



TECHNISCHE
UNIVERSITÄT
DRESDEN

Technische Universität Dresden
Institute for Theoretical Computer Science
Chair for Automata Theory

LTCS–Report

Extending the Description Logic $\tau\mathcal{EL}(deg)$ with Acyclic TBoxes

Franz Baader Oliver Fernández Gil

LTCS-Report 16-02

Postal Address:
Lehrstuhl für Automatentheorie
Institut für Theoretische Informatik
TU Dresden
01062 Dresden

<http://lat.inf.tu-dresden.de>

Visiting Address:
Nöthnitzer Str. 46
Dresden

Extending the Description Logic $\tau\mathcal{EL}(deg)$ with Acyclic TBoxes

Franz Baader¹ and Oliver Fernández Gil^{*1}

¹Theoretical Computer Science, TU Dresden, Germany

Abstract

In a previous paper, we have introduced an extension of the lightweight Description Logic \mathcal{EL} that allows us to define concepts in an approximate way. For this purpose, we have defined a graded membership function deg , which for each individual and concept yields a number in the interval $[0, 1]$ expressing the degree to which the individual belongs to the concept. Threshold concepts $C_{\sim t}$ for $\sim \in \{<, \leq, >, \geq\}$ then collect all the individuals that belong to C with degree $\sim t$. We have then investigated the complexity of reasoning in the Description Logic $\tau\mathcal{EL}(deg)$, which is obtained from \mathcal{EL} by adding such threshold concepts. In the present paper, we extend these results, which were obtained for reasoning without TBoxes, to the case of reasoning w.r.t. acyclic TBoxes. 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, it turns out that, in contrast to the case of \mathcal{EL} , adding acyclic TBoxes to $\tau\mathcal{EL}(deg)$ increases the complexity of reasoning by at least on level of the polynomial hierarchy.

^{*}Supported by DFG Graduiertenkolleg 1763 (QuantLA).

Contents

1	Introduction	3
2	The Description Logic $\tau\mathcal{EL}(deg)$	6
2.1	The Description Logic \mathcal{EL}	6
2.1.1	Characterization of membership in \mathcal{EL}	8
2.2	Adding threshold concepts to \mathcal{EL}	10
2.3	The graded membership function deg	11
3	Acyclic TBoxes for $\tau\mathcal{EL}(m)$	15
4	Reasoning with acyclic $\tau\mathcal{EL}(deg)$ TBoxes	19
4.1	Lower bounds	20
4.2	Normalization	30
4.3	A PSpace upper bound	37
4.4	Reasoning with acyclic knowledge bases	47
5	Conclusion	57
6	Appendix	58

1 Introduction

Description logics (DLs) [BCM⁺03] allow their users to define the important notions of an application domain as concepts by stating necessary and sufficient conditions for an individual to belong to the 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 concept. Basically, such a TBox introduces abbreviations for complex concept descriptions. But even this simple form of TBoxes may increase the complexity of reasoning, as is, for example the case for the DL \mathcal{FL}_0 , for which the complexity of the subsumption problem increases from polynomial-time to coNP-complete if acyclic TBoxes are added [Neb90].

The DL \mathcal{EL} , in which concepts can be built using concept names as well as the concept constructors conjunction (\sqcap), existential restriction ($\exists r.C$), and the top concept (\top),¹ has drawn considerable attention in the last decade since, on the one hand, important inference problems such as the subsumption problem are polynomial in \mathcal{EL} , not only w.r.t. acyclic TBoxes, but also w.r.t. more expressive terminological axioms called GCIs [Bra04]. On the other hand, though quite inexpressive, \mathcal{EL} can be used to define biomedical ontologies, such as the large medical ontology SNOMED CT,² which basically is an acyclic \mathcal{EL} TBox. In \mathcal{EL} we can, for example, define the concept of a *good movie* as a movie that is uplifting, has a simple, but original plot, a likable and an evil character, action and love scenes, and a happy ending.

$$\begin{aligned}
 \text{Movie} \sqcap \text{Uplifting} \sqcap \exists \text{plot} . (\text{Simple} \sqcap \text{Original}) \sqcap \\
 \exists \text{character} . \text{Likeable} \sqcap \exists \text{character} . \text{Evil} \sqcap \\
 \exists \text{scene} . \text{Action} \sqcap \exists \text{scene} . \text{Love} \sqcap \exists \text{ending} . \text{Happy}.
 \end{aligned} \tag{1}$$

For an individual to belong to this concept, all the stated properties need to be satisfied. However, maybe we would still want to call a movie good if most, though not all, of the properties hold.

In [BBG15], we have introduced a DL extending \mathcal{EL} that allows us to define concepts in such an approximate way. The main idea is to use a *graded membership function*, which instead of a Boolean membership value 0 or 1 yields a membership degree from the interval $[0, 1]$. We can then require a good movie to belong

¹In \mathcal{FL}_0 , we have value restrictions ($\forall r.C$) instead of existential restrictions.

²see <http://www.ihtsdo.org/snomed-ct/>

to the \mathcal{EL} concept (1) with degree at least .8. More generally, if C is an \mathcal{EL} concept, then the *threshold concept* $C_{\geq t}$ for $t \in [0, 1]$ collects all the individuals that belong to C with degree at least t . In addition to such upper threshold concepts, also lower threshold concepts $C_{\leq t}$ are considered, and strict inequalities may be used. For example, a bad movie could be required to belong to the \mathcal{EL} concept (1) with a degree less than .2. In contrast to fuzzy logics, which also yield membership degrees, we use classical crisp interpretations to define the semantics of the new logic. The membership degree of an individual d in a concept C is obtained by comparing the properties that the individual has with the properties that the concept requires. There are, of course, different possibilities for how to define a graded membership function m based on this idea, and the semantics of the obtained new logic $\tau\mathcal{EL}(m)$ depends on m . In [BBG15], we have not only introduced this general framework, but have also defined a specific graded membership function deg , and have investigated the complexity of reasoning in $\tau\mathcal{EL}(deg)$ in detail. More precisely, we have shown that the satisfiability and the ABox consistency problem in $\tau\mathcal{EL}(deg)$ are NP-complete, and the subsumption and the instance problem are coNP-complete (the latter w.r.t. data complexity). All these results are shown for the setting without TBoxes.

The main contribution of the present paper is to investigate the complexity of reasoning in $\tau\mathcal{EL}(deg)$ w.r.t. acyclic TBoxes. Surprisingly, this is not as easy as might have been expected. The problem already starts with how to define acyclic TBoxes in $\tau\mathcal{EL}(deg)$. It turns out that simply replacing \mathcal{EL} concepts by $\tau\mathcal{EL}(deg)$ concepts in the definition of an acyclic TBox does not yield the desired result. In fact, acyclic TBoxes are supposed to introduce concept names (defined concepts) as abbreviations for complex concepts, and these complex concepts can be obtained by *unfolding* defined concepts, i.e., by replacing defined concepts by their definitions until no more defined concepts occur. For the straightforward definition of acyclic $\tau\mathcal{EL}(deg)$ TBoxes mentioned above, this unfolding would actually yield concepts that are not syntactically correct $\tau\mathcal{EL}(deg)$ concepts since they may contain nested threshold operators, which is not allowed in $\tau\mathcal{EL}(deg)$.³ Thus, we propose a more sophisticated notion of acyclic TBox for $\tau\mathcal{EL}(deg)$, which consists of an \mathcal{EL} part and a $\tau\mathcal{EL}(deg)$ part satisfying certain properties. These properties ensures that unfolding yields a correct $\tau\mathcal{EL}(deg)$ concept. Of course, from a semantic point of view, we want defined concepts to have the same meaning as their unfolded counterparts. For this to hold, we need to require the graded membership function to “respect” the \mathcal{EL} part of the TBox in an appropriate way. We show how deg can be modified such that it satisfies this requirement. This finally fixes syntax and semantics of acyclic $\tau\mathcal{EL}(deg)$ TBoxes. We then investigate reasoning w.r.t. such TBoxes. We show that, again quite surprisingly, the complexity increases by at least one level in the polynomial hierarchy when acyclic TBoxes are added: satisfiability and consistency are Π_2^P -

³In fact, the semantics of such nested concepts would not be well-defined since the graded membership function can only deal with \mathcal{EL} concepts as input.

hard and subsumption and the instance problem are Σ_2^P -hard. The best upper bound we can currently show for these problems is PSpace.

In the next section, we will sketch how the DL $\tau\mathcal{EL}(deg)$ was defined in [BBG15] (more details and motivating discussions can be found in that paper). In Section 3, we introduce acyclic $\tau\mathcal{EL}(deg)$ TBoxes, and in Section 4 we provide proofs of the mentioned complexity results.

2 The Description Logic $\tau\mathcal{EL}(deg)$

We start by introducing the DL \mathcal{EL} and some related notions that are needed in the rest of the paper. Afterwards, we present the abstract family of DLs $\tau\mathcal{EL}(m)$ that is obtained by extending \mathcal{EL} with threshold concepts defined using a graded membership function m [BBG15]. Finally, we recall the specific graded membership function deg , and briefly discuss the results obtained in [BBG15] concerning the computational complexity of reasoning in $\tau\mathcal{EL}(deg)$.

2.1 The Description Logic \mathcal{EL}

Starting with finite sets of concept names \mathbf{N}_C and role names \mathbf{N}_R , the set $\mathcal{C}_{\mathcal{EL}}$ of \mathcal{EL} concept descriptions is obtained by combining the concept constructors *conjunction* ($C \sqcap D$), *existential restriction* ($\exists r.C$) and *top* (\top), in the following way:

$$C ::= \top \mid A \mid C \sqcap C \mid \exists r.C \quad (2)$$

where $A \in \mathbf{N}_C$, $r \in \mathbf{N}_R$ and $C \in \mathcal{C}_{\mathcal{EL}}$.

The semantics of \mathcal{EL} is given through standard *first-order* logic interpretations. An *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consists of a non-empty domain $\Delta^{\mathcal{I}}$ and an interpretation function $\cdot^{\mathcal{I}}$ that assigns subsets of $\Delta^{\mathcal{I}}$ to concept names in \mathbf{N}_C and binary relations over $\Delta^{\mathcal{I}}$ to role names in \mathbf{N}_R . The function $\cdot^{\mathcal{I}}$ is inductively extended to arbitrary concept descriptions in the usual way, i.e.,

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

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

Information about specific individuals (represented by a set of individual names \mathbf{N}_I) can be expressed in an ABox, which is a finite set of *assertions* of the form $C(a)$ or $r(a, b)$, where $C \in \mathcal{C}_{\mathcal{EL}}$, $r \in \mathbf{N}_R$, and $a, b \in \mathbf{N}_I$. In addition to concept and role names, an interpretation \mathcal{I} now assigns domain elements $a^{\mathcal{I}}$ to individual names a . We say that \mathcal{I} satisfies an assertion $C(a)$ iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$, and $r(a, b)$ iff $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$. Further, \mathcal{I} is a *model* of \mathcal{A} (denoted as $\mathcal{I} \models \mathcal{A}$) iff it satisfies all the assertions of \mathcal{A} . Then, an ABox \mathcal{A} is *consistent* iff $\mathcal{I} \models \mathcal{A}$ for some interpretation \mathcal{I} . Finally, an individual a is an *instance* of C in \mathcal{A} iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$ for all models \mathcal{I} of \mathcal{A} . We denote the set of individual names occurring in \mathcal{A} as $\text{Ind}(\mathcal{A})$.

An \mathcal{EL} TBox \mathcal{T} is a finite set of concept definitions of the form $E \doteq C_E$, where $E \in \mathbf{N}_C$ and C_E is an \mathcal{EL} concept description.⁴ Additionally, 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 of \mathcal{T} are called *defined concepts* while all other concept names are called *primitive concepts*. The sets of defined and primitive concepts of \mathcal{T} are denoted as $\mathbf{N}_d^{\mathcal{T}}$ and $\mathbf{N}_{pr}^{\mathcal{T}}$, respectively. Note that $\mathbf{N}_{pr}^{\mathcal{T}} = \mathbf{N}_C \setminus \mathbf{N}_d^{\mathcal{T}}$, and thus also contains all concept names not occurring in \mathcal{T} . An interpretation \mathcal{I} is a model of \mathcal{T} (in symbols $\mathcal{I} \models \mathcal{T}$) iff $E^{\mathcal{I}} = (C_E)^{\mathcal{I}}$ for all $E \doteq C_E \in \mathcal{T}$. The relations \sqsubseteq and \equiv are now defined *modulo* all models of \mathcal{T} and denoted as $\sqsubseteq_{\mathcal{T}}$ and $\equiv_{\mathcal{T}}$, respectively.

TBoxes can be classified into being *acyclic* or *cyclic*, based on how their defined concepts depend on each other.

Definition 1 (Cyclic/acyclic \mathcal{EL} TBoxes). Let \mathcal{T} be an \mathcal{EL} TBox. We define \rightarrow as a binary relation over the set $\mathbf{N}_d^{\mathcal{T}}$ to represent *direct dependency* between defined concepts in the following way. A defined concept E_1 directly depends on a defined concept E_2 (denoted as $E_1 \rightarrow E_2$) iff $E_1 \doteq C_{E_1} \in \mathcal{T}$ and E_2 occurs in C_{E_1} . Let \rightarrow^+ be the transitive closure of \rightarrow . The TBox \mathcal{T} contains a terminological cycle iff there exists a defined concept E in \mathcal{T} that depends on itself, i.e., $E \rightarrow^+ E$. Then, \mathcal{T} is called *cyclic* if it contains a terminological cycle. Otherwise, it is called *acyclic*.

For acyclic TBoxes, the relation \rightarrow^+ induces a *well-founded* partial order \preceq on the set $\mathbf{N}_d^{\mathcal{T}}$, i.e., $E_1 \preceq E_2$ iff $E_2 \rightarrow^+ E_1$. Furthermore, the unfolding $u_{\mathcal{T}}(C)$ of an \mathcal{EL} concept description C with respect to \mathcal{T} is defined as follows:

$$u_{\mathcal{T}}(C) := \begin{cases} C & \text{if } C \in \mathbf{N}_{pr}^{\mathcal{T}} \\ u_{\mathcal{T}}(C_E) & \text{if } C = E \text{ and } E \doteq C_E \in \mathcal{T} \\ u_{\mathcal{T}}(C_1) \sqcap u_{\mathcal{T}}(C_2) & \text{if } C = C_1 \sqcap C_2 \\ \exists r. u_{\mathcal{T}}(C') & \text{if } C = \exists r. C' \end{cases}$$

Based on this, the meaning of a concept C can always be determined from the meaning of its unfolded description: $C^{\mathcal{I}} = [u_{\mathcal{T}}(C)]^{\mathcal{I}}$ for all models \mathcal{I} of \mathcal{T} , which means that $C \equiv_{\mathcal{T}} u_{\mathcal{T}}(C)$. From a model-theoretical point of view this is captured by the following proposition (see [Neb91]).

Proposition 2. *Let \mathcal{T} be an acyclic \mathcal{EL} TBox. Any interpretation \mathcal{I} of $\mathbf{N}_{pr}^{\mathcal{T}} \cup \mathbf{N}_R$ can be uniquely extended into a model of \mathcal{T} .*

We now define some notions related to \mathcal{EL} concept descriptions that will be useful for subsequent chapters.

⁴In this paper, we do not consider so-called general concept inclusions (GCIs), which are of the form $C \sqsubseteq D$ for $C, D \in \mathcal{C}_{\mathcal{EL}}$.

