ℓ an s-successor of $b^{\mathcal{J}}$, i.e., $(b^{\mathcal{J}}, e_\ell) \in s^{\mathcal{J}}$. Obviously, we still have $e_\ell \in (\widehat{E}_{\ell+1} \sqcap \widehat{D}_{\ell+1})^{\mathcal{J}}$. Hence, $\mathcal{J}$ now satisfies $\widehat{E}_1(b), \ldots, \widehat{E}_{k-1}(b)$. Notice that (35) is preserved in this extension. In fact, the argument given above to justify (35) remains true, if we also consider $\mathcal{J}_1, \ldots, \mathcal{J}_{k-1}$. Therefore, $\mathcal{J} \models \widehat{\mathcal{A}}$.

It remains to show that $a^{\mathcal{J}} \notin \widehat{E}_0^{\mathcal{J}}$. Since $a^{\mathcal{J}} \notin \widehat{E}_k^{\mathcal{J}}$ by assumption, it suffices to show:

$$a^{\mathcal{J}} \notin \widehat{E}_j^{\mathcal{J}} \text{ implies } a^{\mathcal{J}} \notin \widehat{E}_{j-1}^{\mathcal{J}} \text{ for all } j, 1 \le j \le k.$$

Assume $a^{\mathcal{J}} \notin \widehat{E}_j^{\mathcal{J}}$ holds. The implication holds if no s-successor of $a^{\mathcal{J}}$ satisfies $\widehat{E}_j \sqcap \widehat{D}_j$. By definition of $\mathcal{J}$, the only s-successors of $a^{\mathcal{J}}$ are $a^{\mathcal{J}}$ and $b^{\mathcal{J}}$. We have $a^{\mathcal{J}} \notin \widehat{E}_j^{\mathcal{J}}$ by assumption. Regarding $b^{\mathcal{J}}$, we know that $b^{\mathcal{J}} \in (\neg\widehat{D}_j)^{\mathcal{J}}$ since $b^{\mathcal{J}} \in (C_1' \sqcap \cdots \sqcap C_5')^{\mathcal{J}}$ by (35). Hence, since $\widehat{E}_{j-1} = \exists s.(\widehat{E}_j \sqcap \widehat{D}_j)$, we have shown that $a^{\mathcal{J}} \notin \widehat{E}_{j-1}^{\mathcal{J}}$. Thus, it follows that $a^{\mathcal{J}} \notin \widehat{E}_0^{\mathcal{J}}$.

($\Leftarrow$) Suppose $\widehat{\mathcal{A}} \not\models \widehat{E}_0(a)$. Then, there is an interpretation $\mathcal{I}$ such that $\mathcal{I} \models \widehat{\mathcal{A}}$ and $a^{\mathcal{I}} \notin \widehat{E}_0^{\mathcal{I}}$. The definition of $\widehat{\mathcal{A}}$ yields

$$(a^{\mathcal{I}}, a^{\mathcal{I}}) \in s^{\mathcal{I}},\ \ a^{\mathcal{I}} \in \widehat{D}_1^{\mathcal{I}},\ \ldots,\ a^{\mathcal{I}} \in \widehat{D}_k^{\mathcal{I}} \quad \text{and} \quad (a^{\mathcal{I}}, b^{\mathcal{I}}) \in s^{\mathcal{I}},\ \ b^{\mathcal{I}} \in \widehat{E}_1^{\mathcal{I}},\ \ldots,\ b^{\mathcal{I}} \in \widehat{E}_k^{\mathcal{I}}.$$

[14] The only case where $\widehat{D}_j$ may not be satisfiable is when it is of the form $\exists r^n.(\ldots P_i \sqcap (P_i)_{<1} \ldots)$. However, this would mean that a clause in ϕ contains literals x_i and $\neg x_i$, which is a trivial case and can thus be dismissed, i.e., we can restrict the attention to formulas ϕ for which this is not the case.

Hence, since $\widehat{E}_0 = \exists s.(\widehat{E}_1 \sqcap \widehat{D}_1)$, $a^{\mathcal{I}} \notin \widehat{E}_0^{\mathcal{I}}$ implies that $a^{\mathcal{I}} \notin \widehat{E}_1^{\mathcal{I}}$ and $b^{\mathcal{I}} \notin \widehat{D}_1^{\mathcal{I}}$. By iterating this argument over all $j, 1 \leq j \leq k-1$, we obtain that $b^{\mathcal{I}} \in \neg\widehat{D}_\ell^{\mathcal{I}}$ for all $\ell, 1 \leq \ell \leq k$. Further, we also have $b^{\mathcal{I}} \in \widehat{D}^{\mathcal{I}}$ because $\widehat{D}(b) \in \widehat{\mathcal{A}}$. Hence, $b^{\mathcal{I}} \in (C_1' \sqcap \ldots \sqcap C_5')^{\mathcal{I}}$, and Proposition 4.26 yields $b^{\mathcal{I}} \in C_G^{\mathcal{I}}$. Thus, we have shown that $\widehat{\mathcal{A}} \not\models \widehat{E}_0(a)$ implies that C_G is satisfiable. □

The previous lemma and Theorem 4.21 yield the following complexity result.

Theorem 4.28. *Let $\sim \in$ simi-d. In $\tau\mathcal{EL}(m_\sim)$, instance checking is PSpace-complete w.r.t. acyclic $\tau\mathcal{EL}(m_\sim)$ TBoxes. PSpace-hardness holds even for the empty TBox and for arbitrary graded membership functions.*

PSpace-hardness of instance checking in $\tau\mathcal{EL}(m)$ for arbitrary graded membership functions m follows from the fact that the above proof only used the property that $C_{<1}$ is equivalent to $\neg C$, which holds without further restrictions on m (see Proposition 2.13).

4.4. Data complexity of instance checking w.r.t. acyclic $\tau\mathcal{EL}(m_\sim)$ TBoxes

Recall that, in the presence of non-empty TBoxes, we consider the data complexity of answering instance queries for $\tau\mathcal{EL}(m_\sim)$ KBs $(\widehat{\mathcal{T}}, \mathcal{A})$ where $\mathcal{A}$ is a *simple* ABox. This is the usual assumption when investigating the data complexity of query answering over DL knowledge bases [31]. Note that this assumption does not restrict the expressiveness of $\tau\mathcal{EL}(m_\sim)$ KBs since assertions involving complex concepts $\widehat{C}$ can be expressed by introducing a definition of $\widehat{C}$ in the TBox. However, when considering data complexity, the size of the resulting TBox, and thus of $\widehat{C}$, is then assumed to be fixed.

For the case where the TBox is empty we do not restrict the ABox to be simple, but consider non-simple ABoxes. The reason is that restricting to simple ABoxes would restrict the expressiveness of KBs considerably since individuals could no longer be required to belong to complex concepts. The notion of data complexity used for non-simple ABoxes counts the sizes of complex concepts in concept assertions as part of the input size, i.e., the sizes of these concepts are not assumed to be fixed. This is in line with how data complexity is defined in [32,33] for $\mathcal{ALE}$ knowledge bases. The rest of this section analyzes the data complexity of the instance checking problem for these two settings.

4.4.1. Empty TBox

The non-deterministic procedure described in Subsection 4.2.4 to decide non-instance w.r.t. the empty TBox, guesses an interpretation $\mathcal{I}$ of size at most $\mathsf{s}(\mathcal{A}) \cdot \mathsf{s}(\widehat{D})^u$, where $\widehat{D}$ is the query concept and u is the cardinality of $\mathsf{sub}(\widehat{D})$. Since $\widehat{D}$ is assumed to be fixed, the interpretation $\mathcal{I}$ is of size polynomial in $\mathsf{s}(\mathcal{A})$. Therefore, the procedure runs in non-deterministic polynomial time in the size of $\mathcal{A}$. This shows that, w.r.t. the empty TBox, instance checking is in coNP w.r.t. data complexity.

A matching lower bound can be obtained from a result shown in [32]. More precisely, Schaerf proves in [32] that instance checking in the DL $\mathcal{EL}^{\neg A}$ is coNP-hard w.r.t. data complexity. The DL $\mathcal{EL}^{\neg A}$ is the extension of $\mathcal{EL}$ with *atomic negation*, i.e., the use of negated concept names is allowed in the logic. Recall that $\neg A$ is equivalent to $A_{<1}$ in $\tau\mathcal{EL}(m)$ regardless of the considered graded membership function m. Hence, coNP-hardness of instance checking in $\mathcal{EL}^{\neg A}$ implies coNP-hardness of instance checking in $\tau\mathcal{EL}(m_\sim)$ w.r.t. data complexity.

Theorem 4.29. *Let $\sim \in$ simi-d. In $\tau\mathcal{EL}(m_\sim)$, instance checking is coNP-complete w.r.t. the empty TBox for data complexity.*

4.4.2. Arbitrary acyclic $\tau\mathcal{EL}(m_\sim)$ TBoxes

In Algorithm 3, the only non-determinism involving the size of $\mathcal{A}$ consists of guessing the sets v_b at Line 1. In addition, all the other computations performed during an execution of the algorithm are either polynomial in $\mathsf{s}(\mathcal{A})$ or do not depend on $\mathcal{A}$. Hence, Algorithm 3 runs in non-deterministic polynomial time in $\mathsf{s}(\mathcal{A})$. This yields an NP upper bound for the non-instance problem, and therefore instance checking is in coNP regarding data complexity.

As for the case of the empty TBox, the matching lower bound follows from existing results for $\mathcal{EL}^{\neg A}$. In [38], the reduction given in [32] is reused to show that the data complexity of instance checking in $\mathcal{EL}^{\neg A}$ stays coNP-hard if one restricts ABoxes to being simple, but allows for acyclic $\mathcal{EL}^{\neg A}$ TBoxes. Due to the equivalence of $\neg A$ and $A_{<1}$ in $\tau\mathcal{EL}(m_\sim)$, acyclic $\mathcal{EL}^{\neg A}$ TBoxes can be expressed by acyclic $\tau\mathcal{EL}(m_\sim)$ TBoxes. This yields the lower bound in the following complexity result.

Theorem 4.30. *Let $\sim \in$ simi-d. In $\tau\mathcal{EL}(m_\sim)$, instance checking is coNP-complete w.r.t. acyclic $\tau\mathcal{EL}(m_\sim)$ TBoxes for data complexity.*

5. Reasoning with general $\tau\mathcal{EL}(m_\sim)$ TBoxes

Having investigated the complexity of reasoning in threshold logics without TBox and w.r.t. acyclic TBoxes, we now concentrate on general TBoxes. We show that, for all measures $\sim \in$ *simi-d*, all the considered reasoning problems are ExpTime-complete w.r.t. general $\tau\mathcal{EL}(m_\sim)$ TBoxes. In the particular case of instance checking, the problem remains coNP-complete w.r.t. data complexity.

The *ExpTime lower bounds* are an easy consequence of ExpTime-hardness of concept satisfiability w.r.t. general TBoxes in $\mathcal{EL}^{\neg A}$ [7]. The reduction simply replaces each occurrence of $\neg A$ in an $\mathcal{EL}^{\neg A}$ instance of the satisfiability problem by the threshold concept $A_{<1}$. As for the data complexity of instance checking, coNP-hardness follows from Theorem 4.30.

The remainder of this section is devoted to proving the *ExpTime upper bounds*. To this purpose, we devise an algorithm that can decide consistency of KBs of the form $(\mathfrak{T}, \mathcal{A} \cup \{\neg\widehat{D}(a)\})$, where $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$, $a \in \mathsf{NI}$ and $\widehat{D}$ is correctly defined w.r.t $\mathcal{T}$. Like in the case of acyclic TBoxes, this algorithm can also be used to decide the consistency and the non-instance problem. Since a concept $\widehat{C} \sqcap \neg\widehat{D}$ is satisfiable w.r.t. a $\tau\mathcal{EL}(m)$ TBox $\mathfrak{T}$ iff $(\mathfrak{T}, \{\widehat{C}(a)\} \cup \{\neg\widehat{D}(a)\})$ is consistent, we also obtain a decision procedure for the satisfiability and the non-subsumption problem. In the rest of the section, we make the following assumptions regarding the form of $(\mathfrak{T}, \mathcal{A} \cup \{\neg\widehat{D}(a)\})$:

- Similarly to the case of acyclic $\tau\mathcal{EL}(m)$ knowledge bases, we assume that all concept assertions in $\mathcal{A}$ are of the form $\beta(b)$ with $\beta \in \mathsf{NC}_d^{\mathfrak{T}}$, $\widehat{D} = \alpha$ with $\alpha \in \mathsf{NC}_d^{\mathfrak{T}}$, and $a \in \mathsf{Ind}(\mathcal{A})$. Moreover, all threshold concepts occurring in $\widehat{\mathcal{T}}$ are of the form $E_{\bowtie t}$ with $E \in \mathsf{NC}_d^{\mathcal{T}}$, and $\mathcal{T}$ is in normal and reduced form.
- Every role name r occurring in $\mathcal{A}$ also occurs in $\mathfrak{T}$. This is without loss of generality since we can simply add to $\widehat{\mathcal{T}}$ the tautology $\exists r.\top \sqsubseteq \top$.

Our decision procedure consists of a *type elimination algorithm* that uses an extended version of the notion of type introduced in Subsection 4.2.3. Let Γ^+, Γ^- and Γ be as defined in Subsection 4.2.3, and $\Gamma_\tau = \{D_1, \dots, D_m\}$ a set of $\mathcal{EL}$ concepts such that $C_{\bowtie t} \in \Gamma$ implies that $C \in \Gamma_\tau$. In addition, let $\mathcal{V}$ be a set of rational numbers in the interval $[0,1]$. A type for $(\Gamma, \Gamma_\tau, \mathcal{V})$ is a tuple $\mu = (\nu, v_1, \dots, v_m)$ such that ν is a type for Γ, $v_i \in \mathcal{V}$ $(1 \le i \le m)$, and the following condition is satisfied:

t4) for all $C_{\bowtie t} \in \Gamma$: $C_{\bowtie t} \in \nu$ iff $v_i \bowtie t$, where $1 \le i \le m$ and $C = D_i$.

Intuitively, a type μ now additionally contains with $v_1, \dots, v_m$ the degrees of membership in each concept from Γ_τ of the individual d_μ it describes. This is needed to deal with the semantics of threshold concepts since, in contrast to the case of acyclic TBoxes, finite tree models may not exist in the presence of a general $\tau\mathcal{EL}(m)$ TBox, and thus we cannot use the procedure DEGREES to compute the membership degrees while guessing a finite tree model. Condition t4) ensures that these degrees are consistent with the threshold concepts $C_{\bowtie t} \in \Gamma$ for which ν states that d_μ belongs to $C_{\bowtie t}$.

Similarly to the case of acyclic $\tau\mathcal{EL}(m)$ TBoxes, we need to define what is relevant in the presence of $\mathfrak{T}$, which now also involves the sets Γ_τ and $\mathcal{V}$:

- Like before, we define $\Gamma := \mathsf{cl}(\mathfrak{T}, \sim)$, and $\mathsf{cl}(\mathfrak{T}, \sim)$ in terms of $\mathsf{sub}(\mathfrak{T})$ and $\mathfrak{C}(\mathfrak{T}, \sim)$. However, the definitions of $\mathsf{sub}(\mathfrak{T})$ and $\mathfrak{C}(\mathfrak{T}, \sim)$ are slightly modified. More precisely, in the definition of $\mathsf{sub}(\mathfrak{T})$, the sub-descriptions in $\widehat{\mathcal{T}}$ now consist of $\bigcup_{\widehat{C} \sqsubseteq \widehat{D} \in \widehat{\mathcal{T}}} \mathsf{sub}(\widehat{C}) \cup \mathsf{sub}(\widehat{D})$, whereas the ones in $\mathcal{T}$ remain the same. Finally, $\mathfrak{C}(\mathfrak{T}, \sim)$ is defined as the set

$$\{B \in \mathsf{NC} \mid B \in \mathfrak{s}_{pm}(A) \wedge A \in \mathsf{sig}(\mathfrak{T}) \cap \mathsf{NC}\} \cup \{D \in \mathsf{ND} \mid D \in \mathsf{sig}(\mathfrak{T})\}.$$

- The set Γ_τ is defined as $\mathsf{NC}_d^{\mathcal{T}} = \{E_1, \dots, E_m\}$. Given $\Gamma = \mathsf{cl}(\mathfrak{T}, \sim)$, this ensures that $C_{\bowtie t} \in \Gamma$ implies $C \in \Gamma_\tau$, since threshold concepts in $\widehat{\mathcal{T}}$ are assumed to be of the form $E_{\bowtie t}$ with $E \in \mathsf{NC}_d^{\mathcal{T}}$.
- The set $\mathcal{V}$ should contain all possible values for $\widehat{m}_\sim^{\mathcal{I}}(d, E, \mathcal{T})$ where $E \in \mathsf{NC}_d^{\mathcal{T}}$. Recall that $\widehat{m}_\sim^{\mathcal{I}}(d, E, \mathcal{T}) = m_\sim^{\mathcal{I}}(d, u_{\mathcal{T}}(E)) = u_{\mathcal{T}}(E) \sim D$ for some $\mathcal{EL}$ concept D. In Lemma 3.26, we have shown that $u_{\mathcal{T}}(E) \sim D$ is a rational number $y/x_{u_{\mathcal{T}}(E)}$, and determined the form of $x_{u_{\mathcal{T}}(E)}$. For simplicity, we denote $x_{u_{\mathcal{T}}(E)}$ from now on as x_E. As set $\mathcal{V}$ of relevant rational numbers we use the following set $\mathcal{V}_{\mathcal{T},\sim}$:

$$\mathcal{V}_{\mathcal{T},\sim} := \bigcup_{E \in \mathsf{NC}_d^{\mathcal{T}}} \{y/x_E \mid 0 \le y \le x_E \wedge y \in \mathbb{N}\}.$$

We then say that $\mu = (\nu, v_1, \dots, v_m)$ is a *type for* $\mathfrak{T}$ if it is a type for $(\mathsf{cl}(\mathfrak{T}, \sim), \mathsf{NC}_d^{\mathcal{T}}, \mathcal{V}_{\mathcal{T},\sim})$, and ν satisfies the GCIs and definitions in $\mathfrak{T}$, i.e.:

- for all $\widehat{C} \sqsubseteq \widehat{D} \in \widehat{\mathcal{T}}$: $\widehat{C} \in \nu$ implies $\widehat{D} \in \nu$, and
- for all $E \doteq C_E \in \mathcal{T}$: $E \in \nu$ iff $C_E \in \nu$.

Given tuples $\mu, \mu_x, \mu^x \in 2^\Gamma \times \mathcal{V}^{|\Gamma_\tau|}$, where x is a subscript or superscript that shall be determined by the context, we denote their first components as ν, ν_x and ν^x, respectively. Then, the notion of *r-successor candidate* is extended to tuples μ in $2^\Gamma \times \mathcal{V}^{|\Gamma_\tau|}$ by projection onto the first component. More precisely, given $r \in \mathsf{rol}(\nu)$ and a sequence $\psi = \mu^1, \dots, \mu^\ell$ of tuples in $2^\Gamma \times \mathcal{V}^{|\Gamma_\tau|}$ $(\ell \ge 0)$, we say that ψ is an r-successor candidate of μ w.r.t. Γ if $\nu^1, \dots, \nu^\ell$ is an r-successor candidate of ν w.r.t. Γ.

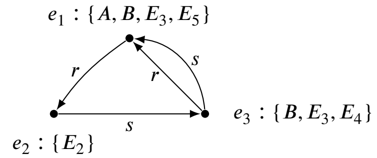
The following example illustrates this new version of the notion of type.

Example 5.1. Consider the logic $\tau\mathcal{EL}(m_{\sim_d^*})$, where $\sim_d^*$ is the CM introduced in Example 3.16. In addition, let $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$ and $\mathcal{I}$ be the $\tau\mathcal{EL}(m)$ TBox and the DL interpretation depicted below.

$$\hat{\mathcal{T}} := \{(E_3)_{\geq 0} \sqsubseteq (E_1)_{\geq 1/2},\ (E_2)_{<1} \sqsubseteq B\}$$

$$\mathcal{T} := \{E_1 \doteq \exists r.E_2 \sqcap \exists s.E_3,\ E_2 \doteq \exists s.E_4,\ E_3 \doteq B,\ E_4 \doteq \exists r.E_5,\ E_5 \doteq A\}$$

$\mathcal{I}$:



It is easy to verify that, in $\tau\mathcal{EL}(m_{\sim_d^*})$, the interpretation $\mathcal{I}$ is a model of $\mathfrak{T}$. In addition, in the presence of $\sim_d^*$ and $\mathfrak{T}$, the components of $(\Gamma, \Gamma_\tau, \mathcal{V})$ correspond to $\Gamma := \mathsf{cl}(\mathfrak{T}, \sim_d^*)$, $\Gamma_\tau := \{E_1, E_2, E_3, E_4, E_5\}$ and $\mathcal{V} := \mathcal{V}_{\mathcal{T},\sim_d^*}$. Hence, a type for $\mathfrak{T}$ is a tuple of the form $\mu = (\nu, v_1, v_2, v_3, v_4, v_5)$ such that:

- ν is a type for $\mathsf{cl}(\mathfrak{T}, \sim_d^*)$ satisfying the GCIs and definitions in $\mathfrak{T}$, $v_i \in \mathcal{V}_{\mathcal{T},\sim_d^*}$ $(i \in \{1, \dots, 5\})$, and the components of μ satisfy t4).

As with the simpler notion of type, the set ν in μ contains the concepts in $\mathsf{cl}(\mathfrak{T}, \sim_d^*)$ satisfied by the individual that μ describes. The difference now is that μ additionally contains the degree of membership in each $E_i \in \Gamma_\tau$ of such individual, i.e., the values of each v_i. Note that the satisfaction of condition t4) ensures that v_i is consistent with whether ν contains or not each threshold concept $(E_i)_{\bowtie t} \in \mathsf{cl}(\mathfrak{T}, \sim_d^*)$. Based on this, the domain elements e_1, e_2, e_3 of $\mathcal{I}$ are described (in order) by the following types:

$$\mu_1 = (\nu_1, 1/2, 0, 1, 0, 1), \qquad \mu_2 = (\nu_2, 1/2, 1, 0, 0, 0), \qquad \mu_3 = (\nu_3, 1/2, 0, 1, 1, 0),$$

where the component v_i $(i \in \{1 \dots, 5\})$ in each μ_j $(1 \leq j \leq 3)$ corresponds to $\hat{m}^{\mathcal{I}}_{\sim_d^*}(e_j, E_i, \mathcal{T})$, whereas each ν_j consists of the set of concepts in $\mathsf{cl}(\mathfrak{T}, \sim_d^*)$ that e_j is an instance of. For example, since $(E_3)_{\geq 0} \equiv \top$ and $\mathcal{I} \models \mathfrak{T}$, it is the case that $(E_3)_{\geq 0} \in \nu_j$ and $(E_1)_{\geq 1/2} \in \nu_j$ for all $j = 1, 2, 3$.

The relational structure of $\mathcal{I}$ is realized by using the notion of role-successor candidates, i.e., by choosing

- the sequence μ_2 as the r-successor candidate of μ_1,
- the sequence μ_3 as the s-successor candidate of μ_2, and
- the sequence μ_1 as, both, the r-successor and s-successor candidates of μ_3.

To verify whether this structure and each $(E_i)_{\bowtie t} \in \nu_j$ are consistent with the threshold semantics, it is enough to check that the value v_i in μ_j is the membership degree computed by procedure DEGREES w.r.t. E_i, ν_j and the tuples of degrees $(v_1, \dots, v_5)$ of the types conforming the role-successor candidates of μ_j. For instance, for $(E_1)_{\geq 1/2}$ and μ_3, we have two such tuples:

$$(r, (1/2, 0, 1, 0, 1)) \quad \text{and} \quad (s, (1/2, 0, 1, 0, 1)),$$

since μ_1 is an $\{r, s\}$-successor candidate of μ_3. One can verify that DEGREES returns $1/2$ (w.r.t. these tuples, E_1 and ν_3), which is the value of v_1 in μ_3. $\triangle$

Using the extended notions of type and r-successor candidate, Algorithm 4 starts by invoking procedure T-ELIM to perform type elimination w.r.t. $\mathfrak{T}$. The subroutine SUCC decides whether a sequence $\mu^1, \dots, \mu^\ell$ is an r-successor candidate of a given type μ. Note that this includes the empty sequence, which we denote as ε. As we will see later, when proving completeness of the algorithm, it suffices to consider sequences of length at most $N_S := 2 \cdot \mathsf{s}(\mathfrak{T}) + |\{r(a,b) \in \mathcal{A}\}|$ as such candidates. The loop at Line 20 and the conditional at Line 22 check whether the degrees $v_1, \dots, v_m$ of μ and a given assignment of candidate successors to roles in $\mathfrak{R}(\mathfrak{T}, \sim)$ are compatible w.r.t. $m_\sim$. Note that, in contrast to Algorithm 2, the call of the procedure DEGREES uses the values that are part of the type rather than already computed values. For this reason, it must be tested whether these values are consistent with how the graded membership function produces such values. Finally, once type elimination is finished, the algorithm tries to find a consistent assignment of types (from the remaining set of types) to the individuals in the ABox.