Definition 3 (sub-description). Let C be an \mathcal{EL} concept description. The set $\text{sub}(C)$ of sub-descriptions of C is defined in the following way:

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

Note that the number of sub-descriptions $|\text{sub}(C)|$ of a concept C is linear in the size of C . Next, we define the *role depth* of a concept description C .

Definition 4 (role depth). The *role depth* $\text{rd}(C)$ of an \mathcal{EL} concept description C is inductively defined as follows:

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

A concept description is called an *atom* iff it is a concept name or an existential restriction. The set of all \mathcal{EL} atoms is denoted by N_A . Additionally, every \mathcal{EL} concept description is a conjunction $C_1 \sqcap \dots \sqcap C_n$ of atoms. These conjuncts are called the *top-level* atoms of C and the set $\{C_1, \dots, C_n\}$ is denoted as $\text{tp}(C)$.

2.1.1 Characterization of membership in \mathcal{EL}

The definition of the graded membership function *deg* that we will recall in Section 2.3 is based on the representation of concepts and interpretations as graphs, and homomorphisms between these graphs.

Definition 5 (\mathcal{EL} description graph). An \mathcal{EL} description graph is a graph of the form $G = (V_G, E_G, \ell_G)$ where:

- V_G is a set of nodes.
- $E_G \subseteq V_G \times \mathbf{N}_R \times V_G$ is a set of edges labeled by role names,
- $\ell_G : V_G \rightarrow 2^{\mathbf{N}_C}$ is a function that labels nodes with sets of concept names.

In [BKM99], it is shown that every \mathcal{EL} concept description C can be translated into an \mathcal{EL} description tree T_C and vice versa. Moreover, in [Baa03] interpretations \mathcal{I} are translated into \mathcal{EL} description graphs $G_{\mathcal{I}} = (V_{\mathcal{I}}, E_{\mathcal{I}}, \ell_{\mathcal{I}})$ in the following way:

- $V_{\mathcal{I}} = \Delta^{\mathcal{I}}$,

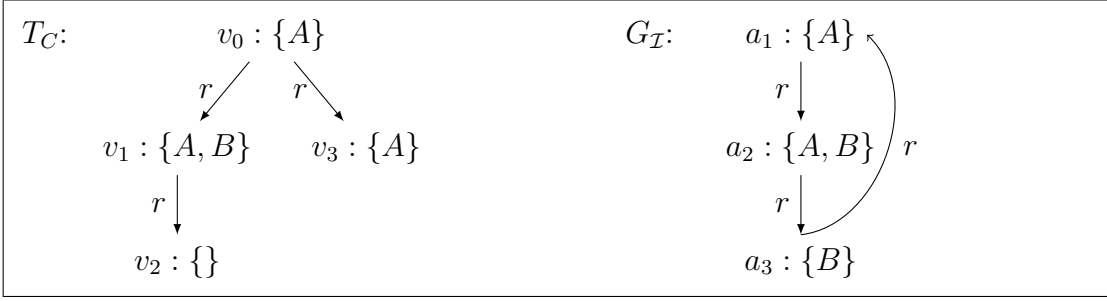


Figure 1: \mathcal{EL} description graphs.

- $E_{\mathcal{I}} = \{(vrw) \mid (v, w) \in r^{\mathcal{I}}\}$,
- $\ell_{\mathcal{I}}(v) = \{A \mid v \in A^{\mathcal{I}}\}$ for all $v \in V_{\mathcal{I}}$.

The following example illustrates the relation between concept descriptions and description trees, and interpretations and description graphs.

Example 6. The \mathcal{EL} concept description

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

yields the \mathcal{EL} description tree T_C depicted on the left-hand side of Figure 1. The description graph on the right-hand side corresponds to the following interpretation:

- $\Delta^{\mathcal{I}} := \{a_1, a_2, a_3\}$,
- $A^{\mathcal{I}} := \{a_1, a_2\}$ and $B^{\mathcal{I}} := \{a_2, a_3\}$,
- $r^{\mathcal{I}} := \{(a_1, a_2), (a_2, a_3), (a_3, a_1)\}$.

Now, homomorphisms between \mathcal{EL} description trees were introduced in [BKM99] to characterize subsumption in \mathcal{EL} : $C \sqsubseteq D$ iff there exists a homomorphism from T_D to T_C mapping the root of T_D to the root of T_C (Thm. 1, [BKM99]). The following definition generalizes such homomorphisms to graphs.

Definition 7. Let $G = (V_G, E_G, \ell_G)$ and $H = (V_H, E_H, \ell_H)$ be two \mathcal{EL} description graphs. A mapping $\varphi : V_G \rightarrow V_H$ is a *homomorphism* from G to H iff the following conditions are satisfied:

1. $\ell_G(v) \subseteq \ell_H(\varphi(v))$ for all $v \in V_G$, and
2. $vrw \in E_G$ implies $\varphi(v)r\varphi(w) \in E_H$.

In Example 6, the mapping φ with $\varphi(v_i) = a_{i+1}$ for $i = 0, 1, 2$ and $\varphi(v_3) = a_2$ is a homomorphism. The proof of the subsumption characterization result in [BKM99] can be easily adapted to characterize *element-hood* in \mathcal{EL} , i.e., whether $d \in C^{\mathcal{I}}$ for some $d \in \Delta^{\mathcal{I}}$.

Theorem 8. 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 φ from T_C to $G_{\mathcal{I}}$ such that $\varphi(v_0) = d$.

2.2 Adding threshold concepts to \mathcal{EL}

In [BBG15], we have extended \mathcal{EL} with a family of concept constructors of the form $C_{\sim t}$, such that C is an \mathcal{EL} concept description, $\sim \in \{<, \leq, >, \geq\}$, and t is a *rational* number in $[0, 1]$. These new constructors can then be combined with the basic \mathcal{EL} concept constructors (2) to form more complex concepts, e.g., $(\exists r.A)_{<1} \sqcap \exists r.(A \sqcap B)_{\geq .8} \sqcap B$. Concepts of the form $C_{\sim t}$ are called *threshold concepts*. The semantics of such concepts is based on a graded membership function m . The idea is that, given an interpretation \mathcal{I} and $d \in \Delta^{\mathcal{I}}$, $m^{\mathcal{I}}(d, C)$ computes a value between 0 and 1 representing the extent to which d belongs to C in \mathcal{I} . For instance, the concept $C_{>.8}$ collects all the individuals that belong to C with degree greater than .8. To indicate which function m is used to obtain the semantics of threshold concepts, we call the extended logic $\tau\mathcal{EL}(m)$. We require such functions m to satisfy the following two properties.

Definition 9. A *graded membership function* m is a family of functions that contains for every interpretation \mathcal{I} a function $m^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \mathcal{C}_{\mathcal{EL}} \rightarrow [0, 1]$ satisfying the following conditions (for $C, D \in \mathcal{C}_{\mathcal{EL}}$):

$$\begin{aligned} M1: & d \in C^{\mathcal{I}} \Leftrightarrow m^{\mathcal{I}}(d, C) = 1 \text{ for all } d \in \Delta^{\mathcal{I}}, \\ M2: & C \equiv D \Leftrightarrow \forall \mathcal{I} \forall d \in \Delta^{\mathcal{I}} : m^{\mathcal{I}}(d, C) = m^{\mathcal{I}}(d, D). \end{aligned}$$

The formal semantics of threshold concepts is then defined in terms of m as follows: $(C_{\sim t})^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid m^{\mathcal{I}}(d, C) \sim t\}$. Taking this into account, $\cdot^{\mathcal{I}}$ is extended in a natural way to interpret complex $\tau\mathcal{EL}(m)$ concept descriptions.

Coming back to Definition 9, on the one hand, property $M2$ expresses the intuition that membership values should not depend on the syntactic form of a concept, but only on its semantics. On the other hand, requiring $M1$ has the following consequences.

Proposition 10. For every \mathcal{EL} concept description C we have $C_{\geq 1} \equiv C$ and $C_{<1} \equiv \neg C$, where the semantics of negation is defined as usual, i.e., $(\neg C)^{\mathcal{I}} := \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$.

The equivalence $C_{<1} \equiv \neg C$ says that negation of \mathcal{EL} concepts is expressible in $\tau\mathcal{EL}(m)$. This does not imply, however, that $\tau\mathcal{EL}(m)$ is closed under negation. Note that nesting of threshold constructors is not allowed. For example, strings like $((\exists r.A)_{<1})_{<1}$ or $(C_{\sim t})_{<1}$ do not constitute well-formed concepts in $\tau\mathcal{EL}(m)$. Thus, negation cannot be nested using these constructors.

Regarding notation, we will sometimes use $C_{=t}$ to abbreviate the concept description $C_{\leq t} \sqcap C_{\geq t}$. Symbols like \widehat{C}, \widehat{D} will be used to refer to $\tau\mathcal{EL}(m)$ concept descriptions. Last, some other notions defined for \mathcal{EL} in the previous section extend naturally to $\tau\mathcal{EL}(m)$ by additionally handling threshold concepts in the following way:

- *role depth*: extends to $\tau\mathcal{EL}(m)$ concept descriptions by defining $\text{rd}(C_{\sim t}) := 0$ for all threshold concepts $C_{\sim t}$,
- *sub-description*: for all threshold concepts $C_{\sim t}$, $\text{sub}(C_{\sim t}) := \{C_{\sim t}\}$.

2.3 The graded membership function deg

In addition to defining the family of DLs $\tau\mathcal{EL}(m)$, in [BBG15] we also define a concrete graded membership function deg and study its induced DL $\tau\mathcal{EL}(deg)$. Since the latter constitutes our main object of study, we shall briefly describe the principal components supporting the definition of deg .

Basically, we use the homomorphism characterization of membership in \mathcal{EL} (Theorem 8) as a starting point. The computation of $deg^{\mathcal{I}}(d, C)$ relies on exploring the search space consisting of all partial mappings from T_C to $G_{\mathcal{I}}$ that map the root of T_C to d and respect the edge structure of T_C . Let us explain the reason for considering such partial mappings using the following example.

Example 11. Figure 2 shows a description tree T_C corresponding to the concept $C := A \sqcap \exists s.(B_1 \sqcap \exists r.B_2 \sqcap \exists r.B_3)$, and the description graph associated to an interpretation \mathcal{I} . Clearly, $d_0 \notin C^{\mathcal{I}}$, and thus there is no homomorphism that maps v_0 to d_0 . Nevertheless, the mappings depicted in the figure (represented by the dashed lines and the dotted ones) provide two different views of how d_0 partially satisfies the properties required by C . The idea is then to calculate to which degree each partial mapping fulfills the homomorphism conditions (see Definition 7), and take the degree of the best one as the membership degree $deg^{\mathcal{I}}(d_0, C)$.

These partial mappings are called *partial tree-to-graph homomorphisms (ptgh)*, and are formally defined as follows.

Definition 12 (Def. 4, [BBG15]). Let $T = (V_t, E_t, \ell_t, v_0)$ and $G = (V_g, E_g, \ell_g)$ be a description tree (with root v_0) and a description graph, respectively. A partial mapping $h : V_t \rightarrow V_g$ is a *partial tree-to-graph homomorphism (ptgh)* from T to G iff the following conditions are satisfied:

1. $\text{dom}(h)$ is a subtree of T with root v_0 , i.e., $v_0 \in \text{dom}(h)$ and if $(v, r, w) \in E_t$ and $w \in \text{dom}(h)$, then $v \in \text{dom}(h)$;

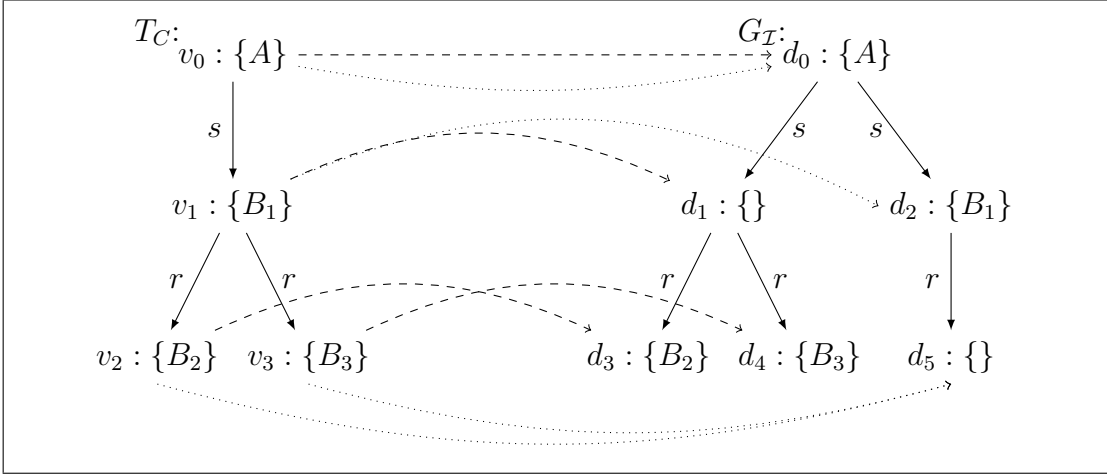


Figure 2: Partial mappings.

2. for all edges $(v, r, w) \in E_t$, $w \in \text{dom}(h)$ implies $(h(v), r, h(w)) \in E_g$.

To measure to which degree a *ptgh* h satisfies the homomorphism conditions, a weighted function $h_w: \text{dom}(h) \rightarrow [0, 1]$ is defined and the value $h_w(v_0)$ considered as the corresponding degree. We use again Figure 2 to sketch how h_w calculates such degrees.

Example 13. Let h denote the mapping represented by the dashed lines and g the other one. To compute $h_w(v_0)$, we basically count the number of properties of v_0 (say ℓ), check how many of those d_0 actually has in \mathcal{I} (say k) and give k/ℓ as the membership degree $h_w(v_0)$. In our example, v_0 has two properties, namely, A and the existence of an s -successor with a certain structure (the node v_1). In particular, the s -successor of d_0 selected by h to match v_1 , does not satisfy all the conditions required by v_1 . Now, instead of assuming that d_0 lacks the second property and setting $h_w(v_0) = 1/2$, $h_w(v_1)$ computes a value that expresses to which degree d_1 satisfies the conditions required by v_1 . This is done by applying the same idea recursively. This procedure stops at nodes of T_C having no successors in $\text{dom}(h)$.

Thus, the real computation is done in a bottom-up manner. First, we have $h_w(v_2) = 1$ and $h_w(v_3) = 1$. Based on these two values and the fact that $d_1 \notin (B_1)^{\mathcal{I}}$, we obtain $h_w(v_1) = 2/3$. Finally, since $d_0 \in A^{\mathcal{I}}$, we get $h_w(v_0) = (1 + h_w(v_1))/2 = 5/6$. Concerning the mapping g , the reader can verify that $g_w(v_0) < h_w(v_0)$, and thus *deg* sees h as a better approximation for membership in C .

The intuition given in the previous example is formally expressed in the following definition.

Definition 14 (Def. 5, [BBG15]). Let T be a finite \mathcal{EL} description tree, G an \mathcal{EL} description graph and $h: V_T \rightarrow V_G$ a *ptgh* from T to G . We define

the *weighted* homomorphism induced by h from T to G as a recursive function $h_w : \text{dom}(h) \rightarrow [0..1]$ in the following way:

$$h_w(v) := \begin{cases} 1 & \text{if } |\ell_T(v)| + k^*(v) = 0 \\ \frac{|\ell_T(v) \cap \ell_G(h(v))| + \sum_{1 \leq i \leq k} h_w(v_i)}{|\ell_T(v)| + k^*(v)} & \text{otherwise.} \end{cases}$$

The elements used to define h_w have the following meaning. For a given $v \in \text{dom}(h)$, $k^*(v)$ denotes the number of successors of v in T , and v_1, \dots, v_k with $0 \leq k \leq k^*(v)$ are the children of v in T such that $v_i \in \text{dom}(h)$.

Based on these ideas, we now defined the graded membership function *deg*. However, in order to satisfy property *M2*, all concept descriptions C are transformed into an appropriate *reduced form* C^r before actually applying the computations sketched above. This reduced form, which was introduced in [Küs01], removes redundancies from concepts, and has the property that $C \equiv D$ iff the description trees of C^r and D^r are isomorphic.

Definition 15 (Def. 6, [BBG15]). Let \mathcal{I} be an interpretation, $d \in \Delta^{\mathcal{I}}$ and C an \mathcal{EL} concept with reduced form C^r . Moreover, let $\mathcal{H}(T_{C^r}, G_{\mathcal{I}}, d)$ be the set of all *ptghs* h from T_{C^r} to $G_{\mathcal{I}}$ with $h(v_0) = d$. Then,

$$\text{deg}^{\mathcal{I}}(d, C) := \max\{q \mid h_w(v_0) = q \text{ and } h \in \mathcal{H}(T_{C^r}, G_{\mathcal{I}}, d)\}$$

We have shown in [BBG15] that the maximum in the above expression always exists. This implies that the function *deg* is *well-defined*. In addition, we could show that the properties *M1* and *M2* are satisfied. Regarding the induced DL $\tau\mathcal{EL}(\text{deg})$, we have investigated the computational complexity of the standard reasoning problems *satisfiability*, *subsumption*, *ABox consistency* and *instance checking*. In particular, the subsumption and the satisfiability problems are tackled by establishing the following polynomial model property for the satisfiability of concepts of the form $\widehat{C} \sqcap \neg \widehat{D}$ for $\tau\mathcal{EL}(\text{deg})$ concepts \widehat{C}, \widehat{D} . Note that this is equivalent to the non-subsumption problem and satisfiability is a special case.

Lemma 16 (Lem. 5, [BBG15]). *Let \widehat{C} and \widehat{D} be $\tau\mathcal{EL}(\text{deg})$ concepts of sizes $\mathfrak{s}(\widehat{C})$ and $\mathfrak{s}(\widehat{D})$ ⁵. If $\widehat{C} \sqcap \neg \widehat{D}$ is satisfiable, then there exists an interpretation \mathcal{I} such that $\widehat{C}^{\mathcal{I}} \setminus \widehat{D}^{\mathcal{I}} \neq \emptyset$ and $|\Delta^{\mathcal{I}}| \leq \mathfrak{s}(\widehat{C}) \times \mathfrak{s}(\widehat{D})$.*

A analogous property has been also proved for consistent ABoxes of the form $\mathcal{A} \cup \{\neg \widehat{C}(a)\}$, thus yielding a bounded model property for non-instance ($\mathcal{A} \not\models \widehat{C}(a)$). Unfortunately, in this case the bound on the model's size has the size of the concept \widehat{C} in the exponent. Nevertheless, since *consistency* is a particular

⁵The size $\mathfrak{s}(\widehat{C})$ of a $\tau\mathcal{EL}(\text{deg})$ concept description \widehat{C} is the number of occurrences of symbols needed to write \widehat{C} , where the numbers used to write threshold values are encoded in binary.

case where \widehat{C} is not present, we have a polynomial model property for ABox consistency. In addition, checking whether a finite interpretation \mathcal{I} satisfies a $\tau\mathcal{EL}(deg)$ concept/ABox can be done in polynomial time. Overall, we can thus employ a standard *guess-and-check* NP-algorithm to decide satisfiability, non-subsumption, and ABox consistency. For non-instance, this algorithm is only in NP if we consider *data complexity* as defined in [DLNS94].

Theorem 17 (Th. 5 and Th. 6, [BBG15]). *In $\tau\mathcal{EL}(deg)$, satisfiability and consistency are NP-complete, whereas subsumption and instance checking (w.r.t. data complexity) are coNP-complete.*

3 Acyclic TBoxes for $\tau\mathcal{EL}(m)$

We now turn to introducing acyclic TBoxes for the whole family of DLs $\tau\mathcal{EL}(m)$, and hence also for $\tau\mathcal{EL}(deg)$. As with acyclic TBoxes in \mathcal{EL} , the purpose is to introduce abbreviations for composite $\tau\mathcal{EL}(m)$ concept descriptions. For instance, the \mathcal{EL} concept definition $E \doteq \exists r.A \sqcap \exists r.B$ can be used to abbreviate the threshold concept $(\exists r.A \sqcap \exists r.B)_{\geq 1/2}$ as $E_{\geq 1/2}$. On top of this, we can then also introduce the abbreviation β for $E_{\geq 1/2}$ and use this abbreviation in other concept definitions, as done in the following TBox:

$$\left\{ \begin{array}{l} \alpha \doteq \exists s.A \sqcap \exists r.\beta \\ \beta \doteq E_{\geq 1/2} \\ E \doteq \exists r.A \sqcap \exists r.B \end{array} \right\} \quad (3)$$

Overall, the concept name α then abbreviates the $\tau\mathcal{EL}(m)$ concept description $\exists s.A \sqcap \exists r.(\exists r.A \sqcap \exists r.B)_{\geq 1/2}$, which can be obtained from α by unfolding.

However, we cannot use arbitrary acyclic sets of $\tau\mathcal{EL}(m)$ concept definitions. For example, suppose that α is now defined in the TBox (3) as $\alpha \doteq \exists s.A \sqcap \exists r.(\beta_{>.8})$ instead. Even though the right-hand side of this definition is a syntactically well-formed $\tau\mathcal{EL}(m)$ concept, unfolding α w.r.t. this new TBox yields

$$\exists s.A \sqcap \exists r.(((\exists r.A \sqcap \exists r.B)_{\geq 1/2})_{>.8}), \quad (4)$$

which is not a well-formed $\tau\mathcal{EL}(m)$ concept description since threshold operators are nested. The following definition is designed to avoid this problem.

Definition 18. An *acyclic $\tau\mathcal{EL}(m)$ TBox* $\widehat{\mathcal{T}}$ is a pair $(\mathcal{T}_\tau, \mathcal{T})$, where \mathcal{T} is an acyclic \mathcal{EL} TBox and \mathcal{T}_τ is a set of concept definitions of the form $\alpha \doteq \widehat{C}_\alpha$ satisfying the following conditions:

- \widehat{C}_α is a $\tau\mathcal{EL}(m)$ concept description,
- α does not *depend* on itself and it does not occur in \mathcal{T} ,
- for all threshold concepts $C_{\sim t}$ occurring in \widehat{C}_α , no defined concept of \mathcal{T}_τ occurs in C .

The TBox (3) can be seen as an acyclic $\tau\mathcal{EL}(m)$ TBox where the first two definitions belong to \mathcal{T}_τ and the last one to \mathcal{T} .

Given an acyclic $\tau\mathcal{EL}(m)$ TBox $\widehat{\mathcal{T}} = (\mathcal{T}_\tau, \mathcal{T})$, we define the set $\mathbf{N}_d^{\widehat{\mathcal{T}}}$ of defined concepts in $\widehat{\mathcal{T}}$ as the union $\mathbf{N}_d^{\mathcal{T}_\tau} \cup \mathbf{N}_d^{\mathcal{T}}$, where $\mathbf{N}_d^{\mathcal{T}_\tau}$ is the set of defined concepts in \mathcal{T}_τ . We denote the set $\mathbf{N}_C \setminus \mathbf{N}_d^{\widehat{\mathcal{T}}}$ as $\mathbf{N}_{pr}^{\widehat{\mathcal{T}}}$. The notion of unfolding is extended to acyclic $\tau\mathcal{EL}(m)$ TBoxes by considering the following two additional rules:

$$\begin{aligned} u_{\widehat{\mathcal{T}}}(\alpha) &:= u_{\widehat{\mathcal{T}}}(\widehat{C}_\alpha), \text{ for all } \alpha \doteq \widehat{C}_\alpha \in \mathcal{T}_\tau \\ u_{\widehat{\mathcal{T}}}(C_{\sim t}) &:= [u_{\mathcal{T}}(C)]_{\sim t} \end{aligned}$$

Then, it is easy to see that the restrictions imposed in the previous definition guarantee that $\alpha \in \mathbf{N}_d^{\mathcal{T}_\tau}$ always unfolds into a well-formed $\tau\mathcal{EL}(m)$ concept description $u_{\widehat{\mathcal{T}}}(\alpha)$, whereas $E \in \mathbf{N}_d^{\mathcal{T}}$ unfolds into an \mathcal{EL} concept. Regarding arbitrary $\tau\mathcal{EL}(m)$ concept descriptions \widehat{C} , we say that \widehat{C} is *correctly defined* w.r.t $\widehat{\mathcal{T}}$ if the pair $(\mathcal{T}_\tau \cup \{\alpha \doteq \widehat{C}\}, \mathcal{T})$ is still an acyclic $\tau\mathcal{EL}(m)$ TBox, where α is a *fresh* concept name not occurring in $\widehat{\mathcal{T}}$ and \widehat{C} .

We are now ready to fix the semantics for this kind of TBoxes. To start with, we say that an interpretation \mathcal{I} is a model of \mathcal{T}_τ (and write $\mathcal{I} \models \mathcal{T}_\tau$) iff $\alpha^{\mathcal{I}} = (\widehat{C}_\alpha)^{\mathcal{I}}$ in $\tau\mathcal{EL}(m)$ for all $\alpha \doteq \widehat{C}_\alpha \in \mathcal{T}_\tau$. Then, \mathcal{I} satisfies $\widehat{\mathcal{T}}$ iff $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \models \mathcal{T}_\tau$. The corresponding subsumption and equivalence relations $\sqsubseteq_{\widehat{\mathcal{T}}}$ and $\equiv_{\widehat{\mathcal{T}}}$ on correctly defined $\tau\mathcal{EL}(m)$ concepts are defined w.r.t. the set of models of $\widehat{\mathcal{T}}$. The next step is to ensure that defined concepts α and their unfolded counterparts $u_{\widehat{\mathcal{T}}}(\alpha)$ have the same meaning in all models of $\widehat{\mathcal{T}}$, i.e.,

$$\alpha \equiv_{\widehat{\mathcal{T}}} u_{\widehat{\mathcal{T}}}(\alpha). \quad (5)$$

Since this equivalence holds for \mathcal{EL} , the only constructor that might lead to a problem is the threshold constructor. More precisely, given a threshold concept $C_{\sim t}$ with $C \in \mathcal{C}_{\mathcal{EL}}$, for all models of $\widehat{\mathcal{T}}$ the following must hold:

$$(C_{\sim t})^{\mathcal{I}} = ((u_{\mathcal{T}}(C))_{\sim t})^{\mathcal{I}}. \quad (6)$$

Thus, we must turn our attention to the graded membership function m since m is providing the semantics for such concepts $C_{\sim t}$. In principle, the graded membership function m is defined on C since C is an \mathcal{EL} concept description. However, this function (e.g., deg) is agnostic of the TBox and thus treats defined and primitive concepts alike: they are just concept names for m . In order to satisfy (6), the function needs to be aware of the TBox. Let us illustrate this using the graded membership function deg :

Example 19. Let $\widehat{\mathcal{T}} = (\mathcal{T}_\tau, \mathcal{T})$ be the $\tau\mathcal{EL}(m)$ acyclic TBox corresponding to the definitions in (3). In addition, let \mathcal{I} be an interpretation such that $\Delta^{\mathcal{I}} = \{d_0, d_r, d_s\}$, $A^{\mathcal{I}} = \{d_s\}$, $B^{\mathcal{I}} = \{d_r\}$, and $r^{\mathcal{I}} = \{(d_0, d_0), (d_0, d_r)\}$, $s^{\mathcal{I}} = \{(d_0, d_s)\}$.

When trying to extend \mathcal{I} to a model of $\widehat{\mathcal{T}}$, we first note that we have $(\exists r.A \sqcap \exists r.B)^{\mathcal{I}} = \emptyset$, and hence $E^{\mathcal{I}}$ must be interpreted as the empty set. Then, since E is treated as a concept name by deg , this means that $deg^{\mathcal{I}}(d, E) = 0$ for all $d \in \Delta^{\mathcal{I}}$. Therefore, $(E_{\geq 1/2})^{\mathcal{I}} = \emptyset$, and consequently we must define $\beta^{\mathcal{I}} := \alpha^{\mathcal{I}} := \emptyset$. To see that (6) fails to hold, one can observe that in contrast to $deg^{\mathcal{I}}(d_0, E) = 0$, for d_0 we obtain $deg^{\mathcal{I}}(d_0, \exists r.A \sqcap \exists r.B) = 1/2$ (recall the ideas defining deg). This means that $(E_{\geq 1/2})^{\mathcal{I}} \neq ((u_{\mathcal{T}}(E))_{\geq 1/2})^{\mathcal{I}}$. Obviously, the problem propagates up to the more general requirement in (5). First, $d_0 \notin \beta^{\mathcal{I}}$ but $d_0 \in (u_{\widehat{\mathcal{T}}}(\beta))^{\mathcal{I}}$. Moreover, it is easy to check that $d_0 \in (u_{\widehat{\mathcal{T}}}(\alpha))^{\mathcal{I}}$, and thus $\alpha \not\equiv_{\widehat{\mathcal{T}}} u_{\widehat{\mathcal{T}}}(\alpha)$.

To avoid the problem demonstrated by this example, the graded membership function m needs to take into account the definitions in \mathcal{T} . This means that \mathcal{T}

must be a parameter of this function. Furthermore, to ensure that property (6) is satisfied, the membership degrees for an \mathcal{EL} concept description C should be the same as for its unfolding $u_{\mathcal{T}}(C)$. Taking this into account, we extend a given graded membership function m such that it takes concept definitions in acyclic \mathcal{EL} TBoxes into account.

Definition 20. For all graded membership functions m (in the sense of Definition 9), the extension of m computing membership degrees w.r.t. acyclic \mathcal{EL} TBoxes is a family of functions containing for every interpretation \mathcal{I} a function $\widehat{m}^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \mathcal{C}_{\mathcal{EL}} \times \mathfrak{T}(\mathcal{I}) \rightarrow [0, 1]$ satisfying

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

where $\mathfrak{T}(\mathcal{I})$ is the set of all acyclic \mathcal{EL} TBoxes \mathcal{T} such that $\mathcal{I} \models \mathcal{T}$.

Clearly, well-definedness of m and acyclicity of \mathcal{T} imply well-definedness of \widehat{m} . For the sake of simplicity, we will from now on use m both to denote the original graded membership function and its extension \widehat{m} .

The use of unfolding in the above definition ensures that, for all interpretations \mathcal{I} and $d \in \Delta^{\mathcal{I}}$, we have $d \in (C_{\sim t})^{\mathcal{I}}$ iff $d \in ((u_{\mathcal{T}}(C))_{\sim t})^{\mathcal{I}}$. Consequently, (6) always holds, as does (5). Finally, it is easy to see that the analogon of Proposition 2 is also valid for acyclic $\tau\mathcal{EL}(m)$ TBoxes.

Proposition 21. *Let $\widehat{\mathcal{T}}$ be an acyclic $\tau\mathcal{EL}(m)$ TBox. Any interpretation \mathcal{I} of $\mathbb{N}_{pr}^{\widehat{\mathcal{T}}} \cup \mathbb{N}_{\mathbb{R}}$ can be uniquely extended into a model of $\widehat{\mathcal{T}}$.*

The following lemma is an easy consequence of Definition 20. It shows that graded membership functions constructed in such a way satisfy a generalization of the conditions stated in Definition 9.