To show that Algorithm 4 is correct, we start with proving soundness. The overall idea of the proof is, starting from a successful run of the algorithm, to first select for each type $\mu \in S$ a mapping ψ of role-successor candidates of μ that passes the tests in procedure SUCC w.r.t. μ. The second step consists of using the types in S and the selected role-successor candidates to define an interpretation $\mathcal{I}$. Finally, we show that $\mathcal{I}$ is a model of $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$.

Lemma 5.2. *If Algorithm 4 answers "yes", then $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is consistent.*

Proof. If the algorithm answers "yes", then there is a mapping t satisfying the test at Line 4. Hence, for each $\mu \in S$ we can select a mapping $\psi^\mu : \mathfrak{R}(\mathfrak{T}, \sim) \to Seq$ that leads to a successful execution of the call

- SUCC$(t(b), S, \mathfrak{T}, b, t)$, if $\mu = t(b)$ for some $b \in \mathsf{Ind}(\mathcal{A})$, and of
- SUCC$(\mu, S, \mathfrak{T}, \perp, \perp)$, otherwise.

The former is the case since the test in Line 4 was successful and the later since μ was not removed during type elimination. Using t and the mappings ψ^μ, we define an interpretation $\mathcal{I}$ as follows:

Algorithm 4 Consistency of $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$ in $\tau\mathcal{EL}(m_\sim)$, where $\sim \in$ *simi-d*.

Input: A $\tau\mathcal{EL}(m)$ KB $(\mathfrak{T}, \mathcal{A})$, $\alpha \in \mathsf{ND}$ and $a \in \mathsf{NI}$.
Output: "yes", if $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is consistent in $\tau\mathcal{EL}(m_\sim)$, "no" otherwise.

1: $S := \text{T-ELIM}(\mathfrak{T})$
2: $T :=$ all mappings $t : \mathsf{Ind}(\mathcal{A}) \to S$ such that $(X(b) \in \mathcal{A} \cup \{\neg\alpha(a)\}) \Rightarrow X \in \nu_b$, where $t(b) = (\nu_b, \dots)$
3: **for all** $t \in T$ **do**
4:     **if** $\text{SUCC}(t(b), S, \mathfrak{T}, b, t) =$ true for all $b \in \mathsf{Ind}(\mathcal{A})$ **then return** "yes"
5: **return** "no"

6: **procedure** $\text{T-ELIM}(\mathfrak{T})$
7:     $S_0 := \{\mu \mid \mu$ is a type for $\mathfrak{T}\}$
8:     $i := 0$
9:     **repeat**
10:         $S_{i+1} := S_i \setminus \{\mu \mid \text{SUCC}(\mu, S_i, \mathfrak{T}, \bot, \bot) = \text{false}\}$
11:         $i := i + 1$
12:     **until** $S_i = S_{i-1}$
13:     **return** S_i

14: **procedure** $\text{SUCC}(\mu, S, \mathfrak{T}, b, t)$
15:     $Seq := \{\mu^1, \dots, \mu^\ell \mid \mu^j \in S \wedge 1 \le j \le \ell \le N_S\} \cup \{\varepsilon\}$
16:     let Ψ be the set of mappings $\psi : \mathfrak{R}(\mathfrak{T}, \sim) \to Seq$.
17:     **iterate over** $\psi \in \Psi$ such that
18:             $r \in \mathsf{rol}(\nu) \Rightarrow \psi(r)$ is an r-successor candidate of μ w.r.t. $\mathsf{cl}(\mathfrak{T}, \sim)$ **and**
19:             $(b \neq \bot$ **and** $r(b, c) \in \mathcal{A}) \Rightarrow t(c)$ is in $\psi(r)$
20:         **for all** $1 \le i \le m$ **do**
21:             $q_i := \text{DEGREES}(\nu \cap \mathsf{NC}, \{(r, (v'_1, \dots, v'_m)) \mid r \in \mathfrak{R}(\mathfrak{T}, \sim) \wedge (\nu', v'_1, \dots, v'_m)$ is in $\psi(r)\}, E_i)$
22:         **if** $q_i = v_i$ for all $1 \le i \le m$ **then return** true
23:     **return** false

- $\Delta^{\mathcal{I}} := \{d_\mu \mid \mu \in S\}$,
- $A^{\mathcal{I}} := \{d_\mu \mid A \in \nu\}$ for all $A \in \mathsf{NC} \cup \mathsf{ND}$,
- $r^{\mathcal{I}} := \{(d_{\mu_1}, d_{\mu_2}) \mid r \in \mathfrak{R}(\mathfrak{T}, \sim)$ and μ_2 is a tuple in $\psi^{\mu_1}(r)\}$ for all $r \in \mathsf{NR}$,
- $b^{\mathcal{I}} := d_{t(b)}$ for all $b \in \mathsf{Ind}(\mathcal{A})$.

We next show that $\mathcal{I}$ is a model of $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$. Let us start with the ABox. Since $t \in T$, the definition of T at Line 2 tells us that $X(b) \in \mathcal{A} \cup \{\neg\alpha(a)\}$ implies that $X \in \nu_b$. Thus, we obtain $d_{t(b)} \in X^{\mathcal{I}}$ by the construction of $\mathcal{I}$ and (if $X = \neg\alpha$) by t3). Hence, $b^{\mathcal{I}} = d_{t(b)}$ implies that $b^{\mathcal{I}} \in X^{\mathcal{I}}$ for all $X(b) \in \mathcal{A} \cup \{\neg\alpha(a)\}$. Next, consider a role assertion $r(b, c) \in \mathcal{A}$. We know that $r \in \mathfrak{R}(\mathfrak{T}, \sim)$, since r occurs in $\mathfrak{T}$ by our assumptions on $\mathcal{A}$ and $\mathfrak{T}$. Therefore, $t(c)$ occurs in $\psi^{t(b)}(r)$ since $\psi^{t(b)}$ leads to a successful execution of $\text{SUCC}(t(b), S, \mathfrak{T}, b, t)$. Hence, $(b^{\mathcal{I}}, c^{\mathcal{I}}) \in r^{\mathcal{I}}$ follows by construction of $\mathcal{I}$. Thus, we have shown that $\mathcal{I} \models \mathcal{A}$.

It remains to show that $\mathcal{I}$ is a model of $\mathfrak{T}$. To this end, we prove first that $\mathcal{I}$ is a model of $\mathcal{T}$, and afterwards that the following holds for all $d_\mu \in \Delta^{\mathcal{I}}$ (the proofs are deferred to Appendix C):

$$\text{for all } \tau\mathcal{EL}(m) \text{ concepts } \widehat{C} \in \mathsf{cl}(\mathfrak{T}, \sim)\text{:} \quad \widehat{C} \in \nu \text{ iff } d_\mu \in \widehat{C}^{\mathcal{I}}. \tag{37}$$

We can then use (37) to show that $\mathcal{I}$ is a model of $\widehat{\mathcal{T}}$. Let $d_\mu \in \Delta^{\mathcal{I}}$ and $\widehat{C} \sqsubseteq \widehat{D} \in \widehat{\mathcal{T}}$. We know that $\{\widehat{C}, \widehat{D}\} \subseteq \mathsf{cl}(\mathfrak{T}, \sim)$. Moreover, since μ is a type for $\mathfrak{T}$, we have that $\widehat{C} \in \nu$ implies $\widehat{D} \in \nu$. Hence, applying (37) twice we obtain that $d_\mu \in \widehat{C}^{\mathcal{I}}$ implies $\widehat{C} \in \nu$ implies $\widehat{D} \in \nu$ implies $d_\mu \in \widehat{D}^{\mathcal{I}}$. Therefore, since d_μ and $\widehat{C} \sqsubseteq \widehat{D}$ were arbitrarily chosen, it follows that $\mathcal{I} \models \widehat{\mathcal{T}}$. Thus, we have shown that $\mathcal{I} \models \mathfrak{T}$. □

We continue by showing completeness of Algorithm 4. This requires to show that, given a satisfiable KB $\mathcal{K}$ of the form $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$, Algorithm 4 has a successful run on input $(\mathfrak{T}, \mathcal{A})$ and a. The proof consists of two main steps. First, we take a finite model $\mathcal{I}$ of $\mathcal{K}$ and extract from it the types and role-candidate successors representing the individuals of $\Delta^{\mathcal{I}}$, their membership degrees and the relational structure of $\mathcal{I}$, as illustrated in Example 5.1. Afterwards, we show how to use these types and role-candidate successors to obtain a successful run of Algorithm 4.

Lemma 5.3. *If $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is consistent, then Algorithm 4 answers "yes".*

Proof. Let $\mathcal{I}$ be a model of $(\mathfrak{T}, \mathcal{A})$ with $a^{\mathcal{I}} \notin \alpha^{\mathcal{I}}$. Due to Proposition 3.13, we can assume that $\mathcal{I}$ is finite, since the DL $\mathcal{ALC}$ enjoys the finite model property [2]. Let us associate a tuple $\mu_d := (\nu_d, v_1^d, \dots, v_m^d)$ with each $d \in \Delta^{\mathcal{I}}$, where $\nu_d := \{X \mid X \in \mathsf{cl}(\mathfrak{T}, \sim)$ and $d \in X^{\mathcal{I}}\}$ and $v_i^d := \widehat{m}_\sim^{\mathcal{I}}(d, E_i, \mathcal{T})$. Given $b \in \mathsf{Ind}(\mathcal{A})$, we denote as μ_b the tuple corresponding to $b^{\mathcal{I}}$. In addition, let $N_{\mathcal{I}} := \bigcup_{d \in \Delta^{\mathcal{I}}} \{\mu_d\}$. To show that Algorithm 4 answers "yes", we prove that:

- $N_{\mathcal{I}} \subseteq S$ and the mapping $t(b) := \mu_b$ $(b \in \mathsf{Ind}(\mathcal{A}))$ is in T and satisfies the condition at Line 4.

Let us start by showing that μ_d is a type for $\mathfrak{T}$ for all $d \in \Delta^{\mathcal{I}}$. By definition, we have that $\nu_d \subseteq \mathsf{cl}(\mathfrak{T}, \sim)$, and the semantics of the concept constructors ensures that ν_d satisfies conditions t1) - t3) w.r.t. $\mathsf{cl}(\mathfrak{T}, \sim)$. Therefore, ν_d is a type for $\mathsf{cl}(\mathfrak{T}, \sim)$. In addition, the definition of $\widehat{m}$ implies that v_i^d is equal to $u_{\mathcal{T}}(E_i) \sim D$ for some $\mathcal{EL}$ concept D. Hence, an application of Lemma 3.26 yields that $v_i^d \in \mathcal{U}_{\mathcal{T},\sim}$ for all $i, 1 \le i \le m$. As for condition t4), by assumption we know that threshold concepts in $\mathsf{cl}(\mathfrak{T}, \sim)$ are of the form $(E_i)_{\bowtie t}$ where $E_i \in \mathsf{NC}_d^{\mathcal{T}}$. The semantics of threshold concepts w.r.t. a TBox states that $d \in [(E_i)_{\bowtie t}]^{\mathcal{I}}$ iff $\widehat{m}_\sim^{\mathcal{I}}(d, E_i, \mathcal{T}) \bowtie t$ (see 8). Hence, the definitions of v_d and v_i^d ensure that μ_d satisfies t4). Overall, we have shown that μ_d is a type for $(\mathsf{cl}(\mathfrak{T}, \sim), \Gamma_\tau, \mathcal{U}_{\mathcal{T},\sim})$. Finally, since $\mathcal{I} \models \mathfrak{T}$, the definition of v_d implies that v_d satisfies the GCIs and definitions in $\mathfrak{T}$. Thus, μ_d is a type for $\mathfrak{T}$. As an immediate consequence we obtain that $N_{\mathcal{I}} \subseteq S_0$.

We continue by proving that, given $d \in \Delta^{\mathcal{I}}$, the execution of $\textsc{succ}(\mu_d, N_{\mathcal{I}}, \mathfrak{T}, \bot, \bot)$ does not fail. This can then be used to show, by induction on i, that $N_{\mathcal{I}} \subseteq S_i$ and the execution of $\textsc{succ}(\mu_d, S_i, \mathfrak{T}, \bot, \bot)$ does not fail. We have $N_{\mathcal{I}} \subseteq S_0$ and non-failure of $\textsc{succ}(\mu_d, N_{\mathcal{I}}, \mathfrak{T}, \bot, \bot)$ implies non-failure also for the call with the superset S_0 of $N_{\mathcal{I}}$. This shows that none of the tuples $\mu_d \in N_{\mathcal{I}}$ is removed in this step, and thus $N_{\mathcal{I}} \subseteq S_1$. If we already know that $N_{\mathcal{I}} \subseteq S_i$, we can show in the same way that the call $\textsc{succ}(\mu_d, S_i, \mathfrak{T}, \bot, \bot)$ does not fail, and thus that $N_{\mathcal{I}} \subseteq S_{i+1}$. This finishes our induction proof, and thus yields that $N_{\mathcal{I}} \subseteq S$ and the execution of $\textsc{succ}(\mu_d, S, \mathfrak{T}, \bot, \bot)$ does not fail, where S is the final result of type elimination.

To show that the call $\textsc{succ}(\mu_d, N_{\mathcal{I}}, \mathfrak{T}, \bot, \bot)$ does not fail, we use d to define a mapping $\psi^d : \mathfrak{R}(\mathfrak{T}, \sim) \to Seq$ as follows:

- For each $r \in \mathfrak{R}(\mathfrak{T}, \sim)$, we consider the set Δ_r^d of r-successors of d in $\mathcal{I}$ defined in Lemma 4.13.
- The mapping ψ^d defines $\psi^d(r)$ as the sequence consisting of all types μ_e such that:
  - $e \in \Delta_r^d$, or $d = b^{\mathcal{I}}$, $e = c^{\mathcal{I}}$ and $r(b, c) \in \mathcal{A}$.

  Since $|\Delta_r^d| \le 2 \cdot \mathsf{s}(\mathfrak{T})$ (see Lemma 4.13), it is easy to see that $\psi^d(r)$ contains at most N_S tuples. Hence, $\psi^d(r)$ is an element in Seq.

We are now ready to show that ψ^d leads to a successful execution of $\textsc{succ}(\mu_d, N_{\mathcal{I}}, \mathfrak{T}, \bot, \bot)$:

- First, we show that, for all $r \in \mathsf{rol}(v_d)$, the sequence $\psi^d(r) = \mu^1, \ldots, \mu^\ell$ is an r-successor candidate of μ_d w.r.t. $\mathsf{cl}(\mathfrak{T}, \sim)$. This requires checking whether the sequence $v^1, \ldots, v^\ell$ satisfies condition c1) for all $\exists r.\widehat{C} \in \mathsf{cl}(\mathfrak{T}, \sim)$. Assume that $\widehat{C} \in v_d^+(r)$. By definition of Δ_r^d, there is $e \in \Delta_r^d$ such that $(d, e) \in r^{\mathcal{I}}$ and $e \in \widehat{C}^{\mathcal{I}}$. This implies that $\mu^j = \mu_e$ for some $j, 1 \le j \le \ell$. Hence, since $\widehat{C} \in \mathsf{cl}(\mathfrak{T}, \sim)$, the definition of v_e yields that $\widehat{C} \in v_e = v^j$. Conversely, suppose that $\widehat{C} \in v^j$ for some $j, 1 \le j \le \ell$. By construction of $\psi^d(r)$ and Δ_r^d, there is $e \in \Delta^{\mathcal{I}}$ such that $\mu_e = \mu^j$ and $(d, e) \in r^{\mathcal{I}}$. Hence, $\widehat{C} \in v^j$ and $v_e = v^j$ imply that $e \in \widehat{C}^{\mathcal{I}}$, and therefore $d \in (\exists r.\widehat{C})^{\mathcal{I}}$ and $\exists r.\widehat{C} \in v_d$. Thus, by definition of $v_d^+(r)$ we obtain that $\widehat{C} \in v_d^+(r)$.
- Next, we show that the mapping ψ^d also satisfies the condition at Line 22. Consider the set $Q := \{(r, (v_1^e, \ldots, v_m^e)) \mid e \in \Delta^{\mathcal{I}} \wedge (d, e) \in r^{\mathcal{I}} \wedge r \in \mathsf{NR}\}$. Since $\mathcal{I}$ is a finite model of $\mathcal{T}$, we can apply Lemma 4.11 to obtain that

  $$\textsc{degrees}(v_d \cap \mathsf{NC}, Q, E_i) = \widehat{m}_\sim^{\mathcal{I}}(d, E_i, \mathcal{T}) = v_i \quad (1 \le i \le m).$$

  Let $Q' := \{(r, (v_1', \ldots, v_m')) \mid r \in \mathfrak{R}(\mathfrak{T}, \sim) \wedge (v', v_1', \ldots, v_m') \text{ is in } \psi^d(r)\}$. To show that ψ^d satisfies the condition at Line 22, it is enough to prove that $\textsc{degrees}(v_d \cap \mathsf{NC}, Q, E_i) = \textsc{degrees}(v_d \cap \mathsf{NC}, Q', E_i)$. The definition of ψ^d yields $Q' \subseteq Q$, which implies that $\textsc{degrees}(v_d \cap \mathsf{NC}, Q', E_i) \le \textsc{degrees}(v_d \cap \mathsf{NC}, Q, E_i)$. Hence, it remains to see that the inequality in the other direction holds as well.

  Clearly, the values added to *aux* at Line 4 of the procedure $\textsc{degrees}$ are the same w.r.t. both Q and Q'. As for Line 6, given $\exists s.E_p \in top(C_{E_i})$, let $(r, q) \in Q$ be a tuple yielding the maximal value added to *aux* w.r.t. Q. We distinguish the following two cases:

  - If $r \notin \mathfrak{R}(\mathfrak{T}, \sim)$, then $pm(s, r) = 0$. Hence, *aux* is not incremented for $\exists s.E_p$.
  - Assume that $r \in \mathfrak{R}(\mathfrak{T}, \sim)$. Since $(r, q) \in Q$, this implies that d has r-successors in $\mathcal{I}$. Therefore, by construction, Δ_r^d contains an r-successor e of d in $\mathcal{I}$ such that:

    $$pm(s, r)[w + (1 - w)v_p^e] = \max\{pm(s, r)[w + (1 - w)v_p^f] \mid (d, f) \in r^{\mathcal{I}}\}.$$

    This means that $\psi^d(r)$ contains the tuple $(v_e, v_1^e, \ldots, v_m^e)$, and hence, Q' contains the tuple $(r, (v_1^e, \ldots, v_m^e))$. Hence, the value added to *aux* for $\exists s.E_p$ w.r.t. Q is not greater than the one added w.r.t. Q'.

  This shows that $\textsc{degrees}(v_d \cap \mathsf{NC}, Q, E_i) \le \textsc{degrees}(v_d \cap \mathsf{NC}, Q', E_i)$.

Overall, we have proved that $\textsc{succ}(\mu_d, N_{\mathcal{I}}, \mathfrak{T}, \bot, \bot)$ returns true for all $d \in \Delta^{\mathcal{I}}$, and thus that $N_{\mathcal{I}} \subseteq S$.

Regarding the mapping t, which we have defined as $t(b) := \mu_b$, first note that it really is a mapping into S since $\mu_b \in N_{\mathcal{I}} \subseteq S$. In addition, since $\mathcal{I}$ is a model of $\mathcal{A} \cup \{\neg\alpha(a)\}$, it follows that v_b satisfies the implication at Line 2 for all $b \in \mathsf{Ind}(\mathcal{A})$. This shows that $t \in T$. We have already shown that $\textsc{succ}(t(b), S, \mathfrak{T}, \bot, \bot)$ does not fail. To see that the same is true for the call $\textsc{succ}(t(b), S, \mathfrak{T}, b, t)$, we need to show that the condition in Line 19 is satisfied. This is the case since $\psi^d(r)$ contains the tuple μ_c whenever $d = b^{\mathcal{I}}$ and $r(b, c) \in \mathcal{A}$. Thus, we can conclude that Algorithm 4 answers "yes". □

We conclude this section by analyzing the complexity of Algorithm 4. Let $n := |S_0|$. It is easy to see that $|T| \le n^{|\mathcal{A}|}$, $|Seq| \le (n+1)^{N_S}$ and $|\Psi| \le |Seq|^{|\mathfrak{R}(\mathfrak{T},\sim)|}$. Moreover, the sets T, Seq and Ψ are obtained starting from a subset S of S_0. Regarding S_0, let $\ell := |\mathsf{cl}(\mathfrak{T}, \sim)|$

and $k := |\mathcal{U}_{\mathcal{T},\sim}|$. Then, S_0 contains at most $2^{\ell} \cdot k^m$ types, where $m = |\mathsf{NC}_d^{\mathcal{T}}|$. By definition, each rational number in $\mathcal{U}_{\mathcal{T},\sim}$ is of the form y/x_E, where $E \in \mathsf{NC}_d^{\mathcal{T}}$ and x_E is the number obtained in Lemma 3.26 for $u_{\mathfrak{T}}(E)$. As explained before, Lemma 3.27 can be reused to show that x_E has a binary representation of size polynomial in $\mathsf{s}(\mathcal{T})$. Hence, $|\mathcal{U}_{\mathcal{T},\sim}| \leq m \cdot 2^{\mathsf{s}(\mathcal{T})}$ and the cardinality of S_0 is exponential in $\mathsf{s}(\mathfrak{T})$. Further, all concept and role names occurring in $u_{\mathcal{T}}(E)$ occur in $\mathcal{T}$ and every conjunction in $\mathcal{S}_{\sqcap}(u_{\mathcal{T}}(E_i))$ corresponds to the unfolding of the right-hand side of a definition in $\mathcal{T}$. Hence, Lemma 3.27 can also be reused to show that x_E can be computed in polynomial time in $\mathsf{s}(\mathcal{T})$. This means that S_0 can be computed in exponential time in $\mathsf{s}(\mathfrak{T})$. Hence, since N_S and $\mathfrak{R}(\mathfrak{T},\sim)$ are computable in polynomial time in $\mathsf{s}(\mathfrak{T}) + \mathsf{s}(\mathcal{A})$ and in $\mathsf{s}(\mathfrak{T})$, respectively, one can see that T, Seq and Ψ are computable in exponential time in $\mathsf{s}(\mathfrak{T})$, as well. Based on this, it is not hard to see that Algorithm 4 runs in exponential time in the size of the input. In addition, since a mapping $t \in T$ and a sequence in Seq can both be guessed in polynomial time in $\mathsf{s}(\mathcal{A})$, Algorithm 4 can be easily turned into an algorithm that decides non-instance in non-deterministic polynomial time in $\mathsf{s}(\mathcal{A})$. Thus, together with the previously stated lower bounds, we obtain the following results.