Lemma 22. Let m be a graded membership function as in Definition 20. Then, for all acyclic \mathcal{EL} TBoxes \mathcal{T} , we have:

$$\begin{aligned} M1^{\mathcal{T}}: d \in C^{\mathcal{I}} &\Leftrightarrow m^{\mathcal{I}}(d, C, \mathcal{T}) = 1 \quad \text{for all } \mathcal{I} \models \mathcal{T} \text{ and } d \in \Delta^{\mathcal{I}} \\ M2^{\mathcal{T}}: C \equiv_{\mathcal{T}} D &\Leftrightarrow \forall \mathcal{I} \models \mathcal{T} \forall d \in \Delta^{\mathcal{I}}: m^{\mathcal{I}}(d, C, \mathcal{T}) = m^{\mathcal{I}}(d, D, \mathcal{T}) \end{aligned}$$

where C and D are \mathcal{EL} concept descriptions.

Proof. Being m a graded membership function in the sense of Definition 9, it satisfies $M1$ and $M2$. Hence, since $C \equiv_{\mathcal{T}} u_{\mathcal{T}}(C)$, the definition of m with respect to \mathcal{T} satisfies $M1^{\mathcal{T}}$. Moreover, $C \equiv_{\mathcal{T}} D$ implies that $u_{\mathcal{T}}(C) \equiv u_{\mathcal{T}}(D)$. From this it is easy to verify that m also satisfies $M2^{\mathcal{T}}$. \square

To sum up, we have introduced a notion of acyclic TBoxes for $\tau\mathcal{EL}(m)$ such that unfolding still works from a syntactic point of view, i.e., the unfolding of a

defined concept is a syntactically correct $\tau\mathcal{EL}(m)$ concept description. To ensure that unfolding is also correct from the semantic point of view (i.e., (5) holds), we had to extend m such that it takes the \mathcal{EL} part of the given acyclic TBox into account. In the following, we consider the case where $m = deg$ and show that the presence of acyclic $\tau\mathcal{EL}(deg)$ TBoxes increases the complexity of reasoning.

4 Reasoning with acyclic $\tau\mathcal{EL}(deg)$ TBoxes

We will not only investigate the satisfiability and the subsumption problem, but also consistency and instance. In the presence of an acyclic $\tau\mathcal{EL}(deg)$ TBox, the concepts occurring in the ABox need to be correctly defined w.r.t. this TBox. An *acyclic $\tau\mathcal{EL}(deg)$ knowledge base* is a pair $\mathcal{K} = (\widehat{\mathcal{T}}, \mathcal{A})$ where $\widehat{\mathcal{T}}$ is an acyclic $\tau\mathcal{EL}(deg)$ TBox, and \mathcal{A} is a finite set of assertions $\widehat{C}(a)$ or $r(a, b)$, where \widehat{C} is correctly defined w.r.t. $\widehat{\mathcal{T}}$.

The size $s(\widehat{\mathcal{T}})$ of an acyclic $\tau\mathcal{EL}(deg)$ TBox $\widehat{\mathcal{T}} = (\mathcal{T}_\tau, \mathcal{T})$ corresponds to the following expression:

$$s(\widehat{\mathcal{T}}) := |\mathbf{N}_d^{\widehat{\mathcal{T}}}| + \sum_{\alpha \doteq \widehat{C}_\alpha \in \mathcal{T}_\tau \cup \mathcal{T}} s(\widehat{C}_\alpha)$$

Furthermore, we define the size $s(\mathcal{A})$ of an ABox \mathcal{A} as:

$$s(\mathcal{A}) := \sum_{\substack{\widehat{C}(a) \in \mathcal{A} \\ a \in \text{Ind}(\mathcal{A})}} s(\widehat{C}) + \sum_{\substack{r(a, b) \in \mathcal{A} \\ a, b \in \text{Ind}(\mathcal{A})}} 1$$

Finally, the size $s(\mathcal{K})$ of a KB $\mathcal{K} = (\widehat{\mathcal{T}}, \mathcal{A})$ is simply $s(\widehat{\mathcal{T}}) + s(\mathcal{A})$.

Proposition 21 together with (5) tell us that reasoning w.r.t. acyclic $\tau\mathcal{EL}(deg)$ TBoxes can be reduced to reasoning in the empty terminology, through unfolding. However, as shown in [Neb90] for the DL \mathcal{FL}_0 , unfolding may produce concept descriptions of exponential size. The following example shows the corresponding version for \mathcal{EL} .

Example 23. The TBox \mathcal{T}_n is inductively defined as follows ($n \geq 0$):

$$\begin{aligned} \mathcal{T}_0 &:= \{\alpha_0 \doteq \top\} \\ \mathcal{T}_1 &:= \mathcal{T}_0 \cup \{\alpha_1 \doteq \exists r.\alpha_0 \sqcap \exists s.\alpha_0\} \\ &\dots \\ \mathcal{T}_n &:= \mathcal{T}_{n-1} \cup \{\alpha_n \doteq \exists r.\alpha_{n-1} \sqcap \exists s.\alpha_{n-1}\} \end{aligned}$$

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

Thus, by applying unfolding and then using the known NP decision procedures for satisfiability/non-subsumption in $\tau\mathcal{EL}(deg)$ [BBG15], we obtain an NExpTime algorithm to decide the same problems w.r.t. acyclic $\tau\mathcal{EL}(deg)$ TBoxes. The natural question to ask is thus: can we do better than NExpTime? We will show that this is indeed the case by providing a PSpace upper bound. At the moment, we do not have a matching lower bound. However, we can show that (unless $\text{NP} = \Pi_2^P$) the complexity of reasoning w.r.t. acyclic $\tau\mathcal{EL}(deg)$ TBoxes is higher than of reasoning in $\tau\mathcal{EL}(deg)$ without a TBox. We start with showing the lower bounds.

4.1 Lower bounds

We reduce the problem $\forall\exists\text{3SAT}$ to concept satisfiability with respect to acyclic $\tau\mathcal{EL}(\text{deg})$ TBoxes. This problem is well-known to be complete for the class Π_2^P (see [Sto76], Section 4).

Definition 24 ($\forall\exists\text{3SAT}$). Let $u = \{u_1, \dots, u_n\}$, $v = \{v_1, \dots, v_\ell\}$ be two disjoint sets of propositional variables, and $\varphi(u, v)$ a 3CNF formula defined over $u \cup v$, i.e., a finite set of propositional clauses $\mathcal{C} = \{c_1, \dots, c_q\}$ such that each c_k is a set of three literals $\{\gamma_1^k, \gamma_2^k, \gamma_3^k\}$ over $u \cup v$. A formula $\forall u \exists v. \varphi(u, v)$ is *valid* iff for all truth assignments \mathbf{t} of the variables in u there is an extension of \mathbf{t} for the variables in v satisfying $\varphi(u, v)$. $\forall\exists\text{3SAT}$ is then the problem of deciding whether a formula $\forall u \exists v. \varphi(u, v)$ is valid or not.

The idea for the reduction goes as follows. Each formula $\forall u \exists v. \varphi(u, v)$ is translated into an acyclic $\tau\mathcal{EL}(\text{deg})$ TBox $\widehat{\mathcal{T}}_n^\varphi$ containing a defined concept α_n such that: $\forall u \exists v. \varphi(u, v)$ is valid iff α_n is satisfiable in $\widehat{\mathcal{T}}_n^\varphi$ (n is the number of variables in u).

The first step consists of encoding $\varphi(u, v)$ into a $\tau\mathcal{EL}(\text{deg})$ concept description \widehat{C}_φ . Literals defined over u and v are represented by concept names A_i, \bar{A}_i ($1 \leq i \leq n$) and B_j, \bar{B}_j ($1 \leq j \leq \ell$), respectively, according to the following mapping:

$$\eta(u_i) := A_i, \quad \eta(\neg u_i) := \bar{A}_i, \quad \eta(v_j) := B_j, \quad \eta(\neg v_j) := \bar{B}_j \quad (7)$$

Based on η , each clause c_k is represented by the \mathcal{EL} concept description:

$$D_k := \eta(\gamma_1^k) \sqcap \eta(\gamma_2^k) \sqcap \eta(\gamma_3^k) \quad (8)$$

The satisfiability of c_k could then be simulated through the threshold concept $(D_k)_{\geq 1/3}$. In fact, by the definition of *deg*, an individual d belongs to $(D_k)_{\geq 1/3}$ iff it belongs to at least one concept name in $\{\eta(\gamma_1^k), \eta(\gamma_2^k), \eta(\gamma_3^k)\}$. Therefore, the $\tau\mathcal{EL}(\text{deg})$ concept $(D_1)_{\geq 1/3} \sqcap \dots \sqcap (D_q)_{\geq 1/3}$ appears as a plausible choice to capture the satisfiability status of $\varphi(u, v)$. Nevertheless, pairs of concepts (A_i, \bar{A}_i) and (B_j, \bar{B}_j) need to be complementary, since they are meant to play the role of a literal u_i (v_j) and its negation. To this end, we define a TBox $\mathcal{T}_{n,\ell}^c$ as follows:

$$\mathcal{T}_{n,\ell}^c := \bigcup_{i=1}^n \{F_i \doteq A_i \sqcap \bar{A}_i\} \cup \bigcup_{j=1}^{\ell} \{G_j \doteq B_j \sqcap \bar{B}_j\}$$

Then, $(F_i)_{=1/2}$ collects the elements that are instances of exactly one concept in $\{A_i, \bar{A}_i\}$ (similarly for $(G_j)_{=1/2}$ and $\{B_j, \bar{B}_j\}$). Putting all these pieces together, \widehat{C}_φ is defined as follows:

$$\widehat{C}_\varphi := \prod_{k=1}^q (D_k)_{\geq 1/3} \sqcap \prod_{i=1}^n (F_i)_{=1/2} \sqcap \prod_{j=1}^{\ell} (G_j)_{=1/2}$$

The following result follows from the construction of \widehat{C}_φ .

Lemma 25. $\varphi(u, v)$ is satisfiable iff \widehat{C}_φ is satisfiable w.r.t. $\mathcal{T}_{n,\ell}^c$.

Proof. (\Rightarrow) Assume that $\varphi(u, v)$ is satisfiable. Then, there exists a truth assignment \mathbf{t} for $u \cup v$ satisfying $\varphi(u, v)$. Using \mathbf{t} we construct a single-pointed interpretation $\mathcal{I}_0 = (\{d_0\}, \cdot^{\mathcal{I}_0})$, where $\cdot^{\mathcal{I}_0}$ interprets the primitive concept names A_i, \bar{A}_i ($1 \leq i \leq n$) and B_j, \bar{B}_j ($1 \leq j \leq \ell$) as follows:

- $d_0 \in (A_i)^{\mathcal{I}_0}$ iff $\mathbf{t}(u_i) = \text{true}$ and $d_0 \in (\bar{A}_i)^{\mathcal{I}_0}$ iff $\mathbf{t}(u_i) = \text{false}$
- $d_0 \in (B_j)^{\mathcal{I}_0}$ iff $\mathbf{t}(v_j) = \text{true}$ and $d_0 \in (\bar{B}_j)^{\mathcal{I}_0}$ iff $\mathbf{t}(v_j) = \text{false}$

Obviously, since \mathbf{t} is a truth assignment it follows that:

$$d_0 \in (A_i)^{\mathcal{I}_0} \text{ iff } d_0 \notin (\bar{A}_i)^{\mathcal{I}_0}, \text{ for all } 1 \leq i \leq n \quad (9)$$

and,

$$d_0 \in (B_j)^{\mathcal{I}_0} \text{ iff } d_0 \notin (\bar{B}_j)^{\mathcal{I}_0}, \text{ for all } 1 \leq j \leq \ell \quad (10)$$

Consequently, its extension into a model of $\mathcal{T}_{n,\ell}^c$ is such that $(F_i)^{\mathcal{I}_0} = (G_j)^{\mathcal{I}_0} = \emptyset$. We now prove that $d_0 \in (\widehat{C}_\varphi)^{\mathcal{I}_0}$. As a direct consequence of (9), (10) and the definition of deg w.r.t. TBoxes, we obtain:

$$d_0 \in \left(\prod_{i=1}^n (F_i)_{=1/2} \sqcap \prod_{j=1}^{\ell} (G_j)_{=1/2} \right)^{\mathcal{I}_0}$$

Moreover, let D_k be an arbitrary concept as constructed in 8, and c_k its corresponding clause in $\varphi(u, v)$. Since \mathbf{t} satisfies $\varphi(u, v)$, this means that there exists a literal γ in c_k such that $\mathbf{t}(\gamma) = \text{true}$. Suppose that γ is of the form u_i . Then, $\eta(\gamma) = A_i$ and by construction of \mathcal{I}_0 we know that $d_0 \in (\eta(\gamma))^{\mathcal{I}_0}$. Conversely, if γ is of the form $\neg u_i$, we then have $\mathbf{t}(u_i) = \text{false}$ and $\eta(\gamma) = \bar{A}_i$. Again, by construction of \mathcal{I}_0 this means that $d_0 \in (\eta(\gamma))^{\mathcal{I}_0}$. The same applies if γ is of the form v_j or $\neg v_j$. Therefore, by the definition of D_k and the definition of deg w.r.t. \mathcal{T} , we further have $d_0 \in ((D_k)_{\geq 1/3})^{\mathcal{I}_0}$. Thus,

$$d_0 \in \left(\prod_{k=1}^q (D_k)_{\geq 1/3} \right)^{\mathcal{I}_0}$$

Overall, we just have shown that $d_0 \in (\widehat{C}_\varphi)^{\mathcal{I}_0}$.

(\Leftarrow) Conversely, assume that \widehat{C}_φ is satisfiable with respect to $\mathcal{T}_{n,\ell}^c$. This means that there exists a model \mathcal{I} of $\mathcal{T}_{n,\ell}^c$ and $d \in \Delta^{\mathcal{I}}$ such that $d \in (\widehat{C}_\varphi)^{\mathcal{I}}$. We define the assignment \mathbf{t}_d for $u \cup v$ as follows:

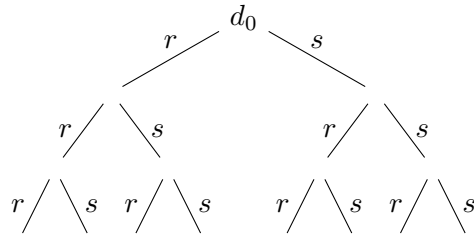
$$\begin{aligned} \mathbf{t}_d(u_i) &= \text{true} \text{ iff } d \in (A_i)^{\mathcal{I}} \quad (1 \leq i \leq n) \\ \mathbf{t}_d(v_j) &= \text{true} \text{ iff } d \in (B_j)^{\mathcal{I}} \quad (1 \leq j \leq \ell) \end{aligned}$$

Let us now show that \mathbf{t}_d satisfies $\varphi(u, v)$. We take an arbitrary clause c_k of $\varphi(u, v)$ and its corresponding concept description D_k (as defined in (8)). Since $d \in (\widehat{C}_\varphi)^{\mathcal{I}}$, this means that $d \in ((D_k)_{\geq 1/3})^{\mathcal{I}}$. Therefore, there exists a literal γ in c_k such that $d \in (\eta(\gamma))^{\mathcal{I}}$. If $\eta(\gamma)$ corresponds to A_i , then by (7) and the construction of \mathbf{t}_d we have that $\gamma = u_i$ and $\mathbf{t}_d(\gamma) = \text{true}$. In case $\eta(\gamma) = \bar{A}_i$, $d \in ((F_i)_{=1/2})^{\mathcal{I}}$ implies that $d \notin (A_i)^{\mathcal{I}}$. This means that $\gamma = \neg u_i$ and $\mathbf{t}_d(u_i) = \text{false}$. One can see that in both cases \mathbf{t}_d satisfies c_k . A similar reasoning can be used with the other two possible forms of $\eta(\gamma)$, namely, B_j and \bar{B}_j for all $1 \leq j \leq \ell$. Finally, since c_k has been chosen arbitrarily, we can conclude that \mathbf{t}_d satisfies $\varphi(u, v)$. Thus, $\varphi(u, v)$ is a satisfiable propositional formula. \square

Obviously, this is not enough to achieve our main goal, since $\forall\exists\text{3SAT}$ asks for the satisfiability of $\varphi(u, v)$ in all truth assignments of u . To mimic this universal quantification, we extend the TBox \mathcal{T}_n (from Example 23) into $\widehat{\mathcal{T}}_n^\varphi$ in such a way that for all models \mathcal{I} of $\widehat{\mathcal{T}}_n^\varphi$, $(\alpha_n)^{\mathcal{I}} \neq \emptyset$ iff for all $X \subseteq u$ there exists $d_X \in \Delta^{\mathcal{I}}$ satisfying:

$$d_X \in (\widehat{C}_\varphi)^{\mathcal{I}} \text{ and for all } i, 1 \leq i \leq n : d_X \in (A_i)^{\mathcal{I}} \text{ iff } u_i \in X \quad (11)$$

Let T_n be the \mathcal{EL} description tree corresponding to the concept description $u_{\mathcal{T}_n}(\alpha_n)$. We denote the interpretation induced by T_n as \mathcal{I}_n . For instance, for $n=3$, the interpretation \mathcal{I}_3 has the following shape⁶:



It is easy to see that the extension of \mathcal{I}_n into a model of \mathcal{T}_n is such that $d_0 \in (\alpha_n)^{\mathcal{I}_n}$. Moreover, a *one-to-one* correspondence can be established between the set of

⁶For an arbitrary n , \mathcal{I}_n has the shape of the full binary tree of depth n , where edges leading to left children are labeled with r and the ones leading to right children are labeled with s .

leaves and the words in $\{r, s\}^n$ as follows. For all words $x = x_1 \dots x_n$ in $\{r, s\}^n$, the corresponding leaf d_x is the one reached from d_0 by the path $d_0 x_1 d_1 \dots x_n d_x$. Then, the desired group of elements satisfying (11) would also exist if: for each word $x \in \{r, s\}^n$ there is at least one path $d_0 x_1 \dots x_n d_x$ such that:

$$d_x \in (\widehat{C}_\varphi)^{\mathcal{I}^n} \quad \text{and} \quad d_x \in (A_i)^{\mathcal{I}^n} \text{ iff } x_i = r \quad (1 \leq i \leq n) \quad (12)$$

The following proposition is an easy consequence of the definition of \mathcal{T}_n .

Proposition 26. *Let \mathcal{I} be a model of \mathcal{T}_n and $d \in \Delta^{\mathcal{I}}$. For all $0 \leq i \leq n$: if $d \in (\alpha_i)^{\mathcal{I}}$, then for each word $x \in \{r, s\}^i$ there exists a path $d x_1 d_1 \dots x_i d_i$ in $G_{\mathcal{I}}$ such that $d_j \in (\alpha_{i-j})^{\mathcal{I}}$ for all $1 \leq j \leq i$.*

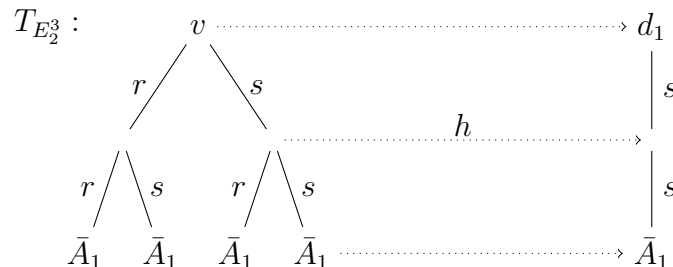
This means that the structure of \mathcal{T}_n certainly guarantees that every model satisfying α_n contains a path $d_0 x_1 \dots x_n d_x$ from a distinguished element d_0 , for all $x \in \{r, s\}^n$. Moreover, the domain elements in such a path satisfy $d_0 \in (\alpha_n)^{\mathcal{I}^n}$, $d_1 \in (\alpha_{n-1})^{\mathcal{I}^n}$, \dots , $d_x \in (\alpha_0)^{\mathcal{I}^n}$. Hence, the first part of (12) can be satisfied by modifying the definition of α_0 in \mathcal{T}_n as:

$$\alpha_0 \doteq \widehat{C}_\varphi \quad (13)$$

Of course, all such models need not have the same shape as \mathcal{I}_n . Hence, to also fulfill the second part of (12), one needs to express within the logic the correct propagation of A_1, \dots, A_n along each path. For simplicity, we explain the intuition of how to do this for $n = 3$. Consider x_1 and A_1 . A solution could be to redefine α_3 as $\alpha_3 \doteq \exists r.(\alpha_2 \sqcap \beta_2^r) \sqcap \exists s.(\alpha_2 \sqcap \beta_2^s)$, where:

$$\beta_2^r := \prod_{x_2, x_3 \in \{r, s\}} \forall x_2 x_3. \neg \bar{A}_1 \quad \beta_2^s := \prod_{x_2, x_3 \in \{r, s\}} \forall x_2 x_3. \neg A_1 \quad (14)$$

The definition of β_2^r implies that if $d_0 \in (\alpha_3)^{\mathcal{I}^3}$, then all paths starting at d_0 following a word x of the form $r.\{r, s\}^2$ must end up at an element d_x such that $d_x \notin (\bar{A}_1)^{\mathcal{I}^3}$. In particular, for a special one where $d_x \in (\alpha_0)^{\mathcal{I}^3}$, this means that $d_x \in (A_1)^{\mathcal{I}^3}$. Now, β_2^r is obviously not a $\tau\mathcal{EL}(\text{deg})$ concept, but its meaning can be equivalently expressed in the logic. We illustrate this with the help of the following diagram.



Let E_2^3 be the \mathcal{EL} concept description associated to the description tree $T_{E_2^3}$ shown on the left-hand side of the diagram. Notice that $T_{E_2^3}$ exhibits all (*and only*) paths falsifying β_2^r . Based on it, one can observe the following:

- if d_0 has an r -successor leading to one such path (d_1 on the *right-hand* side), then there is always a *ptgh* h from $T_{E_2^3}$ to $G_{\mathcal{I}}$ such that $h(v) = d_1$ and $h_w(v) > 0$. Therefore, $d_1 \notin [(E_2^3)_{\leq 0}]^{\mathcal{I}}$.
- Conversely, if no such path exists, then for all partial mappings h from $T_{E_2^3}$ to $G_{\mathcal{I}}$ and all paths $vx_2v_2x_3v_3$ in $T_{E_2^3}$ such that $v_3 \in \text{dom}(h)$, it is the case that $h(v_3) \notin (\bar{A}_1)^{\mathcal{I}}$. Therefore, by definition of h_w , it must be the case that $h_w(v) = 0$. Consequently, $\text{deg}^{\mathcal{I}}(d_1, E_2^3) = 0$ and $d_1 \in [(E_2^3)_{\leq 0}]^{\mathcal{I}}$.

Hence, β_2^r is equivalent to the threshold concept $(E_2^3)_{\leq 0}$. To express β_2^s , we define a dual concept $E_2^{\bar{3}}$, by using A_i instead of \bar{A}_i . Finally, to succinctly represent these “exponentially large” concepts, we employ the \mathcal{EL} TBoxes \mathcal{T}^3 and $\mathcal{T}^{\bar{3}}$ defined as follows:

$$\mathcal{T}^3 := \left\{ \begin{array}{l} E_2^3 \doteq \exists r.E_1^3 \sqcap \exists s.E_1^3 \\ E_1^3 \doteq \exists r.E_0^3 \sqcap \exists s.E_0^3 \\ E_0^3 \doteq \bar{A}_1 \end{array} \right\} \quad \mathcal{T}^{\bar{3}} := \left\{ \begin{array}{l} E_2^{\bar{3}} \doteq \exists r.E_1^{\bar{3}} \sqcap \exists s.E_1^{\bar{3}} \\ E_1^{\bar{3}} \doteq \exists r.E_0^{\bar{3}} \sqcap \exists s.E_0^{\bar{3}} \\ E_0^{\bar{3}} \doteq A_1 \end{array} \right\}$$

The good that comes from this is that we obtain the following equivalences:

$$(E_2^3)_{\leq 0} \equiv_{\mathcal{T}^3} \prod_{x_2, x_3 \in \{r, s\}} \forall x_2 x_3. \neg \bar{A}_1 \quad \text{and} \quad (E_2^{\bar{3}})_{\leq 0} \equiv_{\mathcal{T}^{\bar{3}}} \prod_{x_2, x_3 \in \{r, s\}} \forall x_2 x_3. \neg A_1$$

The same arguments also apply to \bar{A}_2, \bar{A}_3 and A_2, A_3 , by defining similar concept descriptions E_1^2, E_0^1 and $E_1^{\bar{2}}, E_0^{\bar{1}}$, respectively, and their corresponding TBoxes $\mathcal{T}^2, \mathcal{T}^1$ and $\mathcal{T}^{\bar{2}}, \mathcal{T}^{\bar{1}}$. Based on this, one can in general use the threshold concepts $(E_{i-1}^i)_{\leq 0}$ and $(E_{i-1}^{\bar{i}})_{\leq 0}$ to represent the generalization of the value restrictions used in (14) to arbitrary lengths. Let $\mathcal{T}_{n,p}$ be the union of all these TBoxes, i.e.,

$$\mathcal{T}_{n,p} := \bigcup_{i=1}^n (\mathcal{T}^i \cup \mathcal{T}^{\bar{i}})$$

Proposition 27. *For all models \mathcal{I} of $\mathcal{T}_{n,p}$, $d \in \Delta^{\mathcal{I}}$ and $1 \leq i \leq n$:*

1. $d \in [(E_{i-1}^i)_{\leq 0}]^{\mathcal{I}}$ iff $d \in \left(\prod_{x \in \{r, s\}^{i-1}} \forall x. \neg \bar{A}_{n-i+1} \right)^{\mathcal{I}}$.
2. $d \in [(E_{i-1}^{\bar{i}})_{\leq 0}]^{\mathcal{I}}$ iff $d \in \left(\prod_{x \in \{r, s\}^{i-1}} \forall x. \neg A_{n-i+1} \right)^{\mathcal{I}}$.

Proof. We only give the proof for the first statement (the second one can be shown using the same argument). We denote as $T_{E_{i-1}^i}$ the description tree corresponding to the unfolding $u_{\mathcal{T}^i}(E_{i-1}^i)$ of E_{i-1}^i in \mathcal{T}^i . For simplicity, we use just ℓ (without subscript) to refer to the labeling of $T_{E_{i-1}^i}$.

(\Rightarrow) Assume that $d \in [(E_{i-1}^i)_{\leq 0}]^{\mathcal{I}}$. Since E_{i-1}^i is a defined concept in \mathcal{T}^i , this implies:

$$\text{deg}^{\mathcal{I}}(d, E_{i-1}^i, \mathcal{T}^i) = \text{deg}^{\mathcal{I}}(d, u_{\mathcal{T}^i}(E_{i-1}^i)) = 0$$

For a contradiction, suppose that:

$$d \notin \left(\prod_{x \in \{r, s\}^{i-1}} \forall x. \neg \bar{A}_{n-i+1} \right)^{\mathcal{I}}$$

Then, there is a word $x_1 \dots x_{i-1} \in \{r, s\}^{i-1}$ such that $d \notin (\forall x_1 \dots x_{i-1}. \neg \bar{A}_{n-i+1})^{\mathcal{I}}$. The semantics of the value restriction constructor yields the existence of a path of the form $dx_1 d_1 \dots x_{i-1} d_{i-1}$ in $G_{\mathcal{I}}$ such that $d_{i-1} \in (\bar{A}_{n-i+1})^{\mathcal{I}}$.

By definition of E_{i-1}^i in \mathcal{T}^i , there is a path $v_0 x_1 v_1 \dots x_{i-1} v_{i-1}$ in $T_{E_{i-1}^i}$ with $\ell(v_{i-1}) = \{\bar{A}_{n-i+1}\}$, where v_0 is the root of $T_{E_{i-1}^i}$. Therefore, the *ptgh* h from $T_{E_{i-1}^i}$ to $G_{\mathcal{I}}$ with $h(v_0) = d$ and $h(v_j) = d_j$ ($1 \leq j \leq i-1$) induces a weighted homomorphism h_w such that: $h_w(v_0) > 0$. This contradicts our initial assumption since it implies:

$$\text{deg}^{\mathcal{I}}(d, u_{\mathcal{T}^i}(E_{i-1}^i)) > 0$$

Thus, the left to right implication holds.

(\Leftarrow) Assume that

$$d \in \left(\prod_{x \in \{r, s\}^{i-1}} \forall x. \neg \bar{A}_{n-i+1} \right)^{\mathcal{I}}$$

This implies that $d \in (\forall x_1 \dots x_{i-1}. \neg \bar{A}_{n-i+1})^{\mathcal{I}}$ for all words $x_1, \dots, x_{i-1} \in \{r, s\}^{i-1}$. Hence, any path of the form $dx_1 d_1 \dots x_{i-1} d_{i-1}$ in $G_{\mathcal{I}}$ is restricted to have:

$$d_{i-1} \notin (\bar{A}_{n-i+1})^{\mathcal{I}}$$

Let now $v_0 x_1 v_1 \dots x_{i-1} v_{i-1}$ be any path in $T_{E_{i-1}^i}$. By definition of E_{i-1}^i in \mathcal{T}^i we know that $x_1 \dots x_{i-1} \in \{r, s\}^{i-1}$ and $\ell(v_{i-1}) = \{\bar{A}_{n-i+1}\}$. Therefore, for all *ptgh* h from $T_{E_{i-1}^i}$ to $G_{\mathcal{I}}$ having $h(v_0) = d$ and $v_{i-1} \in \text{dom}(h)$, it is the case that $\bar{A}_{n-i+1} \notin \ell_{\mathcal{I}}(h(v_{i-1}))$. Hence, since v_{i-1} is a leaf in $T_{E_{i-1}^i}$, this means that $h_w(v_{i-1}) = 0$.