Theorem 5.4. *Let $\sim \in$ simi-d. In $\tau\mathcal{EL}(m_\sim)$, satisfiability, subsumption, consistency and instance checking w.r.t. general $\tau\mathcal{EL}(m_\sim)$ TBoxes are ExpTime-complete. Regarding data complexity, instance checking is coNP-complete.*

6. Related work

The use of graded membership functions with values in the interval $[0,1]$ in our definition of the semantics of threshold operators may remind the reader of fuzzy logics [39], and in particular of fuzzy description logics [40–42]. However, a closer look reveals that fuzzy DLs and our logics $\tau\mathcal{EL}(m)$ are quite different. The main difference is that the semantics of logics of the form $\tau\mathcal{EL}(m)$ are defined using classical first-order interpretations, whereas fuzzy DLs are based on fuzzy interpretations, where concept and role names are interpreted as fuzzy sets and relations. In a fuzzy DL, the concept name Tall may be interpreted by a fuzzy set that assigns the membership degree .8 to the domain element d that interprets the individual SAM, and the degree .4 in another interpretation. In a threshold logic $\tau\mathcal{EL}(m_\sim)$ for $\sim \in$ *simi-d*, the concept name Tall is interpreted as a crisp set. If the domain element d that interprets the individual SAM belongs to this set, then $m_\sim$ yields the degree 1 for d. Otherwise, the membership degree assigned by $m_\sim$ to d depends on what are the other concepts to which d belongs. For example, if d belongs to the concept Lanky and the primitive measure employed for defining $\sim$ assigns .8 to the pair (Tall, Lanky), then $m_\sim$ may assign .8 to d in case Lanky is the concept that yields the maximum degree. Another difference is that the semantics of fuzzy DLs is compositional in the sense that the degree of membership in a composite concept is determined by the degrees of membership *values* in its subconcepts. For example, the fuzzy membership degree $(\mathsf{Tall} \sqcap \mathsf{Happy})(d)$ is computed as $\mathsf{Tall}(d) \otimes \mathsf{Happy}(d)$, where $\otimes$ is the t-norm used to interpret the conjunction operator. Our graded membership functions need not be compositional. There are various other technical differences, but the take home message is that fuzzy description logics are non-classical logics with a many-valued semantics, which are very often undecidable [43–45], whereas the results of this paper show that the logics $\tau\mathcal{EL}(m_\sim)$ are decidable fragments of classical first-order logic for computable standard CMs that have computable r-reducing and s-reducing functions (see Lemma 3.11 and Theorem 3.14).

As mentioned in Section 3, our way of transforming CMs $\sim$ into graded membership functions $m_\sim$ (see Definition 3.2) is inspired by the approach proposed in [12,29] to relax instance queries using concept similarity measures. This approach proceeds as follows: given a query concept C, a CSiM $\sim$, and a threshold value t, compute all classical answers to query concepts D such that $C \sim D > t$. In this way, individuals a that are not instances of C, but are instances of a different concept D similar enough to C, are considered to be relaxed answers of C. This approximation framework is investigated in [12,29] for the DL $\mathcal{EL}$, i.e., in a setting where the knowledge base consists of an $\mathcal{EL}$ TBox and ABox, and where the query is an $\mathcal{EL}$ concept. In our framework, the relaxed instance queries of [12,29] can be expressed as instance queries w.r.t. threshold concepts of the form $C_{>t}$ in the threshold DL $\tau\mathcal{EL}(m_\sim)$. On the one hand, the DL $\tau\mathcal{EL}(m_\sim)$ yields a more expressive query language since we can formulate threshold concepts using also other comparison operators and can combine such concepts with the basic $\mathcal{EL}$ constructors to form complex query concepts like $(\exists r.A)_{<0.5} \sqcap \exists r.((A \sqcap B)_{\geq 0.8}) \sqcap B$. In addition, we can use such complex concepts of $\tau\mathcal{EL}(m_\sim)$ also in the TBox and ABox. On the other hand, the work in [12,29] also considers relaxed instance queries based on a family of CSiMs that are defined w.r.t. general $\mathcal{EL}$ TBoxes. In contrast, the CMs and graded membership functions considered in our work are defined for $\mathcal{EL}$ concepts only, and are extended to acyclic TBoxes using unfolding. It would be interesting to see whether our approach can be extended to a setting that uses CSiMs w.r.t. general $\mathcal{EL}$ TBoxes as defined in [12,29], but since this would require new ideas and considerable additional work, it is beyond the scope of this article.

In [46], new concept constructors that look similar to our threshold concepts are introduced. The paper proposes an extension to the DL $\mathcal{ALC}$ that uses "prototypes" to define concepts. To accomplish this, it defines the notion of a prototype distance function (pdf), which assigns to each element of an interpretation a distance value in the nonnegative integers (intuitively: the distance of the individual from the prototype). This function induces a new concept constructor of the form $P_{\bowtie t}(d)$ for $\bowtie \in \{<, \leq, >, \geq\}$, which is interpreted as the set of all elements with a distance $\bowtie t$ according to the pdf d. Such distance functions are similar to our graded membership functions, but instead of specifying the prototype by a concept, weighted alternating parity tree automata (wapta) over the nonnegative integers are used in [46] to define pdfs. This allows the authors to show that reasoning in the new DL $\mathcal{ALCP}(\textit{wapta})$, which extends the DL $\mathcal{ALC}$ with the constructors $P_{\bowtie t}(d)$ for pdfs d defined using wapta, can be reduced to decidable decision problems for weighted alternating parity tree automata.

Another approach that exploits distances and prototypes to define approximate concepts is the one introduced in [47]. The authors propose a new DL *sim*-$\mathcal{ALCQO}$ that combines the DL $\mathcal{ALCQO}$ with the logic *sim* introduced in [48] for reasoning about metric spaces. The fusion of $\mathcal{ALCQO}$ and *sim* extends each interpretation $\mathcal{I}$ with a distance function d such that $\langle \Delta^{\mathcal{I}}, d \rangle$ is a metric space. In particular, the logic allows to use a concept C to represent the prototypical individuals of a particular class of elements, and then build concepts of the form $\mathsf{E}^{<t}C$ to capture all the individuals that are *similar* to a prototypical element of such a class within a threshold value $t \in \mathbb{Q}^+$. The notion of similarity is measured using the distance function. A more complex logic with additional constructors can be found in [49]. In principle, one could think about expressing the threshold concept $C_{\geq t}$ of $\tau\mathcal{EL}(m)$ using the *sim*-$\mathcal{ALCQO}$ concept $\mathsf{E}^{\leq(1-t)}C$. However, the main difference between the work in [47,49] and ours is that in [47,49] the authors do not fix a specific distance function. Let us illustrate this difference with an example.

Example 6.1. Consider the graded membership m_s introduced in Example 2.12. Further, let C be the $\mathcal{EL}$ concept $\exists r.A \sqcap \exists s.A \sqcap B$. In the logic $\tau\mathcal{EL}(m_s)$ we have, for all interpretations $\mathcal{I}$ and $e \in \Delta^{\mathcal{I}}$, that:

$$m_s^{\mathcal{I}}(e, C) \geq 1/3 \text{ iff } e \text{ is an instance of at least one top-level atom of } C.$$

Then, the semantics of $\tau\mathcal{EL}(m_s)$ defines $(C_{\geq 1/3})^{\mathcal{I}}$ as the set of all such elements $e \in \Delta^{\mathcal{I}}$. In contrast, in *sim*-$\mathcal{ALCQO}$, the concept $\mathsf{E}^{\leq 2/3}C$ is defined in an interpretation $\mathcal{I}$ as:

$$\text{the set of elements } e \in \Delta^{\mathcal{I}} \text{ such that } d(e, f) \leq 2/3 \text{ for some } f \in C^{\mathcal{I}},$$

where d is the distance function associated to $\mathcal{I}$. Obviously, there can be functions d for which $d(e, f) \leq 2/3$ holds exactly for those elements e in $(C_{\geq 1/3})^{\mathcal{I}}$. However, this will not be the case for the majority of such functions, since they are only required to define a metric space over $\Delta^{\mathcal{I}}$ without taking into account which concepts the elements of $\Delta^{\mathcal{I}}$ satisfy or not. This tells us that $C_{\geq 1/3}$ cannot be expressed by using $\mathsf{E}^{\leq 2/3}C$, nor by using any *sim*-$\mathcal{ALCQO}$ concept of the form $\mathsf{E}^{\bowtie t}D$ where D is an $\mathcal{EL}$ concept. △

To summarize, reasoning in *sim*-$\mathcal{ALCQO}$ is done w.r.t. all possible distance functions, and thus the reasoning algorithms can only take into account general properties of such functions. In contrast, we consider in $\tau\mathcal{EL}(m)$ as specific membership function m, and thus reasoning can (and must) take the specific properties of this function into account.

A new concept constructor called the “tooth operator” is introduced in [50], and its properties are further investigated in [51] and [52]. The authors propose a DL $\mathcal{ALC}_{\mathbb{W}^{\mathbb{R}}}$ that extends $\mathcal{ALC}$ with a new m-ary operator $\mathbb{W}$. This constructor defines concepts of the form $\mathsf{C} := \mathbb{W}_w^t(C_1, \ldots, C_m)$, where $w \in \mathbb{R}^m$ is a vector of weights, $C_1, \ldots, C_m$ are $\mathcal{ALC}$ concepts, and $t \in \mathbb{R}$ is a threshold value. To define the semantics for this concept constructor, they use a function v that, given C and an interpretation $\mathcal{I}$, assigns a value in $\mathbb{R}$ to each element $d \in \Delta^{\mathcal{I}}$ as follows:

$$v_{\mathsf{C}}^{\mathcal{I}}(d) := \sum_{i=1}^{m} \{w_i \mid d \in C_i^{\mathcal{I}}\}.$$

The interpretation of $\mathsf{C} = \mathbb{W}_w^t(C_1, \ldots, C_m)$ under $\mathcal{I}$ is then the following set:

$$\mathbb{W}_w^t(C_1, \ldots, C_m)^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid v_{\mathsf{C}}^{\mathcal{I}}(d) \geq t\}.$$

The general idea is thus similar to our threshold logics, where the function $v_{\mathsf{C}}^{\mathcal{I}}$ acts as the membership function and the semantics of $\mathbb{W}$ follows the same principle as the semantics of our threshold concepts.

From a technical point of view there seem to be two main differences to our approach. The first one is that the tooth operator takes m concepts $C_1, \ldots, C_m$ and an m-tuple of weights as input, whereas our threshold operator is applied to a single concept. The following example provides more insights into this.

Example 6.2. Consider again the graded membership function m_s from Example 2.12. If one assumes that the concepts $C_1, \ldots, C_m$ are subsumption-incomparable $\mathcal{EL}$ atoms and the weights are all equal to $1/m$, then the tooth concept $\mathbb{W}_w^t(C_1, \ldots, C_m)$ and the $\tau\mathcal{EL}(m_s)$ concept $(C_1 \sqcap \cdots \sqcap C_m)_{\geq t}$ for $t \in [0, 1]$ have the same semantics. A concrete example of this are the following “tooth concept” and threshold concept:

$$\mathbb{W}_w^{1/3}(\exists r.A, \exists s.A, B) \quad \text{and} \quad C_{\geq 1/3},$$

where $w = (1/3, 1/3, 1/3)$ and $C = \exists r.A \sqcap \exists s.A \sqcap B$. In case the vector w is not uniform, the different weights in w could partially be accommodated by changing the definition of m_s such that it uses weights of atoms. However, these atom weights would then be fixed by the definition of the graded membership function and could not be changed in different applications of the tooth operator. For instance, consider the following $\mathcal{ALC}_{\mathbb{W}^{\mathbb{R}}}$ and $\tau\mathcal{EL}(m_s)$ concepts:

$$\mathbb{W}_{w_1}^{1/3}(\exists r.A, \exists s.A, B) \sqcap \mathbb{W}_{w_2}^{0.55}(\exists r.A, \exists s.B, A) \quad \text{and} \quad C_{\geq 1/3} \sqcap D_{\geq 0.55},$$

where $w_1 = (1/3, 1/3, 1/3)$, $w_2 = (1/4, 1/2, 1/4)$ and $D = \exists r.A \sqcap \exists s.B \sqcap A$. The different weights in w_2 imply that $D_{\geq 0.55}$ and $\mathbb{W}_{w_2}^{0.55}(\exists r.A, \exists s.B, A)$ do not have the same semantics. To repair this, one would need to adjust m_s to weigh the atoms in D according to w_2. However, this would break the equivalence between $C_{\geq 1/3}$ and $\mathbb{W}_{w_1}^{1/3}(\exists r.A, \exists s.A, B)$, since $\exists r.A$ is shared by both applications of the tooth operator. △

The second difference is that our approach considers a family of graded membership functions that allows the use of more sophisticated notions of membership. One example of this are the membership functions $m_\sim$ induced by CMs $\sim$ in *simi-d*. Unlike the tooth operator and the membership function m_s, such functions $m_\sim$ can take into account how an individual d *partially satisfies* the top-level atoms of a concept C to compute the membership degree of d into C. Let us illustrate this with one final example.

Example 6.3. Recall the threshold logic $\tau\mathcal{EL}(m_{\sim_d^*})$, where $m_{\sim_d^*}$ is the membership function induced by the CM $\sim_d^*$ from Example 3.16. In addition, consider the following threshold concept:

$$C_{\geq 1/2}, \text{ where } C = \exists r.(A \sqcap B) \sqcap \exists s.(A \sqcap B).$$

By definition of $\sim_d^*$, it is easy to see that:

$$C \sim_d^* \exists r.(A \sqcap B) \geq 1/2 \quad \text{and} \quad C \sim_d^* \exists s.(A \sqcap B) \geq 1/2.$$

This implies that, in $\tau\mathcal{EL}(m_{\sim_d^*})$, the interpretation of $C_{\geq 1/2}$ includes all elements that satisfy at least one top-level atom of C. The same is the case for $C_{\geq 1/2}$ in $\tau\mathcal{EL}(m_s)$ and for $\mathbb{W}^{1/2}_{(1/2,1/2)}(\exists r.(A \sqcap B), \exists s.(A \sqcap B))$ in $\mathcal{ALC}_{\mathbb{W}^\mathbb{R}}$. However, the recursive definition of $\sim_d^*$ (as an instance of Definition 3.15) also yields that:

$$C \sim_d^* (\exists r.A \sqcap \exists s.B) \geq 1/2 \quad \text{and} \quad C \sim_d^* (\exists r.B \sqcap \exists s.A) \geq 1/2. \tag{38}$$

This means that $m_{\sim_d^*}$ can take into account different forms of partially satisfying the top-level conjuncts in C to compute membership degrees. For instance, given an interpretation $\mathcal{I}$ and $e \in \Delta^\mathcal{I}$, one can derive from the expressions in (38) that:

$$\left(e \in \exists r.A^\mathcal{I} \text{ and } e \in \exists s.B^\mathcal{I}\right) \text{ or } \left(e \in \exists r.B^\mathcal{I} \text{ and } e \in \exists s.A^\mathcal{I}\right) \text{ implies } m^\mathcal{I}_{\sim_d^*}(e, C) \geq 1/2.$$

Hence, e can be an instance of $C_{\geq 1/2}$, without satisfying at least one top-level atom of C. Consequently, $C_{\geq 1/2}$ cannot be expressed in $\mathcal{ALC}_{\mathbb{W}^\mathbb{R}}$ by just using a concept of the form $\mathbb{W}^t_w(\exists r.(A \sqcap B), \exists s.(A \sqcap B))$, for some value t and tuple w. Overall, the semantics of the tooth operator is not enough to succinctly express arbitrary $\tau\mathcal{EL}(m_{\sim_d^*})$ threshold concepts. $\triangle$

Summing up, one can say that our threshold logics and the logic $\mathcal{ALC}_{\mathbb{W}^\mathbb{R}}$ appear to be incomparable. On the one hand, $\mathcal{ALC}_{\mathbb{W}^\mathbb{R}}$ allows for a more flexible use of weights and extends $\mathcal{ALC}$ rather than $\mathcal{EL}$. On the other hand, the semantics of the tooth operator is based on a graded membership function that is much simpler than the ones of the form $m_\sim$ introduced in the present paper. Strictly speaking, both kinds of logic can be expressed within $\mathcal{ALC}$, but they provide their users with compact representations of certain large disjunctions, which can be handled by dedicated reasoning procedures in a more efficient way than their representation by disjunctions in $\mathcal{ALC}$ by an $\mathcal{ALC}$ reasoner.

7. Conclusions

Motivated by the need for defining concepts in an approximate way, we have introduced an approach for extending the DL $\mathcal{EL}$ by threshold concepts. The semantics of such threshold concepts is defined using the notion of a graded membership function. We have shown that concept measures, which generalize subsumption or equivalence between concepts, can be used to define graded membership functions that yield extensions of $\mathcal{EL}$ with good algorithmic properties. The results presented in this article considerably generalize our previous work on threshold logics presented in three conference publications [11,17,19] by extending the general framework to infinite signatures, dealing both with acyclic and with general TBoxes for a larger class of graded membership functions defined using concept measures, and providing exact complexity results for all the classical reasoning problems. Our results imply that the investigated extensions of $\mathcal{EL}$ can be seen as fragments of the DL $\mathcal{ALC}$. However, employing threshold concepts nevertheless makes sense since they can express huge disjunctions in a very compact manner, which is very useful both for modelling and for reasoning purposes. Our complexity results show that reasoning in our threshold logics has basically the same complexity as reasoning in $\mathcal{ALC}$ (with the exception of the case without TBox, where threshold logics have a lower complexity). However, expressing threshold concepts by disjunctions in $\mathcal{ALC}$ and then employing an $\mathcal{ALC}$ reasoner could cause a non-elementary complexity blowup.

One important direction for future work is to assess the practicality and applicability of reasoning with our threshold logics. We envision two main initial steps to be taken in this direction.

- The first step is to implement a $\tau\mathcal{EL}(m_\sim)$ reasoner. Note that our algorithms are designed with the purpose of obtaining optimal complexity bounds for the considered reasoning problems, and they rely on the use of non-determinism or type elimination. For this reason, a direct implementation of our algorithms is very unlikely to perform well in practice. To overcome this, a first step would be to turn our algorithms into *tableau*-based decision procedures that can handle the semantics of threshold concepts. An implementation of these tableau algorithms can then take advantage of optimization techniques, such as *lazy unfolding*, *absorption* and special forms of *blocking*, which have proven to be very successful in obtaining practical implementations of tableau algorithms for reasoning in expressive DLs [53,54].
- Once a reasoner for $\tau\mathcal{EL}(m_\sim)$ is implemented, the second step would be to evaluate the performance of this reasoner and the usefulness of our threshold logics. This requires to identify an appropriate application domain where the use of threshold concepts

makes sense. As already mentioned in the introduction, match-making is one typical example of such an application domain, where people search for products with many possible configurations, e.g., rental flats or bicycles. In this scenario, when there is no product exactly matching an user request, answering instance queries in a logic $\tau\mathcal{EL}(m_\sim)$ can be useful for finding and recommending products that are similar enough to the ones an user is looking for. Obviously, the quality of the obtained recommendations depends on the measure $\sim$ used to define $m_\sim$. In particular, for the family of CMs *simi-d* that we investigate, the definition of $\sim$ depends on several numerical parameters, as described in Subsection 3.3. The selection of these parameters plays an important role in finding good concept measures for the described approximate query answering scenario. Hence, it would be interesting to investigate whether existing *parameters tuning approaches* (or other techniques) can be used to learn optimal values for these parameters. Overall, these ideas can be combined to design an initial experiment, which would provide preliminary results on the applicability and practicality of reasoning in the family of logics $\tau\mathcal{EL}(m_\sim)$.

From a theoretical point of view, one interesting topic for future research would be to extend our approach for approximately defining concepts to more expressive DLs, starting with extensions of $\mathcal{EL}$ with the *bottom* ($\bot$) and *nominal* constructors, as well as *role inclusions* (i.e., fragments of the DL $\mathcal{EL}^{++}$ [7]), and then continue with the more expressive DL $\mathcal{ALC}$. The following observations provide some insights into these possible extensions.

- In the case of the DL $\mathcal{ELH}_\bot$, which extends $\mathcal{EL}$ with $\bot$ and *role hierarchies*, one can take the *homomorphism degree function* introduced in [55] to compare concepts in the more expressive DL $\mathcal{ALEH}$, and restrict it to obtain CMs that compare $\mathcal{ELH}_\bot$ concepts. Since such a function is defined similarly to the directional measure $\sim_d$ of the *simi* framework, it seems possible to obtain a family of CMs satisfying similar properties as the measures in *simi-d*, to which we can then apply the mechanism described in Section 3 to obtain well-defined graded membership functions that take $\mathcal{ELH}_\bot$ concepts as input. Overall, it should be easy to extend our results to the resulting new family of threshold logics. In fact, our techniques already take into account $\bot$ since it can be expressed in $\tau\mathcal{EL}(m)$, whereas role hierarchies do not fundamentally affect the structure and size of the kind of models our algorithms are based on. In contrast, it is not clear how to extend our approach to take into account nominals and more complex role inclusions of the form $s \circ r \sqsubseteq t$. The definition of CMs in *simi-d* and our bounded model properties largely rely on the fact that concepts can be represented as tree structures. This, however, is no longer possible in the presence of nominals and more expressive role inclusions.
- Regarding $\mathcal{ALC}$, a first step could be to add threshold concepts $C_{\bowtie t}$ for $\mathcal{EL}$ concept descriptions C to $\mathcal{ALC}$. The decidability result in Theorem 3.14, which is based on translating threshold concepts into $\mathcal{ALC}$ concepts, would still hold, but it is not clear how to obtain more precise complexity results for such extensions of $\mathcal{ALC}$. Even more challenging would be to define reasonable graded membership functions that can take $\mathcal{ALC}$ concepts rather than just $\mathcal{EL}$ concepts as input.

Another interesting topic would be to investigate how one can answer conjunctive queries w.r.t. ontologies written in one of our threshold logics. Finally, it would be interesting to investigate threshold logics where graded membership functions are used as parameters for individual threshold concepts rather than for the whole logic. In such an extension, one could, e.g., consider a conjunction $C^{m_1}_{>t_1} \sqcap D^{m_2}_{\le t_2}$, where the semantics of the first threshold concept is defined using the graded membership function m_1 and of the second using m_2.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in Graduiertenkolleg 1763 (QuantLA) and in project number 335448072.

Appendix A. Missing details from Section 3

This appendix is devoted to provide missing details from Subsections 3.3 and 3.4. More precisely, we complete the proof that CSuMs in *simi-d* are standard, revisit the class of CSiMs *simi-mon* and show that they are also standard, and conclude by giving the full proofs of Proposition 3.24 and Lemma 3.26. Some of the proofs that follow use the following property of $\sim_d$ (shown in [14]):

$$D \sqsubseteq E \Rightarrow C \sim_d E \le C \sim_d D. \tag{39}$$

A.1. Properties of CMs in simi-d

Lemma 3.17. *Let $\sim_d$ be a directional instance of simi. Then, $\sim_d$ is role-depth reducing.*

Proof. We prove that setting $\mathfrak{r}_{\sim_d}(C) = \mathsf{rd}(C)$ (for all $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$) yields an r-reducing function for $\sim_d$. Let $C, D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$ and $k = \mathfrak{r}_{\sim_d}(C)$. By Lemma 2.10, we have $\mathsf{rd}(D|_k) \leq \mathfrak{r}_{\sim_d}(C)$ and $D \sqsubseteq D|_k$. Hence, to show that $\mathfrak{r}_{\sim_d}$ is r-reducing it suffices to prove that $C \sim_d D \leq C \sim_d D|_k$. We prove this by well-founded induction on $\mathfrak{r}_{\sim_d}(C)$.

First, note that the case where $C \equiv \top$ is trivial since then $C \sim_d D = 1 = C \sim_d D|_k$. If $C \not\equiv \top$ and $D \equiv \top$, then also $D|_k \equiv \top$, which shows that $C \sim_d D = 0 = C \sim_d D|_k$. Otherwise, the definition of *simi* yields

$$C \sim_d D = \frac{\sum\limits_{C' \in top(C)} \left[g(C') \cdot \max\{simi_a(C', D') \mid D' \in top(D)\} \right]}{\sum\limits_{C' \in top(C)} g(C')}$$

and

$$C \sim_d D|_k = \frac{\sum\limits_{C' \in top(C)} \left[g(C') \cdot \max\{simi_a(C', D^*) \mid D^* \in top([D|_k]^r)\} \right]}{\sum\limits_{C' \in top(C)} g(C')}.$$

Let C' be an arbitrary top-level atom of C, and F' and F^* top-level atoms of D and $[D|_k]^r$, respectively, maximizing the values $simi_a(C', D')$ and $simi_a(C', D^*)$ in the two expressions above. We prove that $simi_a(C', F') \leq simi_a(C', F^*)$ by distinguishing two cases:

- $C' = A \in \mathsf{NC}$. By definition, $simi_a(C', X) = 0$ for all X of the form $\exists r.X'$. Since the concept names occurring as top-level atoms in D and $[D|_k]^r$ are the same, it follows that $simi_a(C', F') \leq simi_a(C', F^*)$.
- $C' = \exists r.C''$. The case $F' \in \mathsf{NC}$ is trivial since then $simi_a(C', F') = 0$. Thus, consider the case where F' is of the form $F' = \exists s.F''$. The definition of $simi_a$ yields

  $$simi_a(C', F') = pm(r, s)[w + (1 - w)(C'' \sim_d F'')].$$

  Clearly, we have $\mathsf{rd}(C'') < \mathsf{rd}(C)$, and thus $\mathfrak{r}_{\sim_d}(C'') < \mathfrak{r}_{\sim_d}(C)$. Hence, we can apply induction to C'' and F'' to obtain $C'' \sim_d F'' \leq C'' \sim_d F''|_\rho$, where $\rho = \mathfrak{r}_{\sim_d}(C'')$. The definition of $\mathfrak{r}$ yields $\rho \leq k - 1$, which implies $F''|_{k-1} \sqsubseteq F''|_\rho$. An application of (39) thus yields

  $$C'' \sim_d F'' \leq C'' \sim_d F''|_\rho \leq C'' \sim_d F''|_{k-1}.$$

  Let us now consider $simi_a(C', F'|_k)$. Since $k = \mathsf{rd}(C) > 0$ and $F' = \exists s.F''$, we know that $F'|_k = \exists s.F''|_{k-1}$. Consequently, the following holds:

  $$simi_a(C', F') \leq pm(r, s)[w + (1 - w)(C'' \sim_d F''|_{k-1})] = simi_a(C', F'|_k).$$

  Finally, since $F'|_k \in top(D|_k)$, either $F'|_k \in top([D|_k]^r)$ or there exists $E \in top([D|_k]^r)$ such that $E \sqsubseteq F'|_k$. By (39), this subsumption yields $C' \sim_d F'|_k \leq C' \sim_d E$, which is the same as saying that $simi_a(C', F'|_k) \leq simi_a(C', E)$ since C', $F'|_k$, and E are atoms. Since F^* was supposed to be the top-level atom in $D|_k$ that yields the maximal value, we have the following sequence of inequations:

  $$simi_a(C', F') \leq simi_a(C', F'|_k) \leq simi_a(C', E) \leq simi_a(C', F^*).$$

Overall, we have thus shown that the following holds for all $C' \in top(C)$:

$$\max\{simi_a(C', D') \mid D' \in top(D)\} \leq \max\{simi_a(C', D^*) \mid D^* \in top([D|_k]^r)\},$$

and thus that $C \sim_d D \leq C \sim_d D|_k$. $\square$

A.2. Properties of CMs in simi-mon

The purpose of this subsection is, on the one hand, to proof Proposition 3.24 and Lemma 3.26. On the other hand, we want to correct a small error in [17] regarding the proof of well-definedness of $m_\sim$ for all $\sim \in$ *simi-mon*. In fact, in [17], a more general well-definedness result was claimed for all instances $\sim$ of *simi* where g assigns 1 to atoms of the form $\exists r.C$. However, Lemma 3.22 shows that this claim is wrong. Below, we give a direct proof of well-definedness for the case of *simi-mon* as introduced in Definition 3.23.

The properties equivalence closedness and equivalence invariance are satisfied by all CSiM $\sim$ obtained from *simi*. The former was already shown in [14], whereas the latter is an easy consequence of the facts that $C \sim D$ is computed using the reduced forms of C and D and that, up to associativity and commutativity of $\sqcap$, equivalent $\mathcal{EL}$ concepts have the same reduced form. Verifying that all measures $\sim \in$ *simi-mon* are also role-depth and signature reducing is, however, more involved. We first need to show that they satisfy some additional properties, which we illustrate in the following example before formulating and proving them in the general case.

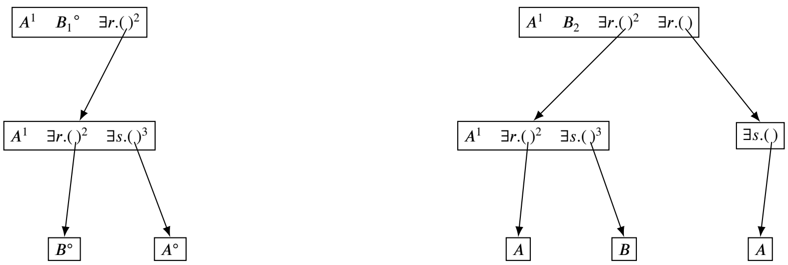


Fig. 6. Computation of $\sim_d$ in *simi*.

Example A.1. We consider the CSM $\sim^{.5}$ whose definition deviates from the one of $\sim^*$ only in one place: we use $w{=}.5$. Consider the concept descriptions C, D of Example 3.16:

$$C := A \sqcap B_1 \sqcap \exists r.(A \sqcap \exists r.B \sqcap \exists s.A),$$

$$D := A \sqcap B_2 \sqcap \exists r.(A \sqcap \exists r.A \sqcap \exists s.B) \sqcap \exists r.\exists s.A.$$

Fig. 6 basically shows the atoms in D chosen by max when computing $C \sim_d^{.5} D$. The *superscripts* are used to denote the corresponding pairings for which the value is > 0. For instance, at the top level of C, A^1 means that A is paired with the top-level atom of D having the same superscript. The symbol $\circ$ on the left-hand side tells us that no match yielding a value > 0 exists. Now, removing the atoms without superscript in D yields the concept $D' := A \sqcap \exists r.(A \sqcap \exists r.\top \sqcap \exists s.\top)$. One can easily verify that $C \sim_d D{=}C \sim_d D'{=}5/9$, and it is clear that C and D are both subsumed by D'. $\triangle$

These properties can be generalized to all pair of concept descriptions and measures in *simi-mon*, as stated in the following lemma.

Lemma A.2. *Let $\sim \in$ simi-mon. For all $\mathcal{EL}$ concept descriptions C and D, there exists an $\mathcal{EL}$ concept description Y such that:*

1. $C \sqsubseteq Y$, $D \sqsubseteq Y$ *and* $\mathsf{s}(Y) \leq \mathsf{s}(C)$,
2. $C \sim_d D = C \sim_d Y$.

Proof. We use induction on the structure of C to prove the claim.

- *C is equal to $A \in \mathsf{NC}$ or $\top$.* If $C = A$, then
$$C \sim_d D = \frac{g(A) \cdot \max\{simi_a(A, D') \mid D' \in top(D)\}}{g(A)}.$$
Since $pm = pm_{df}$, this means that $A \sim_d D = 1$ if $A \in top(D)$, and $A \sim_d D = 0$ otherwise. Choosing $Y := A$ in the first case and $Y := \top$ in the second thus ensures that the properties required for Y in the lemma hold.
If $C \equiv \top$, then the definition of $\sim_d$ implies $C \sim_d X = 1$ for all concept descriptions X. Thus, $Y := \top$ satisfies the required properties.
- *C is of the form $\exists r.E'$.* Let D^* be the top-level atom of D maximizing the value $simi_a(C, D')$. If $D^* \in \mathsf{NC}$, then $simi_a(C, D^*) = 0$ and $C \sim_d D = 0$. Then, $Y := \top$ satisfies the required properties. Otherwise, D^* is of the form $\exists s.F'$. Then,
$$C \sim_d D = pm(r,s)[w + (1-w)(E' \sim_d F')]$$
For $r \neq s$ we have $pm(r,s) = 0$ and $C \sim_d D = 0$. In this case, we can again choose $Y := \top$. Otherwise, applying the induction hypothesis to E' and F' yields a concept description Y' such that:
  - $E' \sqsubseteq Y'$, $F' \sqsubseteq Y'$, and $\mathsf{s}(Y') \leq \mathsf{s}(E')$,
  - $E' \sim_d F' = E' \sim_d Y'$,

  Let $Y := \exists r.Y'$. Since $r = s$, we have $C = \exists r.E' \sqsubseteq Y$ and $D \sqsubseteq \exists r.F' \sqsubseteq Y$. In addition $\mathsf{s}(Y) = 1 + \mathsf{s}(Y') \leq 1 + \mathsf{s}(E') = \mathsf{s}(C)$ and $C \sim_d Y = pm(r,r)[w + (1-w)(E' \sim_d Y')] = C \sim_d D$.
- *$C = C_1 \sqcap \cdots \sqcap C_n$ with $n > 1$, where the C_i are atoms.* In this case we have:
$$C \sim_d D = \frac{\sum_{j=1}^{n}\Big[g(C_j) \cdot \max\{simi_a(C_j, D') \mid D' \in top(D)\}\Big]}{\sum_{j=1}^{n} g(C_j)}.$$

Let D_j $(1 \leq j \leq n)$ be a top-level atom of D that maximizes the value $simi_a(C_j, D')$ among all $D' \in top(D)$. The application of the induction hypothesis to C_j and D_j yields a concept description Y_j such that:

- $C_j \sqsubseteq Y_j$, $D_j \sqsubseteq Y_j$, and $\mathsf{s}(Y_j) \leq \mathsf{s}(C_j)$,
- $C_j \sim_d D_j = C_j \sim_d Y_j$,

Obviously, $C_1 \sqcap \cdots \sqcap C_n \sqsubseteq Y_1 \sqcap \cdots \sqcap Y_n$ and $D_1 \sqcap \cdots \sqcap D_n \sqsubseteq Y_1 \sqcap \cdots \sqcap Y_n$. Therefore, the concept description $Y := Y_1 \sqcap \cdots \sqcap Y_n$ satisfies $C \sqsubseteq Y$, $D \sqsubseteq Y$, and clearly also $\mathsf{s}(Y) \leq \mathsf{s}(C)$. Thus, it remains to show that $C \sim_d D = C \sim_d Y$. The value of $C \sim_d Y$ is given by the following expression:

$$C \sim_d Y = \frac{\sum_{j=1}^{n} \left[g(C_j) \cdot \max\{simi_a(C_j, Y') \mid Y' \in top(Y^r)\} \right]}{\sum_{j=1}^{n} g(C_j)}. \tag{40}$$

Consider an arbitrary $j, 1 \leq j \leq n$, and an atom $Y^* \in top(Y^r)$ such that $simi_a(C_j, Y^*)$ yields the maximum among all the values $simi_a(C_j, Y')$ for $Y' \in top(Y^r)$. We show that $simi_a(C_j, Y_j) = simi_a(C_j, Y^*)$.

- First, we show $simi_a(C_j, Y_j) \leq simi_a(C_j, Y^*)$. Clearly, we only need to consider the case where $Y_j \notin top(Y^r)$. This means that Y_j is removed when computing Y^r from Y. Hence, there exists $X \in top(Y^r)$ such that $X \sqsubseteq Y_j$. By (39), it follows that $C_j \sim_d Y_j \leq C_j \sim_d X$. Since C_j, Y_j and X are atoms, this means that $simi_a(C_j, Y_j) = C_j \sim_d Y_j \leq C_j \sim_d X = simi_a(C_j, X)$. Thus, $simi_a(C_j, Y_j) \leq simi_a(C_j, X) \leq simi_a(C_j, Y^*)$.
- Second, we show that $simi_a(C_j, Y_j) \geq simi_a(C_j, Y^*)$. We only consider the case $Y_j \neq Y^*$. Let D_ℓ be the top-level atom of D from which Y^* was obtained. Then,

$$\begin{aligned}
simi_a(C_j, Y_j) &= C_j \sim_d Y_j && (C_j \text{ and } Y_j \text{ are atoms})\\
&= C_j \sim_d D_j\\
&= simi_a(C_j, D_j) && (C_j \text{ and } D_j \text{ are atoms})\\
&\geq simi_a(C_j, D_\ell)\\
&= C_j \sim_d D_\ell \geq C_j \sim_d Y^* = simi_a(C_j, Y^*). && (D_\ell \sqsubseteq Y^*, (39) \text{ and } C_j, Y^*, D_\ell \text{ are atoms})
\end{aligned}$$

Overall, we have shown (for all $1 \leq j \leq n$) that $simi_a(C_j, Y_j)$ is the maximal value among all the values i $simi_a(C_j, Y')$ in (40). Thus, it follows that $C \sim_d D = C \sim_d Y$.

Since C must be of one of the three forms considered in the above case distinction, this completes the proof of the lemma. □

Using this lemma, we can now show that CSiMs in *simi-mon* satisfy the remaining properties that are required for CMs to be standard.

Proposition A.3. *All measures $\sim \in$ simi-mon are standard CMs.*

Proof. We have already seen that CMs $\sim$ belonging to *simi-mon* are equivalence invariant and equivalence closed. Thus, it is sufficient to show role-depth and signature reduction. We do this by proving that the functions $\mathfrak{s}_\sim(C) := \mathsf{sig}(C)$ and $\mathfrak{r}_\sim(C) := \mathsf{rd}(C)$ (for all $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$) are, respectively, s-reducing and r-reducing functions for $\sim$.

Let $C, D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$. By Lemma A.2, there exists D' such that $C \sim_d D = C \sim_d D'$, $D \sqsubseteq D'$ and $C \sqsubseteq D'$. Since $C \sqsubseteq D'$ and $\sim_d$ is subsumption closed (as shown in [14]), we know that $D' \sim_d C = 1$. Hence, monotonicity of $\otimes$ implies that $C \sim D \leq C \sim D'$. In addition, the fact that $C \sqsubseteq D'$ implies $\mathsf{sig}(D') \subseteq \mathfrak{s}_\sim(C)$ and $\mathsf{rd}(D') \leq \mathfrak{r}_\sim(C)$ by (3). Thus, D' is a witness for signature and role-depth reduction of $\sim$. □

Together with Proposition 3.6, the above proposition implies that all CSiMs $\sim \in$ *simi-mon* induce a well-defined membership function $m_\sim$, and hence a threshold DL $\tau\mathcal{EL}(m_\sim)$. An additional consequence of Lemma A.2 is the following result, which we have used in Section 3 to argue why it is sufficient to concentrate our attention on directional measures in *simi-d*.

Proposition 3.24. *Let $\sim \in$ simi-mon and $\sim_d$ be the associated directional measure. For all $\mathcal{EL}$ concept descriptions C, interpretations $\mathcal{I}$ and $e \in \Delta^\mathcal{I}$, there exists an $\mathcal{EL}$ concept description D such that $e \in D^\mathcal{I}$ and D satisfies the following:*

- $m^\mathcal{I}_\sim(e, C) = (C \sim_d D) \otimes 1$,
- $m^\mathcal{I}_{\sim_d}(e, C) = C \sim_d D$.

Proof. Let $F \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$ be such that $e \in F^{\mathcal{I}}$ and $m_{\sim}^{\mathcal{I}}(e, C) = C \sim F$. Consider the $\mathcal{EL}$ concept description F' obtained by applying Lemma A.2 to C and F. Since $C \sqsubseteq F'$, subsumption closedness of $\sim_d$ implies that $F' \sim_d C = 1$. Moreover, monotonicity of $\otimes$ yields that $C \sim F \leq C \sim F'$. Hence, since $F \sqsubseteq F'$ implies that $d \in F'^{\mathcal{I}}$, it follows that $m_{\sim}^{\mathcal{I}}(e, C) = C \sim F' = (C \sim_d F') \otimes 1$.

Let $G \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$ be such that $e \in G^{\mathcal{I}}$ and $C \sim_d G$ is maximal among all values $C \sim_d X$ for concept descriptions $X \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC}, \mathsf{NR})$ such that $e \in X^{\mathcal{I}}$. We now apply Lemma A.2 to C and G. By using similar arguments as above for C and F, we obtain a concept description G' such that $e \in G'^{\mathcal{I}}$, $C \sim_d G = C \sim_d G'$, and $C \sim G' = (C \sim_d G') \otimes 1$. Our choice of G implies that $C \sim_d F' \leq C \sim_d G = C \sim_d G'$, and thus monotonicity of $\otimes$ yields $C \sim F' \leq C \sim G'$. This shows that $m_{\sim}^{\mathcal{I}}(e, C) = C \sim G' = (C \sim_d G') \otimes 1$. Thus, if we set $D := G'$, then D satisfies all the properties required by the proposition. $\square$

We conclude this subsection with the proof of Lemma 3.26.

Lemma 3.26. *Let $\sim_d \in$ simi-d and let the parameter w be of the form w_n/w_d. In addition, let $C \in \mathcal{C}_{\mathcal{EL}}$, let $k = \mathsf{rd}(C)$, and define*

$$x_C := w_d^k \cdot \prod_{q \in d_{pm}(C)} q^k \cdot \prod_{F \in \mathcal{S}_{\sqcap}(C)} g_F^k.$$

Then, for all $D \in \mathcal{C}_{\mathcal{EL}}$, there exists $y \in \mathbb{N}$ such that $C \sim_d D = y/x_C$.

Proof. We prove the claim by induction on the role depth of C. Let C be of the form

$$A_1 \sqcap \cdots \sqcap A_m \sqcap \exists r_{m+1}.C_{m+1} \sqcap \cdots \sqcap \exists r_{m+\ell}.C_{m+\ell},$$

where $m \geq 0$, $\ell \geq 0$, $m + \ell > 0$, and we assume without loss of generality that C is reduced. For convenience, let us denote as $g_{i,1}/g_{i,2}$ the values $g(A_i)$ $(1 \leq i \leq m)$ and $g(\exists r_i.C_i)$ $(m < i \leq m + \ell)$. In addition, let D be an arbitrary $\mathcal{EL}$ concept. For all $1 \leq i \leq m + \ell$, we select $B_i \in top(D^r) \cap \mathsf{NC}$ (if $i \leq m$) and $\exists s_i.D_i \in top(D^r)$ (if $i > m$) such that

- $simi_a(A_i, B_i)$ and $simi_a(\exists r_i.C_i, \exists s_i.D_i)$ are, respectively, the maximal value of $simi_a(A_i, D')$ and $simi_a(\exists r_i.C_i, D')$ for $D' \in top(D^r)$ in the definition of $C \sim_d D$ in (13).

We then have:

$$C \sim_d D = \frac{\sum_{i=1}^{m} g_{i,1}/g_{i,2} \cdot simi_a(A_i, B_i) + \sum_{i=m+1}^{m+\ell} g_{i,1}/g_{i,2} \cdot simi_a(\exists r_i.C_i, \exists s_i.D_i)}{\sum_{i=1}^{m+\ell} g_{i,1}/g_{i,2}}. \tag{41}$$

To continue, for all $1 \leq i \leq m + \ell$, we define p_i/q_i as follows:

$$p_i/q_i := \begin{cases} pm(A_i, B_i), & \text{if } i \leq m, \\ pm(r_i, s_i), & \text{if } i > m. \end{cases}$$

Further, the application of induction to C_i $(i > m)$ yields that $C_i \sim_d D_i = y_i/x_{C_i}$, and we can then express $simi_a(\exists r_i.C_i, \exists s_i.D_i)$ as

$$\frac{p_i}{q_i}\left[\frac{w_n}{w_d} + (1 - \frac{w_n}{w_d}) \cdot (C_i \sim_d D_i)\right] = \frac{p_i}{q_i}\left[\frac{w_n}{w_d} + (1 - \frac{w_n}{w_d}) \cdot \frac{y_i}{x_{C_i}}\right] = \frac{p_i \cdot (w_n \cdot x_{C_i} + (w_d - w_n) \cdot y_i)}{q_i \cdot w_d \cdot x_{C_i}}.$$

Note that $x_{C_i} = w_d^{k_i} \cdot \prod_{q \in d_{pm}(C_i)} q^{k_i} \cdot \prod_{F \in \mathcal{S}_{\sqcap}(C_i)} g_F^{k_i}$, where $k_i = \mathsf{rd}(C_i)$. Hence, from the previous expression we obtain:

$$simi_a(\exists r_i.C_i, \exists s_i.D_i) = \frac{p_i \cdot (w_n \cdot x_{C_i} + (w_d - w_n) \cdot y_i)}{q_i \cdot w_d \cdot w_d^{k_i} \cdot \prod_{q \in d_{pm}(C_i)} q^{k_i} \cdot \prod_{F \in \mathcal{S}_{\sqcap}(C_i)} g_F^{k_i}}. \tag{42}$$

Using (42), the numerator of the expression in (41) can be rewritten as:

$$\sum_{i=1}^{m} \frac{g_{i,1}}{g_{i,2}} \cdot \frac{p_i}{q_i} + \sum_{i=m+1}^{m+\ell} \frac{g_{i,1}}{g_{i,2}} \cdot \frac{p_i \cdot (w_n \cdot x_{C_i} + (w_d - w_n) \cdot y_i)}{q_i \cdot w_d \cdot w_d^{k_i} \cdot \prod_{q \in d_{pm}(C_i)} q^{k_i} \cdot \prod_{F \in \mathcal{S}_{\sqcap}(C_i)} g_F^{k_i}}.$$

By definition, we know that $q_i \in d_{pm}(C)$, $d_{pm}(C_i) \subseteq d_{pm}(C)$ and $\mathcal{S}_{\sqcap}(C) = \{C\} \cup \bigcup_{i=m+1}^{m+\ell} \mathcal{S}_{\sqcap}(C_i)$. Hence, since $k_i < k$, the following is a *common multiple* of the denominators of the fractions composing the sum in the previous expression:

$$\prod_{i=1}^{m+\ell} g_{i,2} \cdot w_d^k \cdot \prod_{q \in d_{pm}(C)} q^k \cdot \prod_{F \in \mathcal{S}_{\sqcap}(C) \setminus \{C\}} g_F^k.$$

Hence, there is a natural number y such that $C \sim_d D$ can be expressed as

$$C \sim_d D = \frac{y}{\cancel{\prod_{i=1}^{m+\ell} g_{i,2}} \cdot w_d^k \cdot \prod_{q \in d_{pm}(C)} q^k \cdot \prod_{F \in S_{\sqcap}(C) \setminus \{C\}} g_F^k} \cdot \frac{\cancel{\prod_{i=1}^{m+\ell} g_{i,2}}}{\sum_{i=1}^{m+\ell} g_{i,1} \cdot \prod_{j=1, j \neq i}^{m+\ell} g_{j,2}}$$
$$= \frac{y}{w_d^k \cdot \prod_{q \in d_{pm}(C)} q^k \cdot \prod_{F \in S_{\sqcap}(C) \setminus \{C\}} g_F^k \cdot \left(\sum_{i=1}^{m+\ell} g_{i,1} \cdot \prod_{j=1, j \neq i}^{m+\ell} g_{j,2} \right)}.$$

Looking back at (17) we see that the fourth factor in the denominator of this expression is equal to the numerator of (17), and thus equal to g_C. Therefore, $C \sim_d D$ can be expressed as a fraction with a natural number y as numerator and denominator

$$x_C = w_d^k \cdot \prod_{q \in d_{pm}(C)} q^k \cdot \prod_{F \in S_{\sqcap}(C)} g_F^k.$$

Thus, we have shown that $C \sim_d D$ is of the claimed form y / x_C. □

Appendix B. Missing proofs from Section 4

In this appendix we provide the missing proofs from Section 4. We start with the proof of the bounded tree model property stated in Lemma 4.2. This result is then used to complete the proofs missing in Subsection 4.2.4. To conclude, we define the normal and reduced form introduced in Section 4 for acyclic $\tau\mathcal{EL}(m)$ TBoxes and prove related results.

B.1. Bounded tree model property

Lemma 4.2. *Let $\sim \in$ simi-d. In addition, let $\widehat{C}$ and $\widehat{D}$ be two $\tau\mathcal{EL}(m_\sim)$ concept descriptions. If $\widehat{C} \sqcap \neg \widehat{D}$ is satisfiable in $\tau\mathcal{EL}(m_\sim)$, then there is a tree-shaped interpretation $\mathcal{I}$ with root element d_0 such that:*

1. *$d_0 \in (\widehat{C} \sqcap \neg \widehat{D})^{\mathcal{I}}$, $|\Delta^{\mathcal{I}}| \leq \mathsf{s}(\widehat{C}) \cdot \mathsf{s}(\widehat{D})$, and*
2. *$\mathcal{I}$ has depth at most $\mathsf{rd}(\widehat{C} \sqcap \widehat{D})$.*

To prove the lemma, we first need to show some properties about the translation of threshold concepts into $\mathcal{ALC}$.