Overall, we have shown that for all leaves v in $T_{E_{i-1}^i}$ and all *ptgh* h with $v \in \text{dom}(h)$, it holds that $h_w(v) = 0$. Then, since $\ell(v) = \emptyset$ iff v is a non-leaf node, there is no possible way in which $h_w(v_0) > 0$. Consequently, it follows:

$$\text{deg}^{\mathcal{I}}(d, u_{\mathcal{T}^i}(E_{i-1}^i)) = 0$$

Thus, $d \in [(E_{i-1}^i)_{\leq 0}]^{\mathcal{I}}$. □

Having these equivalences, the next step is to integrate the threshold concepts of the form $(E_{i-1}^i)_{\leq 0}$ and $(\bar{E}_{i-1}^i)_{\leq 0}$ into the definitions of \mathcal{T}_n . For all $1 \leq i \leq n$:

$$\alpha_i \doteq \exists r.(\alpha_{i-1} \sqcap (E_{i-1}^i)_{\leq 0}) \sqcap \exists s.(\alpha_{i-1} \sqcap (\bar{E}_{i-1}^i)_{\leq 0})$$

We name the resulting TBox as $\mathcal{T}_{n,\tau}$ (including the modification of α_0 as stated in (13)). Then, the final acyclic $\tau\mathcal{EL}(\text{deg})$ TBox $\widehat{\mathcal{T}}_n^\varphi$ is the pair $(\mathcal{T}_{n,\tau}, \mathcal{T}_{n,\ell}^c \cup \mathcal{T}_{n,p})$. The following proposition is the equivalent of Proposition 26 w.r.t. $\widehat{\mathcal{T}}_n^\varphi$.

Proposition 28. *Let \mathcal{I} be a model of $\widehat{\mathcal{T}}_n^\varphi$ and $d \in \Delta^\mathcal{I}$. For all $0 \leq i \leq n$: if $d \in (\alpha_i)^\mathcal{I}$, then for each word $x \in \{r, s\}^i$ there exists a path $dx_1d_1 \dots x_id_i$ in $G_\mathcal{I}$ such that for all $1 \leq j \leq i$,*

- $d_i \in (\alpha_{i-j})^\mathcal{I}$,
- $d_i \in [(E_{i-j}^{i-j+1})_{\leq 0}]^\mathcal{I}$ if $x_i = r$, otherwise $d_i \in [(\bar{E}_{i-j}^{i-j+1})_{\leq 0}]^\mathcal{I}$.

We now prove an intermediate result that is equivalent to having the wanted properties in (11) and (12).

Lemma 29. *For all $n \geq 0$ and all interpretations \mathcal{I} such that $\mathcal{I} \models \widehat{\mathcal{T}}_n^\varphi$ and $(\alpha_n)^\mathcal{I} \neq \emptyset$ the following holds:*

- for all subsets X of $\{A_1, \dots, A_n\}$ there exists $d_X \in \Delta^\mathcal{I}$ such that:

$$d_X \in (\alpha_0)^\mathcal{I} \text{ and } d_X \in (A_i)^\mathcal{I} \text{ iff } A_i \in X \quad (1 \leq i \leq n) \quad (15)$$

Proof. Let \mathcal{I} be an interpretation such that $\mathcal{I} \models \widehat{\mathcal{T}}_n^\varphi$ and $(\alpha_n)^\mathcal{I} \neq \emptyset$. We fix an arbitrary subset X of $\{A_1, \dots, A_n\}$ and show that X satisfies (15). Let $d \in \Delta^\mathcal{I}$ be an element in $(\alpha_n)^\mathcal{I}$. We define the word $x \in \{r, s\}^n$ corresponding to X as follows:

$$x_i = r \text{ iff } A_i \in X \quad (1 \leq i \leq n) \quad (16)$$

The application of Proposition 28 to d , yields that there is a path $dx_1d_1 \dots x_nd_n$ in $G_\mathcal{I}$ such that for all $1 \leq i \leq n$:

- $d_i \in (\alpha_{n-i})^\mathcal{I}$,
- $d_i \in [(E_{n-i}^{n-i+1})_{\leq 0}]^\mathcal{I}$ if $x_i = r$, otherwise $d_i \in [(\bar{E}_{n-i}^{n-i+1})_{\leq 0}]^\mathcal{I}$

In particular, the suffix $d_ix_{i+1} \dots x_nd_n$ is of length $n - i$. Therefore, we further have:

$$\begin{aligned} x_i = r &\Rightarrow d_i \in [(E_{n-i}^{n-i+1})_{\leq 0}]^\mathcal{I} \\ &\Rightarrow d_i \in (\forall x_{i+1} \dots x_n. \neg \bar{A}_i)^\mathcal{I} \quad (\text{Proposition 27 applied to } d_i \text{ and } E_{n-i}^{n-i+1}) \\ &\Rightarrow d_n \notin (\bar{A}_i)^\mathcal{I} \end{aligned}$$

Symmetrically, $x_i = s$ implies $d_n \notin (A_i)^{\mathcal{I}}$. Now, we know that $F_i \doteq A_i \sqcap \bar{A}_i \in \mathcal{T}_{n,\ell}^c$ and α_0 is of the form:

$$\alpha_0 \doteq \widehat{C}_\varphi$$

Since $d_n \in (\alpha_0)^{\mathcal{I}}$ it follows:

$$d_n \in (A_i)^{\mathcal{I}} \text{ iff } x_i = r \quad (1 \leq i \leq n)$$

From the way the word x is defined in (16), we can conclude that d_n is an element of $\Delta^{\mathcal{I}}$ satisfying (15) with respect to X . \square

Using Lemma 29 we finally show that the described reduction is correct.

Lemma 30. $\forall u \exists v. \varphi(u, v)$ is valid iff α_n is satisfiable in $\widehat{\mathcal{T}}_n^\varphi$.

Proof. (\Rightarrow) Assume that $(\forall u)(\exists v)\varphi(u, v)$ is valid. We take the interpretation \mathcal{I}_n and extend it into a model $\widehat{\mathcal{I}}_n$ of $\widehat{\mathcal{T}}_n^\varphi$ satisfying α_n . By construction, \mathcal{I}_n is tree-shaped and there is a one-to-one correspondence between words in $\{r, s\}^n$ and the leaves in T_n . The leaf d_x corresponding to the word x is the one reached from d_0 through the path $d_0x_1d_1 \dots x_nd_x$. Let \mathcal{L}_n denote the set of leaves of T_n . The interpretation of A_i, \bar{A}_i under $\widehat{\mathcal{I}}_n$ is defined as follows. For all $1 \leq i \leq n$:

$$\begin{aligned} (A_i)^{\widehat{\mathcal{I}}_n} &:= \{d_x \mid d_x \in \mathcal{L}_n \text{ and } x_i = r\} \\ (\bar{A}_i)^{\widehat{\mathcal{I}}_n} &:= \{d_x \mid d_x \in \mathcal{L}_n \text{ and } x_i = s\} \end{aligned} \quad (17)$$

Hence, for all leaves d of T_n and all $i \in \{1 \dots n\}$ we have:

$$d \in (A_i)^{\widehat{\mathcal{I}}_n} \text{ iff } d \notin (\bar{A}_i)^{\widehat{\mathcal{I}}_n} \quad (18)$$

To assign meaning to B_j, \bar{B}_j under $\widehat{\mathcal{I}}_n$, let d be any leaf of T_n . We define the assignment \mathbf{t}_d for $u \cup v$ as follows. First,

$$\mathbf{t}_d(u_i) = \text{true} \text{ iff } d \in (A_i)^{\widehat{\mathcal{I}}_n} \quad (1 \leq i \leq n)$$

Second, \mathbf{t}_d assigns truth values to the variables in v such that it satisfies $\varphi(u, v)$. This is always possible because $(\forall u)(\exists v)\varphi(u, v)$ is valid. Then, for all $1 \leq j \leq \ell$:

$$\begin{aligned} (B_j)^{\widehat{\mathcal{I}}_n} &:= \{d \mid d \in \mathcal{L}_n \text{ and } \mathbf{t}_d(v_j) = \text{true}\} \\ (\bar{B}_j)^{\widehat{\mathcal{I}}_n} &:= \{d \mid d \in \mathcal{L}_n \text{ and } \mathbf{t}_d(v_j) = \text{false}\} \end{aligned} \quad (19)$$

Hence, as in (18), for all leaves d of T_n and all $j \in \{1 \dots \ell\}$ we have:

$$d \in (B_j)^{\widehat{\mathcal{I}}_n} \text{ iff } d \notin (\bar{B}_j)^{\widehat{\mathcal{I}}_n} \quad (20)$$

Now, let us look at the interpretation \mathcal{I}_0 and the element $d_0 \in \Delta^{\mathcal{I}}$ constructed in the proof of Lemma 25. One can observe the following similarities between d_0 and any leaf $d \in \mathcal{L}_n$:

- d_0 satisfies (9) and (10), while d satisfies (18) and (20) w.r.t. $\widehat{\mathcal{I}}_n$.
- the assignment \mathbf{t} used in Lemma 25 and the current truth assignment \mathbf{t}_d satisfy both $\varphi(u, v)$.
- the correspondence established between $\mathbf{t}(u_i)$ and $\mathbf{t}(v_j)$, and the membership of d_0 in concept names A_i and B_j under \mathcal{I}_0 , is the same as the one between \mathbf{t}_d and d under $\widehat{\mathcal{I}}_n$. Similarly for negated literals and membership in \bar{A}_i and \bar{B}_j .

Therefore, it is not hard to conclude that like in Lemma 25 for d_0 and \mathcal{I}_0 , $d \in (\widehat{C}_\varphi)^{\widehat{\mathcal{I}}_n}$ for all $d \in \mathcal{L}_n$.

Since $\widehat{\mathcal{T}}_n^\varphi$ is an acyclic $\tau\mathcal{EL}(\text{deg})$ TBox, there is a unique way to extend $\widehat{\mathcal{I}}_n$ into a model of $\widehat{\mathcal{T}}_n^\varphi$. Having done so, let $\mathbf{h}(d)$ denote the *height* of a domain element d in T_n . We show by induction on $\mathbf{h}(d)$ the following claim:

$$\text{for all } d \in \Delta^{\widehat{\mathcal{I}}_n}: d \in (\alpha_{\mathbf{h}(d)})^{\widehat{\mathcal{I}}_n}$$

Induction Base. $d \in \Delta^{\widehat{\mathcal{I}}_n}$ and $\mathbf{h}(d) = 0$. Then, d is a leaf in T_n . Recall that α_0 is defined in $\mathcal{T}_{n,\tau}$ as:

$$\alpha_0 \doteq \widehat{C}_\varphi$$

As explained above, we have $d \in (\widehat{C}_\varphi)^{\widehat{\mathcal{I}}_n}$. Hence, $d \in (\alpha_0)^{\widehat{\mathcal{I}}_n}$.

Induction Step. Let $d \in \Delta^{\widehat{\mathcal{I}}_n}$ with $0 < \mathbf{h}(d) \leq n$. We assume our claim holds for all $e \in \Delta^{\widehat{\mathcal{I}}_n}$ with $\mathbf{h}(e) < \mathbf{h}(d)$.

To start, $\alpha_{\mathbf{h}(d)}$ is defined in $\mathcal{T}_{n,\tau}$ as:

$$\alpha_{\mathbf{h}(d)} \doteq \exists r. (\alpha_{\mathbf{h}(d)-1} \sqcap (E_{\mathbf{h}(d)-1}^{\mathbf{h}(d)})_{\leq 0}) \sqcap \exists s. (\alpha_{\mathbf{h}(d)-1} \sqcap (\overline{E_{\mathbf{h}(d)-1}^{\mathbf{h}(d)}}})_{\leq 0})$$

By construction of T_n , there exists $e \in \Delta^{\widehat{\mathcal{I}}_n}$ such that $(d, e) \in r^{\widehat{\mathcal{I}}_n}$ and $\mathbf{h}(e) = \mathbf{h}(d) - 1$. The application of induction hypothesis to e yields $e \in (\alpha_{\mathbf{h}(d)-1})^{\widehat{\mathcal{I}}_n}$.

Consider now any word $y \in \{r, s\}^{\mathbf{h}(d)-1}$. Since $\mathbf{h}(e) = \mathbf{h}(d) - 1$, by definition of T_n there is a unique path of the form $ey_1e_1 \dots y_{\mathbf{h}(d)-1}e_{\mathbf{h}(d)-1}$ in T_n , where $e_{\mathbf{h}(d)-1}$ is a leaf. Moreover, such a path is suffix of a path $d_0x_1d_1 \dots d_jx_jex_{j+1} \dots x_nd_n$, where $d_j = d$, $x_j = r$ and $e_{\mathbf{h}(d)-1} = d_n$. Then, we obtain the following equalities:

$$\begin{aligned} n - (j + 1) &= (\mathbf{h}(d) - 1) - 1 \\ n - \mathbf{h}(d) + 1 &= j \end{aligned}$$

Since $x_j = r$, by (17) we obtain that $d_n \in (A_{n-\mathbf{h}(d)+1})^{\widehat{\mathcal{I}}_n}$, and by (18) $d_n \notin (\bar{A}_{n-\mathbf{h}(d)+1})^{\widehat{\mathcal{I}}_n}$. Hence, as y was chosen arbitrarily from $\{r, s\}^{\mathbf{h}(d)-1}$, we have just

shown that:

$$e \in \left(\prod_{y \in \{r,s\}^{h(d)-1}} \forall y. \neg \bar{A}_{n-h(d)+1} \right)^{\mathcal{I}_n}$$

The application of Proposition 27 then yields $e \in [(E_{h(d)-1}^{h(d)})_{\leq 0}]^{\hat{\mathcal{I}}_n}$. For the $\exists s$ restriction in the definition of $\alpha_{h(d)}$, the corresponding result can be shown in the same way. Therefore, $d \in (\alpha_{h(d)})^{\hat{\mathcal{I}}_n}$.

Using this result and the fact that d_0 is of height n in T_n , we can conclude that $d_0 \in (\alpha_n)^{\hat{\mathcal{I}}_n}$. Thus, α_n is satisfiable with respect to $\hat{\mathcal{T}}_n^\varphi$.

(\Leftarrow) Conversely, assume that α_n is satisfiable with respect to $\hat{\mathcal{T}}_n^\varphi$. This means that there exists an interpretation \mathcal{I} such that $\mathcal{I} \models \hat{\mathcal{T}}_n^\varphi$ and $(\alpha_n)^\mathcal{I} \neq \emptyset$. Let us fix a partial truth assignment \mathbf{t} covering all the variables in u . We show that \mathbf{t} can be extended to v in such a way that it satisfies $\varphi(u, v)$. The subset $X_{\mathbf{t}}$ of $\{A_1, \dots, A_n\}$ is induced by \mathbf{t} as follows:

$$X_{\mathbf{t}} := \{A_i \mid \mathbf{t}(u_i) = \text{true}\} \quad (1 \leq i \leq n)$$

By Lemma 29, we know that there exists $d_{\mathbf{t}} \in \Delta^\mathcal{I}$ such that:

- $d_{\mathbf{t}} \in (A_i)^\mathcal{I}$ iff $A_i \in X_{\mathbf{t}}$ (iff $\mathbf{t}(u_i) = \text{true}$),
- $d_{\mathbf{t}} \in (\alpha_0)^\mathcal{I}$.