Lemma B.1. *Let $\sim \in$ simi-d and $C_{\bowtie t}$ be a threshold concept. Then, there exists a disjunction of $\mathcal{EL}$ concepts $\alpha = C_1 \sqcup \cdots \sqcup C_\ell$ such that $C_{\bowtie t} \equiv \alpha$ or $C_{\bowtie t} \equiv \neg\alpha$, and for all $1 \leq p \leq \ell$:*

- *$\mathsf{rd}(C_p) \leq \mathsf{rd}(C)$ and $\mathsf{s}(C_p) \leq \mathsf{s}(C)$.*

Proof. By Lemma 3.11, we know that $C_{\bowtie t} \equiv \alpha$ if $\bowtie \in \{>, \geq\}$, where α is the disjunction in (12). Further, since $C_{<t} \equiv \neg C_{\geq t}$ and $C_{\leq t} \equiv \neg C_{>t}$, we have that $\bowtie \in \{<, \leq\}$ implies $C_{\bowtie t} \equiv \neg\alpha$. Therefore, to show that an α exists satisfying the claimed properties, we can concentrate in the case where $\bowtie \in \{>, \geq\}$.

Each disjunct C_p of α has role depth at most $\mathfrak{r}_\sim(C)$. In Lemma 3.17, we have shown that for all $\sim \in$ *simi-d*, one can assume that $\mathfrak{r}_\sim(C) = \mathsf{rd}(C)$. Hence, $\mathsf{rd}(C_p) \leq \mathsf{rd}(C)$. Further, since $\sim \in$ *simi-d* is a directional instance of *simi*, we can apply to C and C_p the idea illustrated in Example A.1, and reuse the proof of Lemma A.2 to obtain an $\mathcal{EL}$ concept C'_p such that:

- $C \sim C_p = C \sim C'_p$, $C_p \sqsubseteq C'_p$ and $\mathsf{s}(C'_p) \leq \mathsf{s}(C)$.

Finally, $C_p \sqsubseteq C'_p$ and (3) imply $\mathsf{sig}(C'_p) \subseteq \mathsf{sig}(C_p)$ and $\mathsf{rd}(C'_p) \leq \mathsf{rd}(C_p)$. This, together with $C \sim C_p = C \sim C'_p$ and the definition of α in (12), yields that C'_p is a disjunct in α. Thus, it is easy to see that $\alpha' := C'_1 \sqcup \cdots \sqcup C'_\ell$ satisfies our claims. □

We now use the previous lemma to show Lemma 4.2. More precisely, we show a generalization of Lemma 4.2 to conjunctions containing more than one negated $\tau\mathcal{EL}(m)$ concept. This is formulated in the following lemma, which also states some additional properties. This more general claim will be useful later on when considering reasoning w.r.t. acyclic $\tau\mathcal{EL}(m)$ knowledge bases.

Lemma B.2. *Let $\sim \in$ simi-d and let $\widehat{C}$, $\widehat{D}_1, \ldots, \widehat{D}_m$ be $\tau\mathcal{EL}(m_\sim)$ concept descriptions. In addition, let $\mathcal{J}$ be an interpretation such that $d \in (\widehat{C} \sqcap \bigsqcap_{j=1}^{m} \neg \widehat{D}_j)^{\mathcal{J}}$ for some $d \in \Delta^{\mathcal{J}}$. Then, there exists a tree-shaped interpretation $\mathcal{I}$ with root element $e_0 \in (\widehat{C} \sqcap \bigsqcap_{j=1}^{m} \neg \widehat{D}_j)^{\mathcal{I}}$ such that:*

- *there is a homomorphism $\phi : \mathcal{I} \to \mathcal{J}$ with $\phi(e_0) = d$,*
- *$\mathcal{I}$ has depth at most $\mathsf{rd}(\widehat{C} \sqcap \widehat{D}_1 \sqcap \cdots \sqcap \widehat{D}_m)$, and*

- $|\Delta^{\mathcal{I}}| \leq \mathsf{s}(\widehat{C}) \cdot \prod\{\mathsf{s}(\widehat{D}_j) \mid 1 \leq j \leq m \wedge \widehat{D}_j \not\equiv \bot\}$.

Proof. The proof goes by induction on $\mathsf{rd}(\widehat{C} \sqcap \widehat{D}_1 \sqcap \cdots \sqcap \widehat{D}_m)$. Let us assume that $\widehat{C}$ and each $\widehat{D}_j$ are of the form $\widehat{C}_1 \sqcap \cdots \sqcap \widehat{C}_n$ and $\widehat{D}_{j,1} \sqcap \cdots \sqcap \widehat{D}_{j,m_j}$, respectively. By Lemma B.1, we know that each $\widehat{C}_i$ (similarly for each $\widehat{D}_{j,\iota}$) of the form $E_{\bowtie t}$ is equivalent to:

- a disjunction of $\mathcal{EL}$ concepts $E_{i,1} \sqcup \cdots \sqcup E_{i,\ell_i}$, if $\bowtie \in \{>, \geq\}$, or
- a conjunction of negated $\mathcal{EL}$ concepts $\neg E_{i,1} \sqcap \cdots \sqcap \neg E_{i,\ell_i}$, if $\bowtie \in \{<, \leq\}$.

Further, $d \in \widehat{C}^{\mathcal{J}}$ means that $d \in \widehat{C}_i^{\mathcal{J}}$ for all $1 \leq i \leq n$. Hence, $\widehat{C}_i = E_{\{>,\geq\}\,t}$ implies that $d \in E_{i,p}^{\mathcal{J}}$ for some $1 \leq p \leq \ell_i$. In such a case, we fix one such concept $E_{i,p}$ and call it X_i. As for $\widehat{C}_i = E_{\{<,\leq\}\,t}$, we have that $d \notin E_{i,p}^{\mathcal{J}}$ for all $1 \leq p \leq \ell_i$. Regarding each $\widehat{D}_j$, if $\widehat{D}_j \not\equiv \bot$ we select $1 \leq \iota \leq m_j$ such that $d \in \neg \widehat{D}_{j,\iota}^{\mathcal{J}}$ (it exists because $d \in \neg \widehat{D}_j^{\mathcal{J}}$). In case $\widehat{D}_{j,\iota} = E_{\{<,\leq\}\,t}$, we know that $\neg \widehat{D}_{j,\iota}$ is equivalent to a disjunction of $\mathcal{EL}$ concepts $F_1 \sqcup \cdots \sqcup F_\ell$ (since $\neg E_{<t} \equiv E_{\geq t}$ and $\neg E_{\leq t} \equiv E_{>t}$). Again, there is $1 \leq p \leq \ell$ such that $d \in F_p^{\mathcal{J}}$. We pick one such F_p and call it Y_j. In case $\widehat{D}_{j,\iota}$ is of a different form, we set $Y_j = \neg \widehat{D}_{j,\iota}$. Then, we define the set $\mathfrak{G} = \{\exists r_1.\widehat{G}_1, \ldots, \exists r_q.\widehat{G}_q\}$ containing the following existential restrictions:

- $\bigcup_{i=1}^{n} \left\{\exists r.\widehat{G} \mid \widehat{C}_i = \exists r.\widehat{G} \vee \left(\widehat{C}_i = E_{\{>,\geq\}t} \wedge \exists r.\widehat{G} \in top(X_i)\right)\right\}$, and
- $\bigcup_{j=1}^{m} \{\exists r.\widehat{G} \mid \exists r.\widehat{G} \in top(Y_j) \wedge 1 \leq \iota \leq m_j$ was chosen such that $\widehat{D}_{j,\iota} = E_{\{<,\leq\}\,t}\}$.

We next use $\mathfrak{G}$ to construct an interpretation $\mathcal{I}$ satisfying our claims. For each $1 \leq k \leq q$, we define $\widehat{H}_k^0$ and $\widehat{H}_k^1, \ldots, \widehat{H}_k^m$ as the following $\tau\mathcal{EL}(m)$ concepts:

- $\widehat{H}_k^0 = \widehat{G}_k$. If $Y_j = \neg\exists s.\widehat{F}$ was chosen and $s = r_k$, then $\widehat{H}_k^j = \widehat{F}$, else $\widehat{H}_k^j = A \sqcap (A_{<1})$ $(1 \leq j \leq m)$.

Since $d \in (\exists r_k.\widehat{G}_k)^{\mathcal{J}}$ for all $1 \leq k \leq q$, there exists $d_k \in \Delta^{\mathcal{J}}$ such that $(d, d_k) \in r_k^{\mathcal{J}}$ and $d_k \in \widehat{G}_k^{\mathcal{J}}$. Moreover, if $Y_j = \neg\exists s.\widehat{F}$ $(1 \leq j \leq m)$ with $s = r_k$, it must be that $d_k \in \neg\widehat{F}^{\mathcal{J}}$ since $d \in Y_j^{\mathcal{J}}$. Hence, $d_k \in (\widehat{H}_k^0 \sqcap \neg\widehat{H}_k^1 \sqcap \cdots \sqcap \neg\widehat{H}_k^m)^{\mathcal{J}}$. Moreover, by definition of $\mathfrak{G}$ and Lemma B.1, we obtain that:

- $\mathsf{rd}(\widehat{H}_k^0) < \mathsf{rd}(\widehat{C})$ or $\mathsf{rd}(\widehat{H}_k^0) < \mathsf{rd}(\widehat{D}_j)$ for some $j, 1 \leq j \leq m$, and
- $\mathsf{rd}(\widehat{H}_k^j) < \mathsf{rd}(\widehat{D}_j)$ or $\mathsf{rd}(\widehat{H}_k^j) = 0$, for all $j, 1 \leq j \leq m$.

Hence, $\mathsf{rd}(\widehat{H}_k^0 \sqcap \widehat{H}_k^1 \sqcap \cdots \sqcap \widehat{H}_k^m) < \mathsf{rd}(\widehat{C} \sqcap \widehat{D}_1 \sqcap \cdots \sqcap \widehat{D}_m)$ holds. Then, the application of induction to d_k and $\widehat{H}_k^0, \ldots, \widehat{H}_k^m$ yields a tree-shaped interpretation $\mathcal{I}_k$ with root element $e_k \in (\widehat{H}_k^0 \sqcap \bigsqcap_{j=1}^{m} \neg\widehat{H}_k^j)^{\mathcal{I}_k}$ such that:

- there is a homomorphism $\phi_k : \mathcal{I}_k \to \mathcal{J}$ with $\phi(e_k) = d_k$,
- $\mathcal{I}_k$ has depth at most $\mathsf{rd}(\widehat{H}_k^0 \sqcap \widehat{H}_k^1 \sqcap \cdots \sqcap \widehat{H}_k^m)$, and
- $|\Delta^{\mathcal{I}_k}| \leq \mathsf{s}(\widehat{H}_k^0) \cdot \prod\{\mathsf{s}(\widehat{H}_k^j) \mid 1 \leq j \leq m \wedge \widehat{H}_k^j \not\equiv \bot\}$.

We then combine all these interpretations to define $\mathcal{I}$ as follows:

- $\Delta^{\mathcal{I}} = \{e_0\} \cup \bigcup_{k=1}^{q} \Delta^{\mathcal{I}_k}$.
- $A^{\mathcal{I}} := \{e_0 \mid \text{if } d \in A^{\mathcal{I}}\} \cup \bigcup_{k=1}^{q} A^{\mathcal{I}_k}$, for all $A \in \mathsf{NC}$.
- $r^{\mathcal{I}} := \{(e_0, e_k) \mid r_k = r\} \cup \bigcup_{k=1}^{q} r^{\mathcal{I}_k}$, for all $r \in \mathsf{NR}$.

It is clear from its construction that $\mathcal{I}$ is tree-shaped. In addition, since each $\mathcal{I}_k$ has depth at most $\mathsf{rd}(\widehat{H}_k^0 \sqcap \widehat{H}_k^1 \sqcap \cdots \sqcap \widehat{H}_k^m)$ and $\mathsf{rd}(\widehat{H}_k^0 \sqcap \widehat{H}_k^1 \sqcap \cdots \sqcap \widehat{H}_k^m) < \mathsf{rd}(\widehat{C} \sqcap \widehat{D}_1 \sqcap \cdots \sqcap \widehat{D}_m)$, this means that $\mathcal{I}$ has depth at most $\mathsf{rd}(\widehat{C} \sqcap \widehat{D}_1 \sqcap \cdots \sqcap \widehat{D}_m)$. Regarding the size of $\mathcal{I}$, note that if Y_j $(1 \leq j \leq m)$ is not of the form $\neg\exists s.\widehat{F}$, then $\widehat{H}_k^j \equiv \bot$ for all $k, 1 \leq k \leq q$. Let $\mathfrak{A} \subseteq \{1, \ldots, m\}$ be the set of indices j such that Y_j is of the form $\neg\exists s.\widehat{F}$. For such indices, $\widehat{H}_k^j = \widehat{F}$ or $\widehat{H}_k^j \equiv \bot$. Hence, since $\exists s.\widehat{F}$ is a top-level atom of $\widehat{D}_j$, the following holds for all $1 \leq k \leq q$:

$$|\Delta^{\mathcal{I}_k}| \leq \mathsf{s}(\widehat{H}_k^0) \cdot \prod_{j \in \mathfrak{A}} \mathsf{s}(\widehat{D}_j) \tag{43}$$

Let us now look at the concepts $\widehat{H}_k^0$. Each $\exists r_k.\widehat{G}_k \in \mathfrak{G}$ is obtained either from a top-level atom $\widehat{C}_i$ of $\widehat{C}$ or from some $\widehat{D}_j$ such that $j \notin \mathfrak{A}$.

- $\widehat{C}_i$ can be of two forms. If $\widehat{C}_i = \exists r_k.\widehat{G}_k$, it contributes with $\exists r_k.\widehat{G}_k$ to $\mathfrak{G}$, $\widehat{H}^0_k = \widehat{G}_k$, and clearly $\mathsf{s}(\widehat{H}^0_k) < \mathsf{s}(\widehat{C}_i)$. Otherwise, $\widehat{C}_i = E_{\{>,\geq\}\,t}$. Then, a list $\exists r_{k_1}.\widehat{G}_{k_1}, \dots, \exists r_{k_\ell}.\widehat{G}_{k_\ell}$ $(1 \leq k_\ell \leq q)$ is added to $\mathfrak{G}$ such that $\exists r_{k_p}.\widehat{G}_{k_p} \in top(X_i)$ for all $1 \leq p \leq \ell$. By Lemma B.1 one can assume that $\mathsf{s}(X_i) \leq \mathsf{s}(E)$, and this implies that $\sum_{p=1}^{\ell} \mathsf{s}(\widehat{G}_{k_p}) = \sum_{p=1}^{\ell} \mathsf{s}(\widehat{H}^0_{k_p}) < \mathsf{s}(\widehat{C}_i)$.
- $\widehat{D}_j$ has a top-level atom $\widehat{D}_{j,\iota}$ of the form $E_{\{<,\leq\}\,t}$, and a list $\exists r_{k_1}.\widehat{G}_{k_1}, \dots, \exists r_{k_\ell}.\widehat{G}_{k_\ell}$ $(1 \leq k_\ell \leq q)$ is added to $\mathfrak{G}$ such that $\exists r_{k_p}.\widehat{G}_{k_p} \in top(Y_j)$ for all $1 \leq p \leq \ell$. Similarly to the previous case, Lemma B.1 yields that $\mathsf{s}(Y_i) \leq \mathsf{s}(E)$, and this implies that $\sum_{p=1}^{\ell} \mathsf{s}(\widehat{G}_{k_p}) = \sum_{p=1}^{\ell} \mathsf{s}(\widehat{H}^0_{k_p}) < \mathsf{s}(\widehat{D}_j)$.

Let $\mathfrak{B} \subseteq \{1, \dots, m\}$ be the indices j such that $\widehat{D}_j$ contributes to $\mathfrak{G}$. The case distinction tells us that $\sum_{k=1}^{q} \mathsf{s}(\widehat{H}^0_k) < \mathsf{s}(\widehat{C}) + \sum_{j\in\mathfrak{B}} \mathsf{s}(\widehat{D}_j)$ if $\widehat{C}$ contributes to $\mathfrak{G}$, and $\sum_{k=1}^{q} \mathsf{s}(\widehat{H}^0_k) < \sum_{j\in\mathfrak{B}} \mathsf{s}(\widehat{D}_j)$ if not. Further, one can also see that $\mathsf{s}(\widehat{D}_j) > 1$ $(j \in \mathfrak{B})$ and $\mathsf{s}(\widehat{C}) > 1$ (in the first case). Hence, the following inequality holds:

$$\sum_{k=1}^{q} \mathsf{s}(\widehat{H}^0_k) < \mathsf{s}(\widehat{C}) \cdot \prod_{j\in\mathfrak{B}} \mathsf{s}(\widehat{D}_j).$$

Using this inequality, the one in (43) and the construction of $\mathcal{I}$, we obtain the following:

$$|\Delta^{\mathcal{I}}| = 1 + \sum_{k=1}^{q} |\Delta^{\mathcal{I}_k}| < 1 + \mathsf{s}(\widehat{C}) \cdot \prod_{j\in\mathfrak{B}} \mathsf{s}(\widehat{D}_j) \cdot \prod_{j\in\mathfrak{A}} \mathsf{s}(\widehat{D}_j). \tag{44}$$

Last, note that for all $j \in \mathfrak{A} \cup \mathfrak{B}$, a concept Y_j was selected above. This means that $\widehat{D}_j \not\equiv \bot$. Thus, (44) implies that the size of $\Delta^{\mathcal{I}}$ is as required, i.e.,

$$|\Delta^{\mathcal{I}}| \leq \mathsf{s}(\widehat{C}) \cdot \prod\{\mathsf{s}(\widehat{D}_j) \mid 1 \leq j \leq m \wedge \widehat{D}_j \not\equiv \bot\}.$$

To conclude, let us see that $e_0 \in (\widehat{C} \sqcap \bigsqcap_{j=1}^{m} \neg\widehat{D}_j)^{\mathcal{I}}$. To see that $e_0 \in \widehat{C}^{\mathcal{I}}$, we show that $e_0 \in \widehat{C}_i^{\mathcal{I}}$ for all $1 \leq i \leq n$ by considering the possible forms of $\widehat{C}_i$.

- $\widehat{C}_i = A \in \mathsf{NC}$. By definition of $A^{\mathcal{I}}$ and the fact that $d \in \widehat{C}_i^{\mathcal{J}}$, we obtain that $e_0 \in A^{\mathcal{I}}$.
- $\widehat{C}_i = \exists r.\widehat{C}'$. This means that $\exists r.\widehat{C}' = \exists r_k.\widehat{G}_k \in \mathfrak{G}$ and $\widehat{H}^0_k = \widehat{G}_k$. Hence, $e_k \in \widehat{G}_k^{\mathcal{I}_k}$, and the tree-shape of $\mathcal{I}$ yields $e_k \in \widehat{G}_k^{\mathcal{I}}$. Thus, since $(e_0, e_k) \in r_k^{\mathcal{I}}$, we have that $e_0 \in (\exists r_k.\widehat{G}_k)^{\mathcal{I}}$.
- $\widehat{C}_i = E_{\{>,\geq\}\,t}$. Since $d \in \widehat{C}_i^{\mathcal{J}}$, an $\mathcal{EL}$ concept X_i was fixed such that $d \in X_i^{\mathcal{J}}$. Moreover, membership in X_i implies membership in $\widehat{C}_i$. Let $\mathsf{At} \in top(X_i)$. If $\mathsf{At} \in \mathsf{NC}$, like for $\widehat{C}_i = A$, we have $e_0 \in \mathsf{At}^{\mathcal{I}}$. Otherwise, $\mathsf{At} \in \mathfrak{G}$. As in the previous case, it follows that $e_0 \in \mathsf{At}^{\mathcal{I}}$. Thus, $e_0 \in X_i^{\mathcal{I}}$ and $e_0 \in \widehat{C}_i^{\mathcal{I}}$.
- $\widehat{C}_i = E_{\{<,\leq\}\,t}$. We have that $\widehat{C}_i \equiv \neg E_{i,1} \sqcap \dots \sqcap \neg E_{i,\ell_i}$ and $d \notin E_p^{\mathcal{J}}$ for all $1 \leq p \leq \ell_i$. Let $\phi : \Delta^{\mathcal{I}} \to \Delta^{\mathcal{J}}$ be the mapping $\phi(e_0) = d$, and $\phi(e) = \phi_k(e)$ if $e \in \Delta^{\mathcal{I}_k}$. Since each ϕ_k is a homomorphism from $\mathcal{I}_k$ into $\mathcal{J}$, the definition of $\mathcal{I}$ yields that ϕ is a homomorphism from $\mathcal{I}$ into $\mathcal{J}$. Hence, $\phi(e_0) = d$, $d \notin E_p^{\mathcal{J}}$ and Corollary 2.8 imply that $e_0 \notin E_p^{\mathcal{I}}$ for all $1 \leq p \leq \ell_i$. Thus, we can conclude that $e_0 \in \widehat{C}_i^{\mathcal{I}}$.

Regarding each $\widehat{D}_j$, for $\widehat{D}_j \equiv \bot$ it is clear that $e_0 \in \neg\widehat{D}_j^{\mathcal{I}}$. Otherwise, let $\widehat{D}_{j,\iota}$ be the top level atom of $\widehat{D}_j$ chosen above. It suffices to show that $e_0 \notin \widehat{D}_{j,\iota}^{\mathcal{I}}$. If $\widehat{D}_{j,\iota} \in \mathsf{NC}$ or $\widehat{D}_{j,\iota} = E_{\bowtie t}$, we can reuse the cases described above for $\widehat{C}_i$ to show that $e_0 \notin \widehat{D}_{j,\iota}^{\mathcal{I}}$. Otherwise, $\widehat{D}_{j,\iota}$ is of the form $\exists s.\widehat{F}$. Suppose that $(e_0, e_k) \in s^{\mathcal{I}}$ for some $1 \leq k \leq q$. This means that $\exists r_k.\widehat{G}_k \in \mathfrak{G}$, $s = r_k$ and $\widehat{H}^j_k = \widehat{F}$. Hence, $e_k \in (\neg\widehat{H}^j_k)^{\mathcal{I}_k}$ by induction, and $e_k \in (\neg\widehat{H}^j_k)^{\mathcal{I}}$ by construction of $\mathcal{I}$. Since e_k is an arbitrary s-successor of e_0, it then follows that $e_0 \notin (\exists s.\widehat{F})^{\mathcal{I}}$. Thus, we have shown that $e_0 \in (\widehat{C} \sqcap \bigsqcap_{j=1}^{m} \neg\widehat{D}_j)^{\mathcal{I}}$.

This completes the proof of the lemma. □

To continue, we prove Lemma 4.3 by considering its generalization to conjunctions containing more than one negated $\tau\mathcal{EL}(m)$ concept.

Lemma B.3. *Let $\sim \in$ simi-d. In addition, let $\widehat{C}$, $\widehat{D}_1, \dots, \widehat{D}_m$ be $\tau\mathcal{EL}(m_\sim)$ concept descriptions and $\mathcal{J}$ an interpretation satisfying $\widehat{G} := \widehat{C} \sqcap \neg\widehat{D}_1 \sqcap \dots \sqcap \neg\widehat{D}_m$. Then, the interpretation $\mathcal{I}$ obtained from $\mathcal{J}$ in Lemma B.2 only interprets concept names in $\mathfrak{C}(\widehat{G}, \sim) := \{B \in \mathsf{NC} \mid B \in \mathfrak{s}_{pm}(A) \wedge A \in \mathsf{sig}(\widehat{G})\}$ and role names in $\mathfrak{R}(\widehat{G}, \sim) := \{s \in \mathsf{NR} \mid s \in \mathfrak{s}_{pm}(r) \wedge r \in \mathsf{sig}(\widehat{G})\}$ as non-empty.*

Proof. Since $\mathcal{I}$ is homomorphic to $\mathcal{J}$, this implies that $\mathcal{I}$ can only give a non-empty interpretation to a symbol $X \in \mathsf{NC} \cup \mathsf{NR}$ if $\mathcal{J}$ does. Therefore, to prove our claim it suffices to show that there is such a $\mathcal{J}$ interpreting only symbols in $\mathfrak{C}(\widehat{G}, \sim)$ and $\mathfrak{R}(\widehat{G}, \sim)$ as non-empty.