We use $d_{\mathbf{t}}$ to extend \mathbf{t} to v as follows. For all $1 \leq j \leq \ell$:

$$\mathbf{t}(v_j) = \text{true} \text{ iff } d_{\mathbf{t}} \in (B_j)^\mathcal{I}$$

Therefore, since $d_{\mathbf{t}}$ satisfies the complementary restrictions required by $(F_i)_{=1/2}$ and $(G_j)_{=1/2}$ in the definition of \hat{C}_φ for A_i, \bar{A}_i ($1 \leq i \leq n$) and B_j, \bar{B}_j ($1 \leq j \leq \ell$), respectively, we further obtain for all literals γ over $u \cup v$:

$$\mathbf{t}(\gamma) = \text{true} \text{ iff } d_{\mathbf{t}} \in (\eta(\gamma))^\mathcal{I} \quad (21)$$

Moreover, since $d_{\mathbf{t}} \in (\hat{C}_\varphi)^\mathcal{I}$ we have that $d_{\mathbf{t}} \in [(D_k)_{\geq \frac{1}{3}}]^\mathcal{I}$ for all $1 \leq k \leq q$. By definition of deg and D_k there must exist a literal γ in c_k such that $d_{\mathbf{t}} \in (\eta(\gamma))^\mathcal{I}$. It then follows from (21) that \mathbf{t} satisfies every clause $c_k \in \mathcal{C}$, and consequently it satisfies $\varphi(u, v)$. Since the partial truth assignment \mathbf{t} for u was chosen arbitrarily, we thus have shown that $(\forall u)(\exists v)\varphi(u, v)$ is valid. \square

Finally, it is easy to see that $\hat{\mathcal{T}}_n^\varphi$ is an acyclic $\tau\mathcal{EL}(\text{deg})$ TBox and its size is polynomial in the size of $\forall u \exists v. \varphi(u, v)$. Therefore, $\forall \exists \text{3SAT}$ is polynomial-time reducible to concept satisfiability w.r.t. acyclic $\tau\mathcal{EL}(\text{deg})$ TBoxes. In addition, *non-satisfiability* can be reduced to the subsumption and the instance problem, and satisfiability to the consistency problem, by the same argument used in [BBG15] for the setting without TBoxes. Thus, we obtain the following lower bounds.

Lemma 31. *In $\tau\mathcal{EL}(deg)$, satisfiability and consistency are Π_2^P -hard and the subsumption and the instance problems are Σ_2^P -hard, with respect to acyclic $\tau\mathcal{EL}(deg)$ TBoxes.*

4.2 Normalization

To simplify the technical development of the decision procedures presented in the next section, it is convenient to use TBoxes in a special form. We now introduce *normalized $\tau\mathcal{EL}(deg)$ TBoxes* in *reduced* form, and show that one can (without loss of generality) restrict the attention to this kind of TBoxes.

Let us start by recalling the normal form for \mathcal{EL} TBoxes introduced in [Baa02]. An \mathcal{EL} TBox \mathcal{T} is said to be normalized iff $\alpha \doteq C_\alpha \in \mathcal{T}$ implies that C_α is of the form:

$$P_1 \sqcap \dots \sqcap P_k \sqcap \exists r_1.\beta_1 \sqcap \dots \sqcap \exists r_n.\beta_n$$

where $k, n \geq 0$, $P_1, \dots, P_k \in \mathbf{N}_{pr}^{\mathcal{T}}$, and $\beta_1, \dots, \beta_n \in \mathbf{N}_d^{\mathcal{T}}$. We extend this form to $\tau\mathcal{EL}(deg)$, and say that a $\tau\mathcal{EL}(deg)$ TBox $\widehat{\mathcal{T}} = (\mathcal{T}_\tau, \mathcal{T})$ is normalized iff \mathcal{T} is normalized and $\alpha \doteq \widehat{C}_\alpha \in \mathcal{T}_\tau$ implies that \widehat{C}_α is of the form:

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

where $k, n \geq 0$, for all $1 \leq i \leq k$ either $\widehat{P}_i \in \mathbf{N}_{pr}^{\widehat{\mathcal{T}}}$ or it is of the form $E_{\sim t}$ with $E \in \mathbf{N}_d^{\widehat{\mathcal{T}}}$, and $\beta_1, \dots, \beta_n \in \mathbf{N}_d^{\widehat{\mathcal{T}}}$.

To illustrate this normalization process we start with a simpler version of Example 12 in [Baa02].

Example 32. Let \mathcal{T} be the \mathcal{EL} TBox consisting of the following definitions:

$$\begin{aligned} \alpha_1 &\doteq P_1 \sqcap \alpha_2 \sqcap \exists r_1.\exists r_2.\alpha_3 \\ \alpha_2 &\doteq P_2 \sqcap \alpha_3 \sqcap \exists s.(\alpha_3 \sqcap P_3) \\ \alpha_3 &\doteq P_4 \end{aligned}$$

Using auxiliary definitions we obtain a new TBox \mathcal{T}' :

$$\begin{aligned} \alpha_1 &\doteq P_1 \sqcap \alpha_2 \sqcap \exists r_1.\beta_1 \\ \beta_1 &\doteq \exists r_2.\alpha_3 \\ \alpha_2 &\doteq P_2 \sqcap \alpha_3 \sqcap \exists s.\beta_2 \\ \beta_2 &\doteq \alpha_3 \sqcap P_3 \\ \alpha_3 &\doteq P_4 \end{aligned}$$

This step is formalized as the exhaustive application of the rule R_{\exists} .

Condition: applies to concept definitions of the form $\alpha \doteq C_1 \sqcap \dots \sqcap C_n$ if there is an index $i \in \{1, \dots, n\}$ with $C_i = \exists r.D$ and $D \notin \mathbf{N}_d^{\mathcal{T}}$.

Action: its application replaces the conjunct C_i by $\exists r.\beta$, and introduces a new definition $\beta \doteq D$, where β is a fresh concept name.

Since α_1, α_2 and β_2 contain top-level atoms which are defined concepts, \mathcal{T}' is not yet normalized. The original normalization process is devised to handle cyclic \mathcal{EL} TBoxes that can be interpreted by different types of semantics. Consequently, the approach used to overcome this problem varies according to each semantics. In our case, however, this becomes simpler since the \mathcal{EL} TBox \mathcal{T} we are dealing with is acyclic. The solution for this follows from the discussion presented in [Baa02] for the general case, and consists of substituting these occurrences of defined concepts by their definitions. Following the example we obtain the following TBox:

$$\begin{aligned}\alpha_1 &\doteq P_1 \sqcap P_2 \sqcap P_4 \sqcap \exists s.\beta_2 \sqcap \exists r_1.\beta_1 \\ \beta_1 &\doteq \exists r_2.\alpha_3 \\ \alpha_2 &\doteq P_2 \sqcap P_4 \sqcap \exists s.\beta_2 \\ \beta_2 &\doteq P_4 \sqcap P_3 \\ \alpha_3 &\doteq P_4\end{aligned}$$

We name the corresponding rule R_α and formally define it as follows.

Condition: applies to concept definitions of the form $\alpha \doteq C_1 \sqcap \dots \sqcap C_n$ if there is an index $i \in \{1, \dots, n\}$ with $C_i = \beta$ and $\beta \doteq C_\beta \in \mathcal{T}$.

Action: its application replaces C_i by C_β .

Then, once R_\exists can no longer be applied, an exhaustive application of the rule R_α will produce a normalized acyclic \mathcal{EL} TBox. However, to have a polynomial time procedure generating a new TBox of polynomial size, the sequence of applications of R_α should not be arbitrary. This is achieved by following the order \preceq induced by \rightarrow^+ , i.e., R_α can be applied to a concept definition $\alpha \doteq C_\alpha$ only if it has already been applied to all $\beta \in \mathbf{N}_d^{\mathcal{T}}$ such that $\beta \preceq \alpha$.

Each application of R_\exists replaces a top-level atom of the form $\exists r.D$ with a new atom $\exists r.\beta$, and introduces a simpler definition $\beta \doteq D$. Concerning R_α , such an ordered sequence of rule applications will always terminate since we are dealing with acyclic TBoxes. Moreover, the *idempotency* of \sqcap can be exploited to avoid duplications. Hence, R_α is only applied one time for each top-level atom of the form $\beta \in \mathbf{N}_d^{\mathcal{T}}$ occurring in the TBox that results from the application of R_\exists , and it does not cause an exponential *blow-up* of the size of the TBox. Thus, 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 $\tau\mathcal{EL}(deg)$ TBoxes. The rules R_{\exists} and R_{α} can be applied to \mathcal{T}_{τ} in the same way. The only difference is that to apply R_{α} in \mathcal{T}_{τ} , the definition $\beta \doteq C_{\beta}$ may also occur in \mathcal{T} . Additionally, it is required that all occurrences of threshold concepts $E_{\sim t}$ in $\widehat{\mathcal{T}}$ are such that E is a defined concept in \mathcal{T} . For example, α_1 could have been defined as:

$$\alpha_1 \doteq P_1 \sqcap \exists r_1. [(P_2 \sqcap \exists r_2. P_3)_{\leq 8}] \sqcap \exists r_1. \exists r_2. \alpha_3$$

To handle this we use a new rule R_{\sim} .

Condition: applies to concept definitions of the form $\alpha \doteq \widehat{C}_1 \sqcap \dots \sqcap \widehat{C}_n \in \mathcal{T}_{\tau}$ if there is an index $i \in \{1, \dots, n\}$ with $\widehat{C}_i = D_{\sim t}$ and $D \notin \mathbf{N}_d^{\mathcal{T}}$.

Action: its application replaces the conjunct \widehat{C}_i by $(E_D)_{\sim t}$, and adds a new definition $E_D \doteq D$ to \mathcal{T} , being E_D a fresh concept name.

Thus, the normalization will yield the $\tau\mathcal{EL}(deg)$ TBox $\widehat{\mathcal{T}} = (\mathcal{T}_{\tau}, \mathcal{T})$ consisting of the following two sets of definitions:

$$\begin{aligned} \alpha_1 &\doteq P_1 \sqcap \exists r_1. \beta_4 \sqcap \exists r_1. \beta_1 \\ \beta_4 &\doteq E_{\leq 8} \\ \beta_1 &\doteq \exists r_2. \alpha_3 \\ \alpha_2 &\doteq P_2 \sqcap P_4 \sqcap \exists s. \beta_2 \\ \beta_2 &\doteq P_4 \sqcap P_3 \\ \alpha_3 &\doteq P_4 \end{aligned}$$

and \mathcal{T} the following set:

$$E \doteq P_2 \sqcap \exists r_2. P_3$$

Notice that in order to trigger the application of R_{\sim} , the concerned existential restriction in the definition of α_1 had to be first decomposed by applying R_{\exists} . With this in mind, we define the normalization procedure for acyclic $\tau\mathcal{EL}(deg)$ TBoxes as the execution of the following steps.

1. Apply the rule R_{\exists} exhaustively to \mathcal{T}_{τ} .
2. Apply the rule R_{\sim} to \mathcal{T}_{τ} as long as possible.
3. Normalize the augmented \mathcal{EL} TBox \mathcal{T} .
4. Apply the rule R_{α} exhaustively to \mathcal{T}_{τ} .

The applications of R_{\exists} and R_{α} to \mathcal{T}_{τ} modify only \mathcal{T}_{τ} , while no new threshold expressions are introduced. Regarding the second step, as $D_{\sim t}$ is such that no defined concept in \mathcal{T}_{τ} occurs in it, the unfolding of the threshold concept $(E_D)_{\sim t}$ introduced by the application of the rule $R_{\sim t}$ is also such that no defined concept in \mathcal{T}_{τ} occurs in it. Furthermore, adding $E_D \doteq D$ to \mathcal{T} does not introduce any concept name $\alpha \in \mathbf{N}_d^{\mathcal{T}_{\tau}}$ in definitions of \mathcal{T} . Finally, the normalization of \mathcal{T} only transforms the structure of \mathcal{T} . Therefore, $\widehat{\mathcal{T}}'$ satisfies the restrictions required for $\tau\mathcal{EL}(m)$ TBoxes in Definition 18, and it is easy to see that no cycles are introduced in it.

Now, after the first step has been executed, all occurrences of threshold concepts in \mathcal{T}_{τ} appear as top-level atoms on its concept definitions. Consequently, the application of R_{\sim} in the second step will cover all of them. Moreover, the normalization of \mathcal{T} before the final step guarantees that R_{\exists} need not be applied in case R_{α} applies to a defined concept in \mathcal{T} . Overall, this implies that the resulting TBox $\widehat{\mathcal{T}}'$ is normalized.

Last, one can see that the rule R_{\sim} is applied at most one time for each threshold concept $D_{\sim t}$ occurring in a definition of the initial TBox \mathcal{T}_{τ} . Consequently, at most polynomially many new definitions of the form $E_D \doteq D$ are added to \mathcal{T} . Thus, using the same arguments given for the application of R_{\exists} and R_{α} in the \mathcal{EL} setting, the devised normalization procedure runs in polynomial time and yields a normalized acyclic $\tau\mathcal{EL}(deg)$ TBox $\widehat{\mathcal{T}}'$ of size polynomial in the size of $\widehat{\mathcal{T}}$.

We now show that normalization preserves the unfolding of defined concepts.

Lemma 33. *Let $\widehat{\mathcal{T}}$ be an acyclic $\tau\mathcal{EL}(deg)$ TBox and $\widehat{\mathcal{T}}'$ the $\tau\mathcal{EL}(deg)$ TBox that results from a single application of a normalization rule. Then, for all defined concepts α in $\widehat{\mathcal{T}}$, $u_{\widehat{\mathcal{T}}}(\alpha) = u_{\widehat{\mathcal{T}}'}(\alpha)$*

Proof. Let R be a normalization rule and $\beta \doteq \widehat{C}_{\beta} \in \widehat{\mathcal{T}}$ the concept definition that R has been applied to. We use *well-founded* induction on the partial order induced by \rightarrow^+ on $\mathbf{N}_d^{\mathcal{T}_{\tau}} \cup \mathbf{N}_d^{\mathcal{T}}$. For all defined concepts α in $\widehat{\mathcal{T}}$ we distinguish two cases:

- $\alpha \neq \beta$. This means that R was not applied to $\alpha \doteq \widehat{C}_{\alpha}$, and consequently $\alpha \doteq \widehat{C}_{\alpha} \in \widehat{\mathcal{T}}'$. The top-down recursive application of unfolding through the structure of \widehat{C}_{α} with respect to $\widehat{\mathcal{T}}$ and $\widehat{\mathcal{T}}'$ may only result in different concept descriptions if:

$$u_{\widehat{\mathcal{T}}}(\alpha') \neq u_{\widehat{\mathcal{T}}'}(\alpha')$$

for some symbol α' occurring in \widehat{C}_{α} that corresponds to a defined concept name in $\widehat{\mathcal{T}}$. However, $\alpha \rightarrow^+ \alpha'$ and the application of the induction hypothesis to α' imply that this is never the case. Hence, $u_{\widehat{\mathcal{T}}}(\alpha) = u_{\widehat{\mathcal{T}}'}(\alpha)$.

- $\alpha = \beta$. Let \widehat{C}_{β} be of the form $\widehat{C}_1 \sqcap \dots \sqcap \widehat{C}_n$. We analyze the outcome of applying each of the three possible rules to $\beta \doteq \widehat{C}_{\beta}$:

- R_{\sim} : the rule was applied to a conjunct \widehat{C}_i such that $\widehat{C}_i = D_{\sim t}$ and $D \notin \mathbf{N}_d^{\mathcal{T}}$. Its application replaces \widehat{C}_i by $(E_D)_{\sim t}$ in \widehat{C}_β , and adds $E_D \doteq D$ to \mathcal{T}' where E_D is a fresh concept name. By definition of unfolding we have:

$$u_{\widehat{\mathcal{T}}}(\beta) = \prod_{j=1}^{i-1} u_{\widehat{\mathcal{T}}}(\widehat{C}_j) \sqcap [u_{\widehat{\mathcal{T}}}(D)]_{\sim t} \sqcap \prod_{j=i+1}^n u_{\widehat{\mathcal{T}}}(\widehat{C}_j)$$

and,

$$u_{\widehat{\mathcal{T}}'}(\beta) = \prod_{j=1}^{i-1} u_{\widehat{\mathcal{T}}'}(\widehat{C}_j) \sqcap [u_{\widehat{\mathcal{T}}'}(E_D)]_{\sim t} \sqcap \prod_{j=i+1}^n u_{\widehat{\mathcal{T}}'}(\widehat{C}_j)$$

Applying the same inductive argument used above we obtain $u_{\widehat{\mathcal{T}}}(\widehat{C}_j) = u_{\widehat{\mathcal{T}}'}(\widehat{C}_j)$ for all $j \neq i$ (likewise for D). Thus, since $u_{\widehat{\mathcal{T}}'}(E_D) = u_{\widehat{\mathcal{T}}'}(D)$ it follows that $u_{\widehat{\mathcal{T}}}(\beta) = u_{\widehat{\mathcal{T}}'}(\beta)$.