The concepts $\widehat{C}$ and $\widehat{D}_j$ $(1 \leq j \leq m)$ are equivalent to $\mathcal{ALC}$ concepts C and D_j, that result from applying the translation described in Subsection 3.2 to $\widehat{C}$ and $\widehat{D}_j$. Hence, in addition to the role and concept names occurring in $\widehat{C}$ and $\widehat{D}_j$, an interpretation $\mathcal{J}$ satisfying

$C \sqcap \neg D_1 \sqcap \cdots \sqcap \neg D_m$ only needs to consider symbols appearing in disjunctions that result from translating concepts $E_{\bowtie t}$ occurring in $\widehat{C}$ and $\widehat{D}_j$, as described in Lemma 3.11. Such disjunctions are defined over the symbols in $\mathfrak{s}_\sim(E)$, where $\mathfrak{s}_\sim$ is as in Lemma 3.19. Thus, our claim holds, since $\mathsf{sig}(E) \subseteq \mathsf{sig}(\widehat{G})$ and $\mathfrak{s}_\sim(E)$ is defined in terms of $\mathfrak{s}_{pm}(X)$ for all $X \in \mathsf{sig}(E)$. $\square$

We conclude this subsection with the proof of Proposition 4.5. To this end, we use the extension to acyclic $\tau\mathcal{EL}(m)$ TBoxes of the dependency relation $\rightarrow$ from Definition 2.1, i.e., $\rightarrow$ is now defined over $\mathsf{NC}_d^{\mathfrak{T}}$ and considers all definitions in $\mathfrak{T}$.

Proposition 4.5. *Let $\mathfrak{T}$ be an acyclic $\tau\mathcal{EL}(m_\sim)$ TBox and $\alpha \in \mathsf{NC}_d^{\mathfrak{T}}$. Then, $\mathsf{rd}(u_{\mathfrak{T}}(\alpha)) \leq \mathsf{s}(\mathfrak{T})$.*

Proof. Since $|\mathsf{NC}_d^{\mathfrak{T}}| \leq \mathsf{s}(\mathfrak{T})$, it is enough to show that every $\alpha \in \mathsf{NC}_d^{\mathfrak{T}}$ satisfies the following property:

- $\mathsf{rd}(u_{\mathfrak{T}}(\alpha)) \leq |\{\beta \in \mathsf{NC}_d^{\mathfrak{T}} \mid \alpha \rightarrow^+ \beta\}|$.

The proof is by induction on the partial order induced by $\rightarrow$ over $\mathsf{NC}_d^{\mathfrak{T}}$. Let $\alpha \doteq \widehat{C}_\alpha \in \widehat{\mathcal{T}} \cup \mathcal{T}$. To see that α satisfies the property, notice first that if $\alpha \rightarrow \alpha'$ for some $\alpha' \in \mathsf{NC}_d^{\mathfrak{T}}$, then

$$|\{\beta \in \mathsf{NC}_d^{\mathfrak{T}} \mid \alpha' \rightarrow^+ \beta\}| < |\{\beta \in \mathsf{NC}_d^{\mathfrak{T}} \mid \alpha \rightarrow^+ \beta\}|. \tag{45}$$

This follows from the facts that $\alpha' \rightarrow^+ \beta$ implies $\alpha \rightarrow^+ \beta$, and that $\alpha' \rightarrow^+ \alpha'$ does not hold since $\mathfrak{T}$ is acyclic. Second, recall that $\mathsf{rd}(u_{\mathfrak{T}}(\alpha)) = \max\{\mathsf{rd}(u_{\mathfrak{T}}(\mathsf{At})) \mid \mathsf{At} \in top(\widehat{C}_\alpha)\}$. Hence, it suffices to show that $\mathsf{rd}(u_{\mathfrak{T}}(\mathsf{At})) \leq |\{\beta \in \mathsf{NC}_d^{\mathfrak{T}} \mid \alpha \rightarrow^+ \beta\}|$ for all $\mathsf{At} \in top(\widehat{C}_\alpha)$. We distinguish the following cases:

- $\mathsf{At} \in \mathsf{NC}_{pr}^{\mathfrak{T}}$. This means that $\mathsf{At} \in \mathsf{NC} \cup \mathsf{ND}$ and $u_{\mathfrak{T}}(\mathsf{At}) = \mathsf{At}$. Hence, $\mathsf{rd}(u_{\mathfrak{T}}(\mathsf{At})) = \mathsf{rd}(\mathsf{At}) = 0$.
- $\mathsf{At} = E_{\bowtie t}$ with $E \in \mathsf{NC}_d^{\mathcal{T}}$. By definition, $\mathsf{rd}(u_{\mathfrak{T}}(E_{\bowtie t})) = \mathsf{rd}(u_{\mathfrak{T}}(E))$. Since $\alpha \rightarrow E$, by (45) and the application of induction to E we obtain $\mathsf{rd}(u_{\mathfrak{T}}(E)) < |\{\beta \in \mathsf{NC}_d^{\mathfrak{T}} \mid \alpha \rightarrow^+ \beta\}|$.
- $\mathsf{At} = \exists r.\gamma$ with $\gamma \in \mathsf{NC}_d^{\mathfrak{T}}$. We know that $\mathsf{rd}(u_{\mathfrak{T}}(\exists r.\gamma)) = \mathsf{rd}(u_{\mathfrak{T}}(\gamma)) + 1$. Similarly as above, $\alpha \rightarrow \gamma$, (45) and the application of induction to γ yield $\mathsf{rd}(u_{\mathfrak{T}}(\gamma)) + 1 \leq |\{\beta \in \mathsf{NC}_d^{\mathfrak{T}} \mid \alpha \rightarrow^+ \beta\}|$. $\square$

B.2. Assertional reasoning

We now provide the proofs of Lemmas 4.16 and 4.18. We start with the following auxiliary result.

Lemma B.4. *Let $\sim\, \in$ simi-d, $\mathcal{A}$ a consistent $\tau\mathcal{EL}(m)$ ABox, $\mathcal{J}$ a model of $\mathcal{A}$, and $\mathfrak{F}$ a set of $\tau\mathcal{EL}(m)$ concepts closed under sub-descriptions. Define, for all $b \in \mathsf{Ind}(\mathcal{A})$, the set $\mathfrak{F}_b := \{\widehat{F} \in \mathfrak{F} \mid b^{\mathcal{J}} \notin \widehat{F}^{\mathcal{J}}\}$. Then, there is a tree-like model $\mathcal{I}$ of $\mathcal{A}$ such that:*

- *$\mathcal{I}$ only interprets concept names in $\mathfrak{C}(\mathcal{A}, \mathfrak{F}, \sim) := \bigcup\{\mathfrak{s}_{pm}(A) \mid A$ occurs in $\mathcal{A}$ or in a concept in $\mathfrak{F}\}$ and role names in $\mathfrak{R}(\mathcal{A}, \mathfrak{F}, \sim) := \bigcup\{\mathfrak{s}_{pm}(r) \mid r$ occurs in $\mathcal{A}$ or in a concept in $\mathfrak{F}\}$ as non-empty.*
- *For all $b \in \mathsf{Ind}(\mathcal{A})$: $\mathcal{I}_b$ has depth at most $\mathsf{rd}(\widehat{C}_b \sqcap \prod_{\widehat{F} \in \mathfrak{F}_b} \widehat{F})$, $|\Delta^{\mathcal{I}_b}| \leq \mathsf{s}(\widehat{C}_b) \cdot \prod\{\mathsf{s}(\widehat{F}) \mid \widehat{F} \in \mathfrak{F}_b \wedge \widehat{F} \not\equiv \bot\}$, and the following holds for all $\widehat{F} \in \mathfrak{F}$:*

$$\widehat{F} \in \mathfrak{F}_b \text{ implies } b^{\mathcal{I}} \notin \widehat{F}^{\mathcal{I}}. \tag{46}$$

Proof. Since $b^{\mathcal{J}} \in \widehat{C}_b^{\mathcal{J}}$ and $b^{\mathcal{J}} \notin \widehat{F}^{\mathcal{J}}$ for all $\widehat{F} \in \mathfrak{F}_b$, the application of Lemma B.2 yields a finite and tree-shaped interpretation $\mathcal{I}_b$ with root element $d_b \in (\widehat{C}_b \sqcap \prod_{\widehat{F} \in \mathfrak{F}_b} \neg\widehat{F})^{\mathcal{I}_b}$ such that:

- there is a homomorphism $\phi_b : \mathcal{I}_b \rightarrow \mathcal{J}$ with $\phi_b(d_b) = b^{\mathcal{J}}$,
- $\mathcal{I}_b$ has depth at most $\mathsf{rd}(\widehat{C}_b \sqcap \prod_{\widehat{F} \in \mathfrak{F}_b} \widehat{F})$, and
- $|\Delta^{\mathcal{I}_b}| \leq \mathsf{s}(\widehat{C}_b) \cdot \prod\{\mathsf{s}(\widehat{F}) \mid \widehat{F} \in \mathfrak{F}_b \wedge \widehat{F} \not\equiv \bot\}$.

Define an interpretation $\mathcal{I}$ as the disjoint union of all interpretations $\mathcal{I}_b$, where $\cdot^{\mathcal{I}}$ extends the union of all $\cdot^{\mathcal{I}_b}$ by augmenting $r^{\mathcal{I}}$ with $\{(d_b, d_c) \mid r(b, c) \in \mathcal{A}\}$ and defining $b^{\mathcal{I}} := d_b$. Clearly, $\mathcal{I}$ is finite, tree-like, and $|\Delta^{\mathcal{I}_b}|$ and the depth of $\mathcal{I}_b$ are as claimed. Moreover, since $\widehat{C}_b = \prod_{\widehat{C}(b) \in \mathcal{A}} \widehat{C}$, the application of Lemma B.3 yields that $\mathcal{I}$ only interprets symbols in $\mathfrak{C}(\mathcal{A}, \mathfrak{F}, \sim) \cup \mathfrak{R}(\mathcal{A}, \mathfrak{F}, \sim)$ as non-empty.

We next show that $\mathcal{I}$ is a model of $\mathcal{A}$ and satisfies (46). To this end, we use the following property, which is a consequence of the tree shape of each $\mathcal{I}_b$:

$$\text{For all } \tau\mathcal{EL}(m) \text{ concepts } \widehat{C} \text{ and } d \in \Delta^{\mathcal{I}_b} \setminus \{d_b\}\text{: } d \in \widehat{C}^{\mathcal{I}_b} \text{ iff } d \in \widehat{C}^{\mathcal{I}}. \tag{47}$$

Let us start with $\mathcal{A}$. Clearly, $\mathcal{I}$ satisfies all role assertions in $\mathcal{A}$. To see that it also satisfies all $\widehat{C}(b) \in \mathcal{A}$, we distinguish the possible forms of $\widehat{C}$:

- $\widehat{C} = A \in \mathsf{NC}$. Since $d_b \in A^{\mathcal{I}_b}$, we obtain $b^{\mathcal{I}} \in A^{\mathcal{I}}$ as a direct consequence of the definition of $\mathcal{I}$.
- $\widehat{C} = \exists r.\widehat{H}$. Since $d_b \in \widehat{C}^{\mathcal{I}_b}$, there is $e \in \Delta^{\mathcal{I}_b} \setminus \{d_b\}$ such that $(d_b, e) \in r^{\mathcal{I}_b}$ and $e \in \widehat{H}^{\mathcal{I}_b}$. Hence, the application of (47) yields $e \in \widehat{H}^{\mathcal{I}}$, and consequently, $b^{\mathcal{I}} \in (\exists r.\widehat{H})^{\mathcal{I}}$.
- $\widehat{C} = H_{\bowtie t}$. Let $\bowtie \in \{>, \geq\}$. The semantics of $H_{\bowtie t}$ implies that $d_b \in (H_{\bowtie t})^{\mathcal{I}_b}$ iff there is an $\mathcal{EL}$ concept G such that $d_b \in G^{\mathcal{I}_b}$ and $H \sim G \bowtie t$. Since $d_b \in (H_{\bowtie t})^{\mathcal{I}_b}$ actually holds and G is an $\mathcal{EL}$ concept, the construction of $\mathcal{I}$ implies that $b^{\mathcal{I}} \in G^{\mathcal{I}}$. Hence, $b^{\mathcal{I}} \in (H_{\bowtie t})^{\mathcal{I}}$. As for $\bowtie \in \{<, \leq\}$, Lemma 3.11 tells us that $H_{\bowtie t} \equiv \neg H_1 \sqcap \cdots \sqcap \neg H_n$, where each H_i is an $\mathcal{EL}$ concept. Therefore, $b^{\mathcal{J}} \in (H_{\bowtie t})^{\mathcal{J}}$ implies $b^{\mathcal{J}} \notin H_i^{\mathcal{J}}$ for all $1 \leq i \leq n$. Further, it is easy to see that all homomorphisms $\phi_c : \mathcal{I}_c \to \mathcal{J}$ can be combined into a homomorphism $\phi : \mathcal{I} \to \mathcal{J}$ such that $\phi(c^{\mathcal{I}}) = c^{\mathcal{J}}$ for all $c \in \mathsf{Ind}(\mathcal{A})$. Hence, Corollary 2.8 yields that $b^{\mathcal{I}} \notin H_i^{\mathcal{I}}$ for all $1 \leq i \leq n$. Thus, $b^{\mathcal{I}} \in (H_{\bowtie t})^{\mathcal{I}}$.
- $\widehat{C}$ is a conjunction. This is an easy consequence of the previous cases and the semantics of $\sqcap$.

It remains to show that (46) holds. We use induction on the structure of concepts in $\mathfrak{F}$. Let $b \in \mathsf{Ind}(\mathcal{A})$ and $\widehat{F} \in \mathfrak{F}$ such that $\widehat{F} \in \mathfrak{F}_b$. This means that $b^{\mathcal{J}} \notin \widehat{F}^{\mathcal{J}}$. We distinguish the following cases:

- $\widehat{F} = \exists r.\widehat{H}$. Suppose $(b^{\mathcal{I}}, e) \in r^{\mathcal{I}}$ for some $e \in \Delta^{\mathcal{I}}$. If $e \in \Delta^{\mathcal{I}_b} \setminus \{d_b\}$, then $d_b \notin (\exists r.\widehat{H})^{\mathcal{I}_b}$ implies $e \notin \widehat{H}^{\mathcal{I}_b}$. The application of (47) yields $e \notin \widehat{H}^{\mathcal{I}}$. Otherwise, $e = c^{\mathcal{I}}$ for some $c \in \mathsf{Ind}(\mathcal{A})$ and $r(b, c) \in \mathcal{A}$. Since $\mathcal{J} \models \mathcal{A}$ and $b^{\mathcal{J}} \notin \exists r.\widehat{H}^{\mathcal{J}}$, we have $c^{\mathcal{J}} \notin \widehat{H}^{\mathcal{J}}$. Further, $\widehat{H} \in \mathfrak{F}$ because $\mathfrak{F}$ is closed under sub-descriptions. Hence, $\widehat{H} \in \mathfrak{F}_c$, and the application of induction yields $c^{\mathcal{I}} \notin \widehat{H}^{\mathcal{I}}$. Thus, $b^{\mathcal{I}} \notin (\exists r.\widehat{H})^{\mathcal{I}}$ because no r-successor of $b^{\mathcal{I}}$ is in $\widehat{H}^{\mathcal{I}}$.
- $\widehat{F} = A \in \mathsf{NC}$ or $\widehat{F} = H_{\bowtie t}$ or $\widehat{F}$ is a conjunction. Since $b^{\mathcal{J}} \notin \widehat{F}^{\mathcal{J}}$, this means that $d_b \notin \widehat{F}^{\mathcal{I}_b}$ by construction of $\mathcal{I}_b$. If $\widehat{F} = A$ then $d_b \notin A^{\mathcal{I}_b}$ and the definition of $\mathcal{I}$ yield $b^{\mathcal{I}} \notin A^{\mathcal{I}}$. The case where $\widehat{F} = H_{\bowtie t}$ can be shown as in the previous case distinction, since $\neg H_{\bowtie t} \equiv H_{\bowtie' t}$. The case where $\widehat{F}$ is a conjunction follows as a direct consequence of the semantics of $\sqcap$ and the previous cases. $\square$

Lemma 4.16. *Let $\sim \in$ simi-d, $\widehat{D}$ a $\tau\mathcal{EL}(m)$ concept description, $\mathcal{A}$ a $\tau\mathcal{EL}(m)$ ABox, and $a \in \mathsf{Ind}(\mathcal{A})$. If the ABox $\mathcal{A} \cup \{\neg\widehat{D}(a)\}$ is consistent in $\tau\mathcal{EL}(m_\sim)$, then it has a tree-like model $\mathcal{I}$ such that*

- $|\Delta^{\mathcal{I}_b}| \leq \mathsf{s}(\widehat{C}_b) \cdot \mathsf{s}(\widehat{D})^u$ *for all* $b \in \mathsf{Ind}(\mathcal{A})$, *where* $u = |\mathsf{sub}(\widehat{D})|$,
- $\mathcal{I}$ *only interprets concept names in* $\mathfrak{C}(\mathcal{A}, \widehat{D}, \sim) := \{B \in \mathsf{NC} \mid B \in \mathfrak{s}_{pm}(A) \wedge A \in \mathsf{sig}(\mathcal{A}) \cup \mathsf{sig}(\widehat{D})\}$ *and role names in* $\mathfrak{R}(\mathcal{A}, \widehat{D}, \sim) := \{s \in \mathsf{NR} \mid s \in \mathfrak{s}_{pm}(r) \wedge r \in \mathsf{sig}(\mathcal{A}) \cup \mathsf{sig}(\widehat{D})\}$ *as non-empty.*

Proof. Let $\mathcal{J}$ be a model of $\mathcal{A} \cup \{\neg\widehat{D}(a)\}$. The application of Lemma B.4 w.r.t. $\mathfrak{F} = \mathsf{sub}(\widehat{D})$ yields a tree-like model $\mathcal{I}$ of $\mathcal{A}$ satisfying (46) w.r.t. $\mathsf{sub}(\widehat{D})$. Hence, $a^{\mathcal{I}} \notin \widehat{D}^{\mathcal{I}}$, since $a^{\mathcal{J}} \notin \widehat{D}^{\mathcal{J}}$ and $\widehat{D} \in \mathsf{sub}(\widehat{D})$. Further, since $\mathfrak{F} = \mathsf{sub}(\widehat{D})$, Lemma B.4 ensures that $|\Delta^{\mathcal{I}_b}|$ is as claimed, and that $\mathcal{I}$ only interprets symbols in $\mathfrak{C}(\mathcal{A}, \widehat{D}, \sim)$ and $\mathfrak{R}(\mathcal{A}, \widehat{D}, \sim)$ as non-empty. $\square$

Lemma 4.18. *Let $\sim \in$ simi-d, $(\mathfrak{T}, \mathcal{A})$ an acyclic $\tau\mathcal{EL}(m)$ KB, $\alpha \in \mathsf{NC}_d^{\mathfrak{T}}$ and $a \in \mathsf{Ind}(\mathcal{A})$. In $\tau\mathcal{EL}(m_\sim)$, if $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is consistent, then it has a finite tree-like model $\mathcal{I}$ such that for all $b \in \mathsf{Ind}(\mathcal{A})$:*

1. *$\mathcal{I}_b$ is a model of $\mathfrak{T}$, has depth at most $\mathsf{s}(\mathfrak{T})$, and it only interprets symbols in $\mathfrak{C}(\mathfrak{T}, \sim)$, $\mathfrak{R}(\mathfrak{T}, \sim)$ and $\mathsf{NC}_d^{\mathfrak{T}}$ as non-empty.*
2. *For all $X \in \mathsf{cl}(\mathfrak{T}, \sim)$: $b^{\mathcal{I}} \in X^{\mathcal{I}}$ iff $b^{\mathcal{I}} \in X^{\mathcal{I}_b}$.*

Proof. Let $\mathcal{J}$ be a model of $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$. Consider the following ABox:

$$\mathcal{A}^* := \{u_{\mathfrak{T}}(X)(b) \mid b \in \mathsf{Ind}(\mathcal{A}) \wedge b^{\mathcal{J}} \in X^{\mathcal{J}} \wedge X \in \mathsf{sub}(\mathfrak{T}) \cup \mathfrak{C}(\mathfrak{T}, \sim)\}.$$

By Lemma 2.28, we know that $\beta^{\mathcal{I}} = u_{\mathfrak{T}}(\beta)^{\mathcal{I}}$ for all $\beta \in \mathsf{NC}_d^{\mathfrak{T}}$ and all models $\mathcal{I}$ of $\mathfrak{T}$. Hence, $\mathcal{J}$ is a model of $\mathcal{A}^*$. Further, since $\mathcal{J} \models \mathcal{A} \cup \{\neg\alpha(a)\}$, we have:

$$X(b) \in \mathcal{A} \text{ implies } u_{\mathfrak{T}}(X)(b) \in \mathcal{A}^* \quad \text{and} \quad a^{\mathcal{J}} \notin u_{\mathfrak{T}}(\alpha)^{\mathcal{J}}. \tag{48}$$

Define $\mathfrak{F} := \{u_{\mathfrak{T}}(X) \mid X \in \mathsf{sub}(\mathfrak{T}) \cup \mathfrak{C}(\mathfrak{T}, \sim)\}$. The application of Lemma B.4 to $\mathcal{A}^*$ w.r.t. $\mathfrak{F}$ yields a finite tree-like model $\mathcal{I}^*$ of $\mathcal{A}^*$. Notice that all concepts in $\mathfrak{F}$ and $\mathcal{A}^*$ are of the form $u_{\mathfrak{T}}(X)$ for some $X \in \mathsf{sub}(\mathfrak{T}) \cup \mathfrak{C}(\mathfrak{T}, \sim)$. As a consequence, we have that:

- $\mathsf{rd}(u_{\mathfrak{T}}(X)) \leq \mathsf{s}(\mathfrak{T})$ due to Proposition 4.5, and all concept and role names occurring in $\mathcal{A}^*$ or in a concept in $\mathfrak{F}$ belong to $\mathsf{sig}(\mathfrak{T})$.

Hence, Lemma B.4 also tells us that $\mathcal{I}_b^*$ has depth at most $\mathsf{s}(\mathfrak{T})$ for all $b \in \mathsf{Ind}(\mathcal{A})$, and $\mathcal{I}^*$ only interprets symbols in $\mathfrak{C}(\mathfrak{T}, \sim)$ and $\mathfrak{R}(\mathfrak{T}, \sim)$ as non-empty. Furthermore, we can show that given $b \in \mathsf{Ind}(\mathcal{A})$, the following claim holds for all $X \in \mathsf{sub}(\mathfrak{T}) \cup \mathfrak{C}(\mathfrak{T}, \sim)$:

$$b^{\mathcal{I}^*} \in u_{\mathfrak{T}}(X)^{\mathcal{I}^*} \text{ iff } b^{\mathcal{I}^*} \in u_{\mathfrak{T}}(X)^{\mathcal{I}_b^*} \tag{49}$$

The proof goes as follows. Assume that $b^{\mathcal{I}^*} \in u_{\mathfrak{T}}(X)^{\mathcal{I}^*}$. The contrapositive of (46) yields $b^{\mathcal{J}} \in u_{\mathfrak{T}}(X)^{\mathcal{J}}$, and consequently $b^{\mathcal{J}} \in X^{\mathcal{J}}$. The latter implies that $u_{\mathfrak{T}}(X)(b) \in \mathcal{A}^*$. Thus, the construction of $\mathcal{I}_b^*$ in Lemma B.4 yields that $b^{\mathcal{I}^*} \in u_{\mathfrak{T}}(X)^{\mathcal{I}_b^*}$. Conversely, suppose that $b^{\mathcal{I}^*} \in u_{\mathfrak{T}}(X)^{\mathcal{I}_b^*}$. It must be that $b^{\mathcal{J}} \in u_{\mathfrak{T}}(X)^{\mathcal{J}}$, for otherwise $b^{\mathcal{J}} \notin u_{\mathfrak{T}}(X)^{\mathcal{J}}$ implies $u_{\mathfrak{T}}(X) \in \mathfrak{F}_b$, and the latter yields $b^{\mathcal{I}^*} \notin u_{\mathfrak{T}}(X)^{\mathcal{I}_b^*}$ by construction of $\mathcal{I}_b^*$ in Lemma B.4. Hence, $b^{\mathcal{J}} \in X^{\mathcal{J}}$ and $u_{\mathfrak{T}}(X)(b) \in \mathcal{A}^*$. Thus, $b^{\mathcal{I}^*} \in u_{\mathfrak{T}}(X)^{\mathcal{I}^*}$.

$$
\begin{array}{llcl}
R_{\exists}: & E \doteq \exists r.C \sqcap \tilde{F} \text{ with } C \notin \mathsf{NC}_d^{\mathcal{T}} & \longrightarrow & \{E \doteq \exists r.E_C \sqcap \tilde{F}, E_C \doteq C\} \\
R_E: & E \doteq F \sqcap \tilde{F} \text{ with } F \in \mathsf{NC}_d^{\mathcal{T}} & \longrightarrow & \{E \doteq C_F \sqcap \tilde{F}\} \\
R_{\bowtie}: & \alpha \doteq C_{\bowtie t} \sqcap \tilde{F} & \longrightarrow & \{\alpha \doteq (E_C)_{\bowtie t} \sqcap \tilde{F}, E_C \doteq C\}
\end{array}
$$

Fig. 7. Normalization rules.

Now, by Proposition 2.29, $\mathcal{I}^*$ can be uniquely extended into a model $\mathcal{I}$ of $\mathfrak{T}$. Since $\mathcal{I}^* \models \mathcal{A}^*$ and satisfies (46) in Lemma B.4, we obtain from (48) that $b^{\mathcal{I}} \in u_{\mathfrak{T}}(X)^{\mathcal{I}}$ for all $X(b) \in \mathcal{A}$ and $a^{\mathcal{I}} \notin u_{\mathfrak{T}}(\alpha)^{\mathcal{I}}$. Thus, the application of Lemma 2.28 yields that $\mathcal{I}$ is a model of $(\mathfrak{T}, \mathcal{A} \cup \{\neg\alpha(a)\})$. Notice that $\mathcal{I}$ is obtained from $\mathcal{I}^*$ by appropriately interpreting the concept names in $\mathsf{NC}_d^{\mathfrak{T}}$. Therefore, $\mathcal{I}$ only interprets symbols in $\mathfrak{C}(\mathfrak{T},\sim)$, $\mathfrak{R}(\mathfrak{T},\sim)$ and $\mathsf{NC}_d^{\mathfrak{T}}$ as non-empty, and the interpretations $\mathcal{I}_b$ have depth at most $\mathsf{s}(\mathfrak{T})$. Hence, it remains to show that all interpretations $\mathcal{I}_b$ are models of $\mathfrak{T}$ and satisfy 2. This can be shown in three steps.

1. Since $\mathcal{I} \models \mathfrak{T}$, the application of Lemma 2.28 can be used to extend (49) into:

   $$b^{\mathcal{I}} \in X^{\mathcal{I}} \text{ iff } b^{\mathcal{I}} \in u_{\mathfrak{T}}(X)^{\mathcal{I}} \text{ iff } b^{\mathcal{I}^*} \in u_{\mathfrak{T}}(X)^{\mathcal{I}^*} \text{ iff } b^{\mathcal{I}^*} \in u_{\mathfrak{T}}(X)^{\mathcal{I}^*_b}.$$

   Further, we know that $\cdot^{\mathcal{I}}$ and $\cdot^{\mathcal{I}_b}$ coincide on $\mathsf{NC} \cup \mathsf{ND}$ w.r.t. $\Delta^{\mathcal{I}_b}$, and on NR w.r.t. $\Delta^{\mathcal{I}_b} \setminus \{b^{\mathcal{I}}\}$. Hence, using the previous equivalences and the normal form of $\mathfrak{T}$, it is not hard to prove that for all $\mathcal{EL}$ concepts $X \in \mathsf{sub}(\mathfrak{T}) \cup \mathfrak{C}(\mathfrak{T},\sim)$: $b^{\mathcal{I}} \in X^{\mathcal{I}}$ iff $b^{\mathcal{I}} \in X^{\mathcal{I}_b}$. This ensures that $\mathcal{I}_b$ is a model of $\mathcal{T}$.
2. Once we know that $\mathcal{I}_b \models \mathcal{T}$, (49) can be used to show that $b^{\mathcal{I}} \in (E_{\bowtie t})^{\mathcal{I}}$ iff $b^{\mathcal{I}} \in (E_{\bowtie t})^{\mathcal{I}_b}$ for all $E_{\bowtie t} \in \mathsf{sub}(\mathfrak{T})$. Afterwards, the argument used in the previous step can be applied to obtain that $b^{\mathcal{I}} \in X^{\mathcal{I}}$ iff $b^{\mathcal{I}} \in X^{\mathcal{I}_b}$, for all $X \in \mathsf{sub}(\mathfrak{T}) \cup \mathfrak{C}(\mathfrak{T},\sim)$. Thus, it follows that $\mathcal{I} \models \widehat{\mathcal{T}}$ and $\mathcal{I} \models \mathfrak{T}$.
3. Since $\mathsf{cl}(\mathfrak{T},\sim)$ is the closure under negation of $\mathsf{sub}(\mathfrak{T}) \cup \mathfrak{C}(\mathfrak{T},\sim)$, the equivalence obtained in the previous step yields that 2 is fulfilled.

This completes the proof of the lemma. □

B.3. Normalization

We now introduce *normalized* $\tau\mathcal{EL}(m)$ TBoxes in *reduced* form, and show that one can (without loss of generality) restrict the attention to this kind of TBoxes.[15] Let us start by recalling the normal form introduced in [56] for $\mathcal{EL}$ TBoxes. An $\mathcal{EL}$ TBox $\mathcal{T}$ is said to be normalized iff $E \doteq C_E \in \mathcal{T}$ implies that C_E is of the form

$$P_1 \sqcap \cdots \sqcap P_k \sqcap \exists r_1.F_1 \sqcap \cdots \sqcap \exists r_n.F_n,$$

where $k, n \geq 0$, $P_1, \ldots, P_k \in \mathsf{NC}_{pr}^{\mathcal{T}}$ and $F_1, \ldots, F_n \in \mathsf{NC}_d^{\mathcal{T}}$. We extend this to $\tau\mathcal{EL}(m)$ and say that a $\tau\mathcal{EL}(m)$ TBox $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$ is normalized iff $\mathcal{T}$ is normalized and $\alpha \doteq \widehat{C}_\alpha \in \widehat{\mathcal{T}}$ implies that $\widehat{C}_\alpha$ is of the form

$$\widehat{P}_1 \sqcap \cdots \sqcap \widehat{P}_k \sqcap \exists r_1.\beta_1 \sqcap \cdots \sqcap \exists r_n.\beta_n,$$

where $k, n \geq 0$, for all $1 \leq i \leq k$ either $\widehat{P}_i \in \mathsf{NC}_{pr}^{\mathfrak{T}}$ or $\widehat{P}_i = E_{\bowtie t}$ with $E \in \mathsf{NC}_d^{\mathcal{T}}$, and $\beta_1, \ldots, \beta_n \in \mathsf{NC}_d^{\mathfrak{T}}$.

The normalization procedure for $\mathcal{EL}$ consists in applying the rules $R_\exists$ and R_E from Fig. 7. To illustrate the process we start with a simpler version of Example 12 in [56].

Example B.5. Let $\mathcal{T}$ be the $\mathcal{EL}$ TBox consisting of the following definitions:

$$E_1 \doteq P_1 \sqcap E_2 \sqcap \exists r_1.\exists r_2.E_3$$

$$E_2 \doteq P_2 \sqcap E_3 \sqcap \exists s.(E_3 \sqcap P_3)$$

$$E_3 \doteq P_4$$

The exhaustive application of rule $R_\exists$ yields a new TBox $\mathcal{T}'$:

$$E_1 \doteq P_1 \sqcap E_2 \sqcap \exists r_1.E_1' \qquad E_1' \doteq \exists r_2.E_3$$

$$E_2 \doteq P_2 \sqcap E_3 \sqcap \exists s.E_2' \qquad E_2' \doteq E_3 \sqcap P_3$$

[15] This is the case regardless of the chosen graded membership function.

$E_3 \doteq P_4$

Each newly introduced definition $E' \doteq C_{E'}$ is such that E' is a *fresh* concept name from ND. Nevertheless, $\mathcal{T}'$ is not yet normalized since E_1, E_2 and E_2' contain top-level atoms which are defined concepts. To deal with this, one can substitute these occurrences of defined concepts by their definitions, i.e., applying rule R_E, where $F \doteq C_F \in \mathcal{T}$. The correctnesses of it follows from the discussion presented in [56] for the more general case of cyclic $\mathcal{EL}$ TBoxes. Continuing with the example we obtain the normalized TBox:

$$E_1 \doteq P_1 \sqcap P_2 \sqcap P_4 \sqcap \exists s.E_2' \sqcap \exists r_1.E_1' \qquad E_1' \doteq \exists r_2.E_3$$

$$E_2 \doteq P_2 \sqcap P_4 \sqcap \exists s.E_2' \qquad E_2' \doteq P_4 \sqcap P_3$$

$$E_3 \doteq P_4$$

To have a polynomial time procedure, R_E should be applied following the partial order induced by the relation $\rightarrow^+$ introduced in Definition 2.1. More precisely, R_E can be applied to a concept definition $E \doteq C_E$ only if it has already been applied to all $F \in \mathsf{NC}_d^{\mathcal{T}}$ such that $E \rightarrow^+ F$. △

Overall, each application of $R_\exists$ replaces a top-level atom of the form $\exists r.C$ with a new atom $\exists r.E_C$, and introduces a simpler definition $E_C \doteq C$. Concerning R_E, such an ordered sequence of rule applications will always terminate since we are dealing with acyclic TBoxes. Moreover, *idempotency* of $\sqcap$ can be exploited to avoid duplications. This ensures that the described normalization procedure runs in polynomial time and produces a TBox $\mathcal{T}'$ of size polynomial in the size of $\mathcal{T}$.

This procedure can be easily adapted to normalize acyclic $\tau\mathcal{EL}(m)$ TBoxes: the rules $R_\exists$ and R_E can be applied to $\widehat{\mathcal{T}}$ in the same way, with the difference that applications of R_E need to consider that $F \doteq C_F$ may occur in $\mathcal{T}$. In addition, it needs to be ensured that threshold concepts $C_{\bowtie t}$ occurring in $\mathfrak{T}$ are such that C is a defined concept in $\mathcal{T}$. For example, consider the following definition in $\widehat{\mathcal{T}}$:

$$\alpha \doteq P_1 \sqcap \exists r_1.[(P_2 \sqcap \exists r_2.P_3)_{\leq .8}]. \tag{50}$$

To treat $(P_2 \sqcap \exists r_2.P_3)_{\leq .8}$, we introduce the rule $R_\bowtie$ whose application replaces a threshold concept $C_{\bowtie t}$ by $(E_C)_{\bowtie t}$ and adds a new definition $E_C \doteq C$ to $\mathcal{T}$, where E_C is a fresh concept name from ND. Then, to normalize (50), one first applies $R_\exists$ to obtain $\alpha \doteq P_1 \sqcap \exists r_1.\beta$ and a new definition $\beta \doteq (P_2 \sqcap \exists r_2.P_3)_{\leq .8}$ in $\widehat{\mathcal{T}}$. Second, $R_\bowtie$ is applied to simplify β into $\beta \doteq F_{\leq .8}$ and add $F \doteq P_2 \sqcap \exists r_2.P_3$ to $\mathcal{T}$.

To summarize, we define the normalization procedure for acyclic $\tau\mathcal{EL}(m)$ TBoxes as the execution of the following steps.

1. Apply the rules $R_\exists$ and $R_\bowtie$ exhaustively to $\widehat{\mathcal{T}}$.
2. Normalize the augmented $\mathcal{EL}$ TBox $\mathcal{T}$.
3. Apply the rule R_E exhaustively to $\widehat{\mathcal{T}}$.

Notice that applications of $R_\exists$ and R_E to $\widehat{\mathcal{T}}$ modify only $\widehat{\mathcal{T}}$, do not introduce threshold concepts and do not modify existing ones. In addition, since a threshold concept $C_{\bowtie t}$ occurring in $\widehat{\mathcal{T}}$ satisfies that $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{NC} \cup \mathsf{NC}_d^{\mathcal{T}}, \mathsf{NR})$, the definition $E_C \doteq C$ introduced into $\mathcal{T}$ by applying $R_\bowtie$ satisfies condition 1 in Definition 2.20. Finally, the normalization of $\mathcal{T}$ transforms only $\mathcal{T}$ and introduces only concept names from ND. Therefore, it is not hard to see that $\mathfrak{T}'$ is acyclic and satisfies the restrictions required for acyclic $\tau\mathcal{EL}(m)$ TBoxes in Definition 2.27. Regarding the running time of the procedure, $R_\bowtie$ is applied at most once to each threshold concept $C_{\bowtie t}$ in the initial TBox $\widehat{\mathcal{T}}$. This means that at most polynomially many new definitions of the form $E_C \doteq C$ are added to $\mathcal{T}$. Thus, using the same arguments as for the $\mathcal{EL}$ setting, the extended normalization procedure runs in polynomial time and yields a normalized acyclic $\tau\mathcal{EL}(m)$ TBox $\mathfrak{T}'$ of size polynomial in the size of $\mathfrak{T}$.

We next show that reasoning w.r.t. an acyclic $\tau\mathcal{EL}(m)$ TBox $\mathfrak{T}$ can be reduced to reasoning w.r.t. its normal form. As mentioned in Subsection 4.2, one can restrict the attention to satisfiability of concepts of the form $\alpha_1 \sqcap \neg\alpha_2$ such that $\alpha_1, \alpha_2 \in \mathsf{NC}_d^{\mathfrak{T}}$, and this can in turn be reduced to satisfiability of $u_{\mathfrak{T}}(\alpha_1) \sqcap \neg u_{\mathfrak{T}}(\alpha_2)$ w.r.t. $\mathfrak{T} = \emptyset$. Hence, the following proposition implies our claim.

Proposition B.6. *Let $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$ be an acyclic $\tau\mathcal{EL}(m)$ TBox and $\mathfrak{T}'$ the normal form of $\mathfrak{T}$. Then, for all $\alpha \in \mathsf{NC}_d^{\mathfrak{T}}$ we have that $u_{\mathfrak{T}}(\alpha) = u_{\mathfrak{T}'}(\alpha)$.*

To prove this proposition, it suffices to show that unfolding of a defined concept in $\mathfrak{T}$ is preserved under any rule application. The following lemma proves this property.

Lemma B.7. *Let $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$ be an acyclic $\tau\mathcal{EL}(m)$ TBox and $\mathfrak{T}' = (\widehat{\mathcal{T}}', \mathcal{T}')$ the TBox that results from a single application of a normalization rule. Then, for all $\alpha \in \mathsf{NC}_d^{\mathfrak{T}}$ it holds that $u_{\mathfrak{T}}(\alpha) = u_{\mathfrak{T}'}(\alpha)$.*

Proof. Let $\beta \doteq \widehat{C}_\beta \in \widehat{\mathcal{T}} \cup \mathcal{T}$ be the definition to which a normalization rule is applied to. We use induction on the partial order induced by $\rightarrow^+$ on $\mathsf{NC}_d^{\mathfrak{T}}$. For all $\alpha \in \mathsf{NC}_d^{\mathfrak{T}}$ we distinguish two cases:

- $\alpha \neq \beta$. This means that $\alpha \doteq \widehat{C}_\alpha \in \widehat{\mathcal{T}}' \cup \mathcal{T}'$. By definition of unfolding, $u_{\mathfrak{T}}(\alpha) \neq u_{\mathfrak{T}'}(\alpha)$ can only happen if $u_{\mathfrak{T}}(\alpha') \neq u_{\mathfrak{T}'}(\alpha')$ for some symbol $\alpha' \in \mathsf{NC}_d^{\mathfrak{T}}$ occurring in $\widehat{C}_\alpha$. However, $\alpha \to^+ \alpha'$ and the application of the induction hypothesis to α' imply that this is never the case. Hence, $u_{\mathfrak{T}}(\alpha) = u_{\mathfrak{T}'}(\alpha)$.
- $\alpha = \beta$. Let $\widehat{C}_\beta = \widehat{C}_1 \sqcap \ldots \sqcap \widehat{C}_n$. We analyze the outcome of applying each possible rule to $\beta \doteq \widehat{C}_\beta$.

  – $R_{\bowtie}$: the rule is applied to a conjunct $\widehat{C}_i = C_{\bowtie t}$ such that $C \notin \mathsf{NC}_d^{\mathcal{T}}$. Its application replaces $\widehat{C}_i$ with $(E_C)_{\bowtie t}$ in $\widehat{C}_\beta$, and adds $E_C \doteq C$ to $\mathcal{T}'$ where $E_C \in \mathsf{ND}$ is a fresh concept name. By definition of unfolding we have:

  $$u_{\mathfrak{T}}(\beta) = \prod_{j=1}^{i-1} u_{\mathcal{T}}(\widehat{C}_j) \sqcap [u_{\mathcal{T}}(C)]_{\bowtie t} \sqcap \prod_{j=i+1}^{n} u_{\mathfrak{T}}(\widehat{C}_j)$$

  and

  $$u_{\mathfrak{T}'}(\beta) = \prod_{j=1}^{i-1} u_{\mathfrak{T}'}(\widehat{C}_j) \sqcap [u_{\mathcal{T}'}(E_C)]_{\bowtie t} \sqcap \prod_{j=i+1}^{n} u_{\mathfrak{T}'}(\widehat{C}_j)$$

  The application of induction (as in the previous case) yields $u_{\mathfrak{T}}(\widehat{C}_j) = u_{\mathfrak{T}'}(\widehat{C}_j)$ for all $j \neq i$, and $u_{\mathcal{T}}(C) = u_{\mathcal{T}'}(C)$. Thus, $u_{\mathcal{T}'}(E_C) = u_{\mathcal{T}'}(C)$ implies $u_{\mathfrak{T}}(\beta) = u_{\mathfrak{T}'}(\beta)$.

  – $R_\exists$: the rule is applied to $\widehat{C}_i = \exists r.\widehat{D}$ such that $\widehat{D} \notin \mathsf{NC}_d^{\mathfrak{T}}$. This means that $\widehat{C}_i$ is substituted in $\widehat{C}_\beta$ by $\exists r.\beta_1$, where β_1 is a fresh concept name from ND, and $\beta_1 \doteq \widehat{D}$ is added to $\mathfrak{T}'$. Since $u_{\mathfrak{T}'}(\beta_1) = u_{\mathfrak{T}'}(\widehat{D})$, the same arguments used for $R_{\bowtie}$ yield $u_{\mathfrak{T}}(\exists r.\widehat{D}) = u_{\mathfrak{T}'}(\exists r.\beta_1)$ and $u_{\mathfrak{T}}(\beta) = u_{\mathfrak{T}'}(\beta)$.

  – R_E: there is $1 \leq i \leq n$ such that $\widehat{C}_i = \beta_1$ and $\beta_1 \doteq \widehat{C}_{\beta_1} \in \mathfrak{T}$. The application of R_E replaces β_1 in $\widehat{C}_\beta$ with $\widehat{C}_{\beta_1}$. Since $\beta \to^+ \beta_1$, the application of induction yields $u_{\mathfrak{T}}(\beta_1) = u_{\mathfrak{T}'}(\beta_1) = u_{\mathfrak{T}'}(\widehat{C}_{\beta_1})$. Again, $u_{\mathfrak{T}}(\widehat{C}_j) = u_{\mathfrak{T}'}(\widehat{C}_j)$ for all $j \neq i$, and the rest follows from applying the definition of unfolding on β. □

Let us continue with the reduced form for acyclic $\tau\mathcal{EL}(m)$ TBoxes, which has been already introduced in Definition 4.9. To transform an acyclic $\mathcal{EL}$ TBox into reduced form, we reuse the polynomial time algorithm sketched in [27] (derived from Proposition 6.3.1) to compute the reduced form of $\mathcal{EL}$ concepts. The idea is to use $\sqsubseteq_{\mathcal{T}}$ instead of $\sqsubseteq$, when applying the rewrite rule described in Section 2. As a result, given an $\mathcal{EL}$ concept C, the modified procedure yields a concept C^r such that $C \equiv_{\mathcal{T}} C^r$. Moreover, since $\sqsubseteq_{\mathcal{T}}$ is decidable in polynomial time in $\mathcal{EL}$ [5], this procedure also runs in polynomial time. Based on this we devise a very simple polynomial time transformation that takes an acyclic $\mathcal{EL}$ TBox $\mathcal{T}$ and outputs an equivalent TBox $\mathcal{T}'$ in reduced form. This transformation and its correctness are given in the following lemma.

Lemma B.8. *Let $\mathcal{T}$ be a normalized acyclic $\mathcal{EL}$ TBox. The TBox $\mathcal{T}'$ obtained from $\mathcal{T}$ by substituting $E \doteq C_E$ with $E \doteq C_E^r$ (for all $E \doteq C_E \in \mathcal{T}$) satisfies the following.*

1. *For all interpretations $\mathcal{I}$: $\mathcal{I} \models \mathcal{T}$ iff $\mathcal{I} \models \mathcal{T}'$.*
2. *$\mathcal{T}'$ is in reduced form.*

Proof. 1. Let $\mathcal{I}$ be an interpretation such that $\mathcal{I} \models \mathcal{T}$. We know that $E^{\mathcal{I}} = C_E^{\mathcal{I}}$ for all $E \doteq C_E \in \mathcal{T}$, and $C_E \equiv_{\mathcal{T}} C_E^r$. Hence, $E^{\mathcal{I}} = {C_E^r}^{\mathcal{I}}$ for all $E \doteq C_E^r \in \mathcal{T}'$, and thus $\mathcal{I} \models \mathcal{T}'$.

Conversely, assume that $\mathcal{I}$ is a model of $\mathcal{T}'$. Since $\mathcal{T}$ is normalized, a top-level atom of C_E is of the form $A \in \mathsf{NC}_{pr}^{\mathcal{T}}$ or $\exists r.F$, where $F \doteq C_F \in \mathcal{T}$. This implies that $top(C_E^r) \subseteq top(C_E)$ for all $E \doteq C_E \in \mathcal{T}$. Therefore, we have that $C_E^{\mathcal{I}} \subseteq {C_E^r}^{\mathcal{I}}$ and $C_E^{\mathcal{I}} \subseteq E^{\mathcal{I}}$. To see that $\mathcal{I} \models \mathcal{T}$, it then remains to prove that $E^{\mathcal{I}} \subseteq C_E^{\mathcal{I}}$. To this end, we show that $d \in E^{\mathcal{I}}$ implies $d \in \mathsf{At}^{\mathcal{I}}$ for all $\mathsf{At} \in top(C_E)$. We make a case distinction based on whether At is a top-level atom of C_E^r.

- $\mathsf{At} \in top(C_E^r)$. Since $E \doteq C_E^r \in \mathcal{T}'$ and $\mathcal{I} \models \mathcal{T}'$, then $d \in E^{\mathcal{I}}$ implies $d \in \mathsf{At}^{\mathcal{I}}$.
- $\mathsf{At} \notin top(C_E^r)$. In this case, there must exist $\mathsf{At}' \in top(C_E)$ such that $\mathsf{At}' \sqsubseteq_{\mathcal{T}} \mathsf{At}$ and At' is a top-level atom of C_E^r. The previous case tells us that $d \in E^{\mathcal{I}}$ implies $d \in \mathsf{At}'^{\mathcal{I}}$. Thus, $d \in E^{\mathcal{I}}$ implies $d \in \mathsf{At}^{\mathcal{I}}$.

2. Assume that $\mathcal{T}'$ is not in reduced form. Then, there is $E \doteq C_E^r \in \mathcal{T}'$ such that C_E^r is not reduced w.r.t. $\mathcal{T}'$. This means that there are $\mathsf{At}_1, \mathsf{At}_2 \in top(C_E^r)$ such that $\mathsf{At}_1 \sqsubseteq_{\mathcal{T}'} \mathsf{At}_2$. Since we just have shown that $\mathcal{T}$ and $\mathcal{T}'$ are equivalent from a model-theoretic point of view, we also have $\mathsf{At}_1 \sqsubseteq_{\mathcal{T}} \mathsf{At}_2$. Hence, we obtain a contradiction against the fact that C_E^r is reduced w.r.t. $\mathcal{T}$. Thus, $\mathcal{T}'$ must be in reduced form. □

Overall, given an acyclic $\tau\mathcal{EL}(m)$ TBox $\mathfrak{T} = (\widehat{\mathcal{T}}, \mathcal{T})$ we have that:

- $\mathfrak{T}$ can be normalized in polynomial time into an acyclic TBox $\mathfrak{T}' = (\widehat{\mathcal{T}}', \mathcal{T}')$ such that reasoning w.r.t. $\mathfrak{T}$ can be reduced to reasoning w.r.t. $\mathfrak{T}'$.
- The TBox $\mathcal{T}'$ can be translated in polynomial time into an equivalent $\mathcal{EL}$ TBox $\mathcal{T}''$ in reduced form.

- Since computing $\mathcal{T}''$ only removes atoms from concept definitions, $(\widehat{\mathcal{T}}', \mathcal{T}'')$ remains normalized.

Thus, reasoning in $\tau\mathcal{EL}(m)$ w.r.t. acyclic TBoxes can be restricted to normalized acyclic TBoxes in reduced form.

Proposition B.9. *Satisfiability and subsumption on concepts defined in an acyclic $\tau\mathcal{EL}(m)$ TBox can be reduced in polynomial time to satisfiability and subsumption on concepts defined in a normalized acyclic $\tau\mathcal{EL}(m)$ TBox in reduced form.*

We conclude by showing that unfolding w.r.t. normalized $\mathcal{EL}$ TBoxes in reduced form always yields $\mathcal{EL}$ concept descriptions in reduced form.

Lemma 4.10. *Let $\mathcal{T}$ be a normalized acyclic $\mathcal{EL}$ TBox in reduced form. Then, for all concept names $E \in \mathsf{NC}_d^{\mathcal{T}}$, the unfolded $\mathcal{EL}$ concept description $u_{\mathcal{T}}(E)$ is reduced.*

Proof. We use induction on the partial order induced by $\to^+$ on $\mathsf{NC}_d^{\mathcal{T}}$. Since $\mathcal{T}$ is normalized, C_E is of the form $P_1 \sqcap \cdots \sqcap P_k \sqcap \exists r_1.E_1 \sqcap \cdots \sqcap \exists r_n.E_n$. Then, the application of induction to each E_i yields that $u_{\mathcal{T}}(E_i)$ is reduced. In addition, since C_E is reduced w.r.t. $\mathcal{T}$, for all pairs $(\exists r_i.E_i, \exists r_j.E_j)$ we have:

- $r_i \neq r_j$ or ($E_i \not\sqsubseteq_{\mathcal{T}} E_j$ and $E_j \not\sqsubseteq_{\mathcal{T}} E_i$).

Further, we know that $E_i \equiv_{\mathcal{T}} u_{\mathcal{T}}(E_i)$ and $E_j \equiv_{\mathcal{T}} u_{\mathcal{T}}(E_j)$. Hence, $r_i = r_j$ implies that $u_{\mathcal{T}}(E_i) \not\sqsubseteq u_{\mathcal{T}}(E_j)$ and $u_{\mathcal{T}}(E_j) \not\sqsubseteq u_{\mathcal{T}}(E_i)$. Thus, since $u_{\mathcal{T}}(E)$ corresponds to $P_1 \sqcap \cdots P_k \sqcap \exists r_1.u_{\mathcal{T}}(E_1) \sqcap \cdots \sqcap \exists r_n.u_{\mathcal{T}}(E_n)$, we can conclude that $u_{\mathcal{T}}(E)$ is reduced. □

Appendix C. Missing proofs from Section 5

This appendix contains the missing proofs from Section 5. More precisely, we show the equivalence in (37) used in the proof of Lemma 5.2. To this end, we first need to prove two auxiliary results.

Lemma C.1. *The interpretation $\mathcal{I}$ constructed in the proof of Lemma 5.2 is a model of $\mathcal{T}$.*

Proof. First, we show that the following holds for all $\mathcal{EL}$ concepts occurring in $\mathcal{T}$ and for all $d_\mu \in \Delta^{\mathcal{I}}$:

$$C \in \nu \text{ iff } d_\mu \in C^{\mathcal{I}}.$$

Based on the normal form of $\mathcal{T}$, we distinguish three cases corresponding to the possible form of C:

- $C = A \in \mathsf{NC} \cup \mathsf{ND}$. In this case, the equivalence holds as a direct consequence of the definition of $\mathcal{I}$.
- $C = \exists r.E$ where $E \in \mathsf{NC}_d^{\mathcal{T}}$. Consider the mapping ψ^μ selected in Lemma 5.2. Since it leads to a successful run of SUCC, the sequence $\psi^\mu(r)$ is an r-successor candidate of μ w.r.t. $\mathsf{cl}(\mathfrak{T}, \sim)$. Suppose that $\exists r.E \in \nu$. We have $r \in \mathsf{rol}(\nu)$ and $E \in \nu^+(r)$. Since $\psi^\mu(r)$ satisfies c1), it must contain a type μ' such that $E \in \nu'$. The previous case yields $d_{\mu'} \in E^{\mathcal{I}}$. In addition, by definition of $\mathcal{I}$ we have $(d_\mu, d_{\mu'}) \in r^{\mathcal{I}}$. Thus, $d_\mu \in (\exists r.E)^{\mathcal{I}}$. Conversely, assume that $d_\mu \in (\exists r.E)^{\mathcal{I}}$. Then, there is $d_{\mu'} \in \Delta^{\mathcal{I}}$ such that $(d_\mu, d_{\mu'}) \in r^{\mathcal{I}}$ and $d_{\mu'} \in E^{\mathcal{I}}$. The construction of $\mathcal{I}$ tells us that μ' is in $\psi^\mu(r)$, and the previous case yields $E \in \nu'$. Again, by property c1), it follows that $E \in \nu^+(r)$ and thus $\exists r.E \in \nu$.
- C is a conjunction of concept names and existential restrictions. Our claim is then an easy consequence of the previous cases and the fact that ν satisfies t1).

Hence, given $d_\mu \in \Delta^{\mathcal{I}}$ and $E \doteq C_E \in \mathcal{T}$, we have that $d_\mu \in E^{\mathcal{I}}$ iff $E \in \nu$ iff $C_E \in \nu$ iff $d_\mu \in C_E^{\mathcal{I}}$. The equivalence in the middle holds because μ is a type for $\mathfrak{T}$. Thus, we have shown that $\mathcal{I} \models \mathcal{T}$. □

Lemma C.2. *For all $d_\mu \in \Delta^{\mathcal{I}}$ and $1 \leq i \leq m$, it holds that $\widehat{m}_\sim^{\mathcal{I}}(d_\mu, E_i, \mathcal{T}) = v_i$.*

Proof. The proof is by induction along the partial order on $\mathsf{NC}_d^{\mathcal{T}}$ induced by the dependency relation $\to$ of $\mathcal{T}$ (see Definition 2.1). Let $d_\mu \in \Delta^{\mathcal{I}}$ and $E_i \in \mathsf{NC}_d^{\mathcal{T}}$. In addition, let ψ^μ be the mapping selected for μ to construct $\mathcal{I}$. Since ψ^μ leads to a successful run of SUCC in which $q_i = v_i$ holds for all $i, 1 \leq i \leq m$ (see Line 22), it suffices to show that the value q_i computed by DEGREES at Line 21 is actually $\widehat{m}_\sim^{\mathcal{I}}(d_\mu, E_i, \mathcal{T})$.

Let $Q := \{(r, (q_1^e, \ldots, q_m^e)) \mid e \in \Delta^{\mathcal{I}} \wedge (d_\mu, e) \in r^{\mathcal{I}} \wedge r \in \mathsf{NR}\}$, where $q_i^e := \widehat{m}^{\mathcal{I}}(e, E_i, \mathcal{T})$. By construction, $\mathcal{I}$ is clearly a finite interpretation and $\nu \cap \mathsf{NC} = \{A \mid d_\mu \in A^{\mathcal{I}} \wedge A \in \mathsf{NC}\}$. Hence, since $\mathcal{I} \models \mathcal{T}$ by Lemma C.1, we can apply Lemma 4.11 to obtain

$$\textsc{degrees}(\nu \cap \mathsf{NC}, Q, E_i) = \widehat{m}_\sim^{\mathcal{I}}(d_\mu, E_i, \mathcal{T}) \quad (1 \leq i \leq m).$$

Note that, in this application of Lemma 4.11, the subsets Q_i are all equal to the whole set Q, and thus the second condition on the sets Q_i is trivially satisfied.

In Line 21 of Algorithm 4, the procedure DEGREES is not called with input Q, but with input $Q' := \{(r, (v'_1, \ldots, v'_m)) \mid r \in \mathfrak{R}(\mathfrak{T}, \sim) \wedge (v', v'_1, \ldots, v'_m) \text{ is in } \psi^\mu(r)\}$. By definition of $\mathcal{I}$, the following holds for all $r \in \mathsf{NR}$:

$$(d_\mu, d_{\mu'}) \in r^{\mathcal{I}} \text{ iff } r \in \mathfrak{R}(\mathfrak{T}, \sim) \text{ and } \mu' = (v', v'_1, \ldots, v'_m) \text{ is a tuple in } \psi^\mu(r).$$

Hence, there is a bijection between Q and Q', i.e., $(r, (q^e_1, \ldots, q^e_m)) \in Q$ corresponds to $(r, (v'_1, \ldots, v'_m)) \in Q'$ where $e = d_{\mu'}$. Let $\exists s.E_p \in C_{E_i}$. Since $E_i \to E_p$, we can apply induction to E_p and e to obtain that $q^e_p = \widehat{m}^{\mathcal{I}}_\sim(e, E_p, \mathcal{T}) = v'_p$. Since a computation of DEGREES(_, _, E_i) only uses values q_p such that $\exists s.E_p \in top(C_{E_i})$, it follows that DEGREES($v \cap \mathsf{NC}, Q, E_i$) = DEGREES($v \cap \mathsf{NC}, Q', E_i$). This completes the proof. □

We are finally ready to show the equivalence in (37).

Proof of (37). The proof is by induction on the structure of $\widehat{C}$. We distinguish the following cases:

- $\widehat{C} = (E_i)_{\bowtie t}$. We have that $(E_i)_{\bowtie t} \in v$ iff $v_i \bowtie t$, since μ satisfies t4) w.r.t. $\mathsf{cl}(\mathfrak{T}, \sim)$ and $\mathsf{NC}^{\mathcal{T}}_d$. In addition, the semantics of threshold concepts yields that $d_\mu \in [(E_i)_{\bowtie t}]^{\mathcal{I}}$ iff $\widehat{m}^{\mathcal{I}}_\sim(d_\mu, E_i, \mathcal{T}) \bowtie t$. Hence, since $\widehat{m}^{\mathcal{I}}_\sim(d_\mu, E_i, \mathcal{T}) = v_i$ by Lemma C.2, it thus follows that $(E_i)_{\bowtie t} \in v$ iff $d_\mu \in [(E_i)_{\bowtie t}]^{\mathcal{I}}$.
- The remaining cases, i.e., $\widehat{C} \in \mathsf{NC} \cap \mathsf{ND}$, $\widehat{C} = \exists r.\widehat{D}$, and $\widehat{C}$ is a conjunction, can be shown similarly to the proof of Lemma C.1, but now using induction rather than referring to the previous case(s). For instance, for $\widehat{C} = \exists r.\widehat{D}$, induction is used to obtain that $d_{\mu'} \in \widehat{D}^{\mathcal{I}}$ (*only if* direction) and $\widehat{D} \in v'$ (*if* direction). □

References

[1] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, P.F. Patel-Schneider (Eds.), The Description Logic Handbook: Theory, Implementation, and Applications, Cambridge University Press, 2003.
[2] F. Baader, I. Horrocks, C. Lutz, U. Sattler, An Introduction to Description Logic, Cambridge University Press, 2017.
[3] R. Hoehndorf, P.N. Schofield, G.V. Gkoutos, The role of ontologies in biological and biomedical research: a functional perspective, Brief. Bioinform. 16 (6) (2015) 1069–1080.
[4] M. Schmidt-Schauß, G. Smolka, Attributive concept descriptions with complements, Artif. Intell. 48 (1) (1991) 1–26.
[5] F. Baader, Terminological cycles in a description logic with existential restrictions, in: G. Gottlob, T. Walsh (Eds.), IJCAI-03, in: Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, 2003, Morgan Kaufmann, 2003, pp. 325–330.
[6] S. Brandt, Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and - what else?, in: R.L. de Mántaras, L. Saitta (Eds.), Proceedings of the 16th Eureopean Conference on Artificial Intelligence, ECAI'2004, Including Prestigious Applicants of Intelligent Systems, PAIS 2004, IOS Press, 2004, pp. 298–302.
[7] F. Baader, S. Brandt, C. Lutz, Pushing the $\mathcal{EL}$ envelope, in: IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30 - August 5, 2005, Professional Book Center, July 2005, pp. 364–369.
[8] J. Austen Pride, Prejudice, T. Egerton, London, 1813.
[9] E.D. Sciascio, F.M. Donini, T.D. Noia, Introduction to the special issue: semantic matchmaking and resource retrieval on the web, Int. J. Electron. Commer. 12 (2) (2007) 5–9.
[10] T.D. Noia, E.D. Sciascio, F.M. Donini, Semantic matchmaking as non-monotonic reasoning: a description logic approach, J. Artif. Intell. Res. 29 (2007) 269–307.
[11] F. Baader, G. Brewka, O. Fernández Gil, Adding threshold concepts to the description logic $\mathcal{EL}$, in: C. Lutz, S. Ranise (Eds.), Frontiers of Combining Systems - 10th International Symposium, FroCoS 2015. Proceedings, in: Lecture Notes in Computer Science, vol. 9322, Springer, 2015, pp. 33–48.
[12] A. Ecke, R. Peñaloza, A.-Y. Turhan, Similarity-based relaxed instance queries, J. Appl. Log. 13 (4) (2015) 480–508.
[13] C. d'Amato, N. Fanizzi, F. Esposito, A semantic similarity measure for expressive description logics, in: A. Pettorossi (Ed.), Proceedings of Convegno Italiano di Logica Computazionale (CILC05), 2005.
[14] K. Lehmann, A.-Y. Turhan, A framework for semantic-based similarity measures for $\mathcal{ELH}$-concepts, in: L.F. del Cerro, A. Herzig, J. Mengin (Eds.), Logics in Artificial Intelligence - 13th European Conference, JELIA 2012. Proceedings, in: Lecture Notes in Computer Science, vol. 7519, Springer, 2012, pp. 307–319.
[15] B. Suntisrivaraporn, A similarity measure for the description logic $\mathcal{EL}$ with unfoldable terminologies, in: 2013 5th International Conference on Intelligent Networking and Collaborative Systems, 2013, IEEE, 2013, pp. 408–413.
[16] F. Baader, O. Fernández Gil, P. Marantidis, Approximation in description logics: how weighted tree automata can help to define the required concept comparison measures in $\mathcal{FL}_0$, in: F. Drewes, C. Martín-Vide, B. Truthe (Eds.), Language and Automata Theory and Applications - 11th International Conference, LATA 2017, Proceedings, in: Lecture Notes in Computer Science, vol. 10168, 2017, pp. 3–26.
[17] F. Baader, O. Fernández Gil, Decidability and complexity of threshold description logics induced by concept similarity measures, in: A. Seffah, B. Penzenstadler, C. Alves, X. Peng (Eds.), Proceedings of the Symposium on Applied Computing, SAC 2017, ACM, 2017, pp. 983–988.
[18] F. Baader, P. Koopmann, F. Michel, A. Turhan, B. Zarrieß, Efficient TBox reasoning with value restrictions using the $\mathcal{FL}_0$wer reasoner, Theory Pract. Log. Program. 22 (2) (2022) 162–192.
[19] F. Baader, O. Fernández Gil, Extending the description logic $\tau\mathcal{EL}(deg)$ with acyclic TBoxes, in: G.A. Kaminka, M. Fox, P. Bouquet, E. Hüllermeier, V. Dignum, F. Dignum, F. van Harmelen (Eds.), ECAI 2016 - 22nd European Conference on Artificial Intelligence - Including Prestigious Applications of Artificial Intelligence (PAIS 2016), in: Frontiers in Artificial Intelligence and Applications, vol. 285, IOS Press, 2016, pp. 1096–1104.
[20] F. Baader, R. Küsters, R. Molitor, Computing least common subsumers in description logics with existential restrictions, in: T. Dean (Ed.), Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, IJCAI 99, 1999, Morgan Kaufmann, 1999, pp. 96–103, long version available as [21].
[21] F. Baader, R. Küsters, R. Molitor, Computing Least Common Subsumers in Description Logics with Existential Restrictions, LTCS-Report, http://lat.inf.tu-dresden.de/research/reports.html, 1998, LTCS-98-09. Chair for Automata Theory, Institute for Theoretical Computer Science Dresden University of Technology, Germany.
[22] A. Borgida, T.J. Walsh, H. Hirsh, Towards measuring similarity in description logics, in: I. Horrocks, U. Sattler, F. Wolter (Eds.), Proceedings of the 2005 International Workshop on Description Logics (DL2005), vol. 147, 2005, CEUR-WS.org.
[23] T. Racharak, B. Suntisrivaraporn, S. Tojo, Personalizing a concept similarity measure in the description logic $\mathcal{ELH}$ with preference profile, Comput. Inform. 37 (3) (2018) 581–613.
[24] T. Racharak, On approximation of concept similarity measure in description logic $\mathcal{ELH}$ with pre-trained word embedding, IEEE Access 9 (2021) 61429–61443.
[25] B. Nebel, Terminological cycles: semantics and computational properties, in: J.F. Sowa (Ed.), Principles of Semantic Networks: Explorations in the Representation of Knowledge, in: Morgan Kaufmann Ser. Represent. Reason., Morgan Kaufmann, 1991, pp. 331–361.

[26] F. Baader, B. Sertkaya, A. Turhan, Computing the least common subsumer w.r.t. a background terminology, J. Appl. Log. 5 (3) (2007) 392–420.
[27] R. Küsters, Non-Standard Inferences in Description Logics, Lecture Notes in Computer Science, vol. 2100, Springer, 2001.
[28] F. Baader, B. Morawska, Unification in the description logic $\mathcal{EL}$, Log. Methods Comput. Sci. 6 (3) (2010).
[29] A. Ecke, R. Peñaloza, A. Turhan, Answering instance queries relaxed by concept similarity, in: C. Baral, G.D. Giacomo, T. Eiter (Eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20-24, 2014, AAAI Press, 2014, pp. 248–257.
[30] K. Lehmann, A framework for semantic invariant similarity measures for $\mathcal{ELH}$ concept descriptions, Master's thesis, TU Dresden, Germany, 2012, available from: http://lat.inf.tu-dresden.de/research/mas.
[31] D. Calvanese, G.D. Giacomo, D. Lembo, M. Lenzerini, R. Rosati, Data complexity of query answering in description logics, Artif. Intell. 195 (2013) 335–360.
[32] A. Schaerf, On the complexity of the instance checking problem in concept languages with existential quantification, J. Intell. Inf. Syst. 2 (3) (1993) 265–278.
[33] F.M. Donini, M. Lenzerini, D. Nardi, A. Schaerf, Deduction in concept languages: from subsumption to instance checking, J. Log. Comput. 4 (4) (1994) 423–452.
[34] B. Nebel, Terminological reasoning is inherently intractable, Artif. Intell. 43 (2) (1990) 235–249.
[35] W.J. Savitch, Relationships between nondeterministic and deterministic tape complexities, J. Comput. Syst. Sci. 4 (2) (1970) 177–192.
[36] K. Schild, A correspondence theory for terminological logics: preliminary report, in: J. Mylopoulos, R. Reiter (Eds.), Proceedings of the 12th International Joint Conference on Artificial Intelligence, 1991, Morgan Kaufmann, 1991, pp. 466–471.
[37] L.J. Stockmeyer, A.K. Chandra, Provably difficult combinatorial games, SIAM J. Comput. 8 (2) (1979) 151–174.
[38] A. Krisnadhi, C. Lutz, Data complexity in the EL family of description logics, in: Logic for Programming, Artificial Intelligence, and Reasoning, 14th International Conference, LPAR 2007, Yerevan, Armenia, October 15-19, 2007, Proceedings, in: Lecture Notes in Computer Science, vol. 4790, Springer, 2007, pp. 333–347.
[39] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (3) (1965) 338–353.
[40] J. Yen, Generalizing term subsumption languages to fuzzy logic, in: J. Mylopoulos, R. Reiter (Eds.), Proceedings of the 12th International Joint Conference on Artificial Intelligence, 1991, Morgan Kaufmann, 1991, pp. 472–477.
[41] U. Straccia, Reasoning within fuzzy description logics, J. Artif. Intell. Res. 14 (2001) 137–166.
[42] P. Hájek, Making fuzzy description logic more general, Fuzzy Sets Syst. 154 (1) (2005) 1–15.
[43] S. Borgwardt, R. Peñaloza, Positive subsumption in fuzzy $\mathcal{EL}$ with general t-norms, in: IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013, IJCAI/AAAI, 2013, pp. 789–795.
[44] F. Baader, S. Borgwardt, R. Peñaloza, On the decidability status of fuzzy $\mathcal{ALC}$ with general concept inclusions, J. Philos. Log. 44 (2) (2015) 117–146.
[45] S. Borgwardt, F. Distel, R. Peñaloza, The limits of decidability in fuzzy description logics with general concept inclusions, Artif. Intell. 218 (2015) 23–55.
[46] F. Baader, A. Ecke, Reasoning with prototypes in the description logic $\mathcal{ALC}$ using weighted tree automata, in: Language and Automata Theory and Applications - 10th International Conference, LATA 2016, Prague, Czech Republic, March 14-18, 2016, Proceedings, in: Lecture Notes in Computer Science, vol. 9618, Springer, 2016, pp. 63–75.
[47] C. Lutz, F. Wolter, M. Zakharyaschev, A tableau algorithm for reasoning about concepts and similarity, in: Automated Reasoning with Analytic Tableaux and Related Methods, International Conference, TABLEAUX 2003, Rome, Italy, September 9-12, 2003. Proceedings, in: Lecture Notes in Computer Science, vol. 2796, Springer, 2003, pp. 134–149.
[48] F. Wolter, M. Zakharyaschev, Reasoning about distances, in: IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003, Morgan Kaufmann, 2003, pp. 1275–1282.
[49] M. Sheremet, D. Tishkovsky, F. Wolter, M. Zakharyaschev, A logic for concepts and similarity, J. Log. Comput. 17 (3) (2007) 415–452.
[50] D. Porello, O. Kutz, G. Righetti, N. Troquard, P. Galliani, C. Masolo, A toothful of concepts: towards a theory of weighted concept combination, in: Proceedings of the 32nd International Workshop on Description Logics, Oslo, Norway, June 18–21, 2019, in: CEUR Workshop Proc., vol. 2373, 2019, CEUR-WS.org.
[51] P. Galliani, O. Kutz, D. Porello, G. Righetti, N. Troquard, On knowledge dependence in weighted description logic, in: GCAI 2019. Proceedings of the 5th Global Conference on Artificial Intelligence, Bozen/Bolzano, Italy, 17–19 September 2019, in: EPiC Series in Computing, vol. 65, EasyChair, 2019, pp. 68–80.
[52] P. Galliani, G. Righetti, O. Kutz, D. Porello, N. Troquard, Perceptron connectives in knowledge representation, in: Knowledge Engineering and Knowledge Management - 22nd International Conference, EKAW 2020, Bolzano, Italy, September 16-20, 2020, Proceedings, in: Lecture Notes in Computer Science, vol. 12387, Springer, 2020, pp. 183–193.
[53] D. Tsarkov, I. Horrocks, P.F. Patel-Schneider, Optimizing terminological reasoning for expressive description logics, J. Autom. Reason. 39 (3) (2007) 277–316.
[54] B. Glimm, I. Horrocks, B. Motik, G. Stoilos, Z. Wang, Hermit: an OWL 2 reasoner, J. Autom. Reason. 53 (3) (2014) 245–269.
[55] B. Suntisrivaraporn, S. Tongphu, A structural subsumption based similarity measure for the description logic ALEH, in: H.J. van den Herik, J. Filipe (Eds.), Proceedings of the 8th International Conference on Agents and Artificial Intelligence (ICAART 2016), Volume 2, Rome, Italy, February 24-26, 2016, SciTePress, 2016, pp. 204–212.
[56] F. Baader, Terminological cycles in a description logic with existential restrictions, LTCS-Report LTCS-02-02, Chair for Automata Theory, Institute for Theoretical Computer Science, Dresden University of Technology, Germany, 2002, see http://lat.inf.tu-dresden.de/research/reports.html.