- R_{\exists} : the rule has been applied to an atom \widehat{C}_i of the form $\exists r.\widehat{D}$ such that $\widehat{D} \notin \mathbf{N}_d^{\mathcal{T}}$. Hence, \widehat{C}_i is substituted in \widehat{C}_β by $\exists r.\beta_1$ with β_1 being a fresh concept name and $\beta_1 \doteq \widehat{D} \in \widehat{\mathcal{T}}'$. By definition of unfolding we have:

$$u_{\widehat{\mathcal{T}}}(\beta) = \prod_{j=1}^{i-1} u_{\widehat{\mathcal{T}}}(\widehat{C}_j) \sqcap \exists r.u_{\widehat{\mathcal{T}}}(\widehat{D}) \sqcap \prod_{j=i+1}^n u_{\widehat{\mathcal{T}}}(\widehat{C}_j)$$

and,

$$u_{\widehat{\mathcal{T}}'}(\beta) = \prod_{j=1}^{i-1} u_{\widehat{\mathcal{T}}'}(\widehat{C}_j) \sqcap \exists r.u_{\widehat{\mathcal{T}}'}(\beta_1) \sqcap \prod_{j=i+1}^n u_{\widehat{\mathcal{T}}'}(\widehat{C}_j)$$

We know that $u_{\widehat{\mathcal{T}}'}(\beta_1) = u_{\widehat{\mathcal{T}}'}(\widehat{D})$. Hence, the same reasoning used for R_{\sim} yields $u_{\widehat{\mathcal{T}}}(\alpha) = u_{\widehat{\mathcal{T}}'}(\alpha)$.

- R_α : there is an index $1 \leq i \leq n$ such that \widehat{C}_i is of the form β_1 , and $\beta_1 \doteq \widehat{C}_{\beta_1} \in \widehat{\mathcal{T}}$. The application of R_α to β_1 replaces it in \widehat{C}_β with \widehat{C}_{β_1} . Since $\beta \rightarrow^+ \beta_1$, the application of induction hypothesis yields $u_{\widehat{\mathcal{T}}}(\beta_1) = u_{\widehat{\mathcal{T}}'}(\beta_1) = u_{\widehat{\mathcal{T}}'}(\widehat{C}_{\beta_1})$. Again, $u_{\widehat{\mathcal{T}}}(\widehat{C}_j) = u_{\widehat{\mathcal{T}}'}(\widehat{C}_j)$ for all $j \neq i$, and the rest follows from the definition of unfolding on β .

□

This property is then invariant under any number of rule applications. Therefore, the following proposition is a direct consequence of Lemma 33.

Proposition 34. *Let $\widehat{\mathcal{T}}$ be an acyclic $\tau\mathcal{EL}(\text{deg})$ TBox and $\widehat{\mathcal{T}}'$ the normal form of $\widehat{\mathcal{T}}$. For all defined concepts α in $\widehat{\mathcal{T}}$, $u_{\widehat{\mathcal{T}}}(\alpha) = u_{\widehat{\mathcal{T}}'}(\alpha)$.*

Proposition 34 implies that reasoning with respect to an acyclic $\tau\mathcal{EL}(deg)$ TBox $\widehat{\mathcal{T}}$ can be reduced to reasoning with respect to its normal form $\widehat{\mathcal{T}}'$. Therefore, from now on we only consider normalized TBoxes.

We still require one more transformation. Recall that for acyclic \mathcal{EL} TBoxes, the value of $deg^{\mathcal{I}}(d, C, \mathcal{T})$ is defined in terms of applying the basic definition of deg to the unfolding of C in \mathcal{T} . Moreover, deg needs to further translate $u_{\mathcal{T}}(C)$ into its reduced form $[u_{\mathcal{T}}(C)]^r$. Since $u_{\mathcal{T}}(C)$ may result in a concept of exponential size, it is certainly not a good idea to unfold and then compute the reduced form. To have this issue handled in a more transparent way by the decision procedures presented in the next section, we introduce the reduced form for acyclic \mathcal{EL} TBoxes. The ideas that follow are based on the results shown by Küsters in [Küs01].

Definition 35. Let \mathcal{T} be an acyclic \mathcal{EL} TBox and C an \mathcal{EL} concept description. Then, C is *reduced* with respect to \mathcal{T} iff:

- C is *reduced* according to Küsters' definition modulo $\sqsubseteq_{\mathcal{T}}$ (i.e., $\sqsubseteq_{\mathcal{T}}$ is used to identify redundancies instead of \sqsubseteq).

We say that \mathcal{T} is in reduced form iff for all $\alpha \doteq C_{\alpha} \in \mathcal{T}$ the concept C_{α} is reduced with respect to \mathcal{T} .

The benefit of using these type of TBoxes is that the unfolding of a defined concept will always result in a reduced concept description.

Lemma 36. *Let \mathcal{T} be a normalized acyclic \mathcal{EL} TBox in reduced form. Then, for all $\alpha \doteq C_{\alpha}$ the \mathcal{EL} concept description $u_{\mathcal{T}}(\alpha)$ is reduced.*

Proof. We use *well-founded* induction on \rightarrow^+ over $\mathbf{N}_d^{\mathcal{T}}$. Since \mathcal{T} is normalized, C_{α} has the following structure:

$$P_1 \sqcap \dots \sqcap P_k \sqcap \exists r_1.\alpha_1 \sqcap \dots \sqcap \exists r_n.\alpha_n$$

Clearly, $\alpha \rightarrow^+ \alpha_i$ for all $1 \leq i \leq n$. Therefore, the application of induction hypothesis yields that $u_{\mathcal{T}}(\alpha_i)$ is reduced. Now, since C_{α} is reduced with respect to \mathcal{T} , for all pairs $(\exists r_i.\alpha_i, \exists r_j.\alpha_j)$ we have:

- $r_i \neq r_j$, or
- $\alpha_i \not\sqsubseteq_{\mathcal{T}} \alpha_j$ and $\alpha_j \not\sqsubseteq_{\mathcal{T}} \alpha_i$.

In addition, we know that $\alpha_i \equiv_{\mathcal{T}} u_{\mathcal{T}}(\alpha_i)$ and $\alpha_j \equiv_{\mathcal{T}} u_{\mathcal{T}}(\alpha_j)$. This means that having $r_i = r_j$, it will be the case that $u_{\mathcal{T}}(\alpha_i) \not\sqsubseteq u_{\mathcal{T}}(\alpha_j)$ and $u_{\mathcal{T}}(\alpha_j) \not\sqsubseteq u_{\mathcal{T}}(\alpha_i)$. Finally, since $u_{\mathcal{T}}(\alpha)$ is the following concept description

$$P_1 \sqcap \dots \sqcap P_k \sqcap \exists r_1.u_{\mathcal{T}}(\alpha_1) \sqcap \dots \sqcap \exists r_n.u_{\mathcal{T}}(\alpha_n)$$

we can conclude that $u_{\mathcal{T}}(\alpha)$ is reduced. □

To translate acyclic \mathcal{EL} TBoxes into its reduced form, the algorithm sketched in [Küs01] (derived from Proposition 6.3.1.) to compute the reduced form of \mathcal{EL} concept descriptions comes in handy. By using $\sqsubseteq_{\mathcal{T}}$ instead of \sqsubseteq , it will be able to compute the reduced form $C^{r(\mathcal{T})}$ of a concept C with respect to \mathcal{T} . Since $\sqsubseteq_{\mathcal{T}}$ is decidable in polynomial time in \mathcal{EL} [Baa03], the modified procedure also runs in polynomial time. Moreover, the concept $C^{r(\mathcal{T})}$ satisfies $C \equiv_{\mathcal{T}} C^{r(\mathcal{T})}$.

Based on this we can devise a very simple polynomial time transformation that given an acyclic \mathcal{EL} TBox \mathcal{T} outputs and equivalent TBox \mathcal{T}' in reduced form. The translation and its correctness are given in the following lemma.

Lemma 37. *Let \mathcal{T} be a normalized acyclic \mathcal{EL} TBox. The TBox \mathcal{T}' obtained from \mathcal{T} by the substitution of $\alpha \doteq C_{\alpha}$ for $\alpha \doteq (C_{\alpha})^{r(\mathcal{T})}$ (for all $\alpha \doteq C_{\alpha} \in \mathcal{T}$) satisfies the following:*

1. \mathcal{T} and \mathcal{T}' are equivalent.
2. \mathcal{T}' is in reduced form.

Proof. 1) We show that every model of \mathcal{T} is a model of \mathcal{T}' and vice versa. Let \mathcal{I} be a model of \mathcal{T} , then $\alpha^{\mathcal{I}} = (C_{\alpha})^{\mathcal{I}}$ for all $\alpha \doteq C_{\alpha} \in \mathcal{T}$. Since $C_{\alpha} \equiv_{\mathcal{T}} (C_{\alpha})^{r(\mathcal{T})}$, this means that $\alpha^{\mathcal{I}} = \left[(C_{\alpha})^{r(\mathcal{T})} \right]^{\mathcal{I}}$ for all $\alpha \doteq (C_{\alpha})^{r(\mathcal{T})} \in \mathcal{T}'$. Hence, $\mathcal{I} \models \mathcal{T}'$.

Conversely, let \mathcal{I}' be a model of \mathcal{T}' . We take a model \mathcal{I} of \mathcal{T} such that $\Delta^{\mathcal{I}} = \Delta^{\mathcal{I}'}$ and $X^{\mathcal{I}} = X^{\mathcal{I}'}$, for all $X \in \mathbf{N}_{pr}^{\mathcal{T}} \cup \mathbf{N}_{\mathbf{R}}$. Such a model exists because of Proposition 2. We prove that $\alpha^{\mathcal{I}} = \alpha^{\mathcal{I}'}$ for all $\alpha \doteq C_{\alpha} \in \mathcal{T}$. The proof goes by induction on the partial order induced by \rightarrow^+ . Since \mathcal{T} is normalized, each top-level atom of C_{α} is of the form $A \in \mathbf{N}_{pr}^{\mathcal{T}}$ or $\exists r.\beta$, where $\beta \doteq C_{\beta} \in \mathcal{T}$. Moreover, the set of atoms occurring in $(C_{\alpha})^{r(\mathcal{T})}$ is a subset of the corresponding set for C_{α} . Therefore, we distinguish two cases for all top-level atoms At of C_{α} :

- At occurs in $(C_{\alpha})^{r(\mathcal{T})}$. If $\text{At} = A$, by selection of \mathcal{I} we have $A^{\mathcal{I}} = A^{\mathcal{I}'}$. Otherwise, $\text{At} = \exists r.\beta$ and $\alpha \rightarrow^+ \beta$. The application of induction yields $\beta^{\mathcal{I}} = \beta^{\mathcal{I}'}$ and thus $(\exists r.\beta)^{\mathcal{I}} = (\exists r.\beta)^{\mathcal{I}'}$. Hence, it is not hard to see that for all $d \in \Delta^{\mathcal{I}}$, $d \in \alpha^{\mathcal{I}}$ implies $d \in \alpha^{\mathcal{I}'}$.
- At only occurs in C_{α} . There must be a top-level atom At' in C_{α} such that $\text{At}' \sqsubseteq_{\mathcal{T}} \text{At}$ and At' does occur in $(C_{\alpha})^{r(\mathcal{T})}$. From the previous point we know that $(\text{At}')^{\mathcal{I}} = (\text{At}')^{\mathcal{I}'}$. Therefore, if $d \in (\text{At}')^{\mathcal{I}'}$ we also have $d \in (\text{At}')^{\mathcal{I}}$ and $d \in \text{At}'$. Hence, $d \in \alpha^{\mathcal{I}'}$ implies $d \in \alpha^{\mathcal{I}}$.

Thus, we have shown that $\alpha^{\mathcal{I}} = \alpha^{\mathcal{I}'}$. This implies the following equalities:

$$\alpha^{\mathcal{I}'} = \alpha^{\mathcal{I}} = (C_{\alpha})^{\mathcal{I}} = (u_{\mathcal{T}}(C_{\alpha}))^{\mathcal{I}}$$

Then, since \mathcal{I} and \mathcal{I}' have the same interpretation for $\mathbf{N}_{pr}^{\mathcal{T}} \cup \mathbf{N}_{\mathbf{R}}$, this means that $[u_{\mathcal{T}}(C_{\alpha})]^{\mathcal{I}} = [u_{\mathcal{T}}(C_{\alpha})]^{\mathcal{I}'}$. Hence, for all $\alpha \doteq C_{\alpha} \in \mathcal{T}$ we have:

$$\alpha^{\mathcal{I}'} = [u_{\mathcal{T}}(C_{\alpha})]^{\mathcal{I}'}$$

Consequently, $\alpha^{\mathcal{I}'} = (C_{\alpha})^{\mathcal{I}'}$ and \mathcal{I}' is a model of \mathcal{T} .

2) Assume that \mathcal{T}' is not in reduced form. Then, there exists $\alpha \doteq (C_{\alpha})^{r(\mathcal{T})} \in \mathcal{T}'$ such that $(C_{\alpha})^{r(\mathcal{T})}$ is reducible with respect to \mathcal{T}' . This means that there are two top-level atoms At_1 and At_2 in $(C_{\alpha})^{r(\mathcal{T})}$ such that $\text{At}_1 \sqsubseteq_{\mathcal{T}'} \text{At}_2$. Since we just have shown that \mathcal{T} and \mathcal{T}' are equivalent from a model-theoretic point of view, we also have $\text{At}_1 \sqsubseteq_{\mathcal{T}} \text{At}_2$. Hence, we obtain a contradiction against the fact that $(C_{\alpha})^{r(\mathcal{T})}$ is reduced with respect to \mathcal{T} . Thus, \mathcal{T}' is in reduced form. \square

To sum up, given an acyclic $\tau\mathcal{EL}(\text{deg})$ TBox $\widehat{\mathcal{T}} = (\mathcal{T}_{\tau}, \mathcal{T})$, we have demonstrated the following along this section:

- $\widehat{\mathcal{T}}$ can be normalized in polynomial time into an acyclic TBox $\widehat{\mathcal{T}}' = (\mathcal{T}'_{\tau}, \mathcal{T}')$, such that reasoning w.r.t. $\widehat{\mathcal{T}}$ can be reduced to reasoning w.r.t. $\widehat{\mathcal{T}}'$.
- The new TBox \mathcal{T}' can be translated in polynomial time into an equivalent \mathcal{EL} TBox \mathcal{T}'' in reduced form.
- The computation of the reduced form only removes atoms from concept definitions. Therefore, \mathcal{T}'' remains normalized.

Hence, reasoning in $\tau\mathcal{EL}(\text{deg})$ with respect to acyclic TBoxes can be restricted to normalized acyclic TBoxes in reduced form.

Proposition 38. *Satisfiability and subsumption on concepts defined in an acyclic $\tau\mathcal{EL}(\text{deg})$ TBox can be reduced in polynomial time to satisfiability and subsumption on concepts defined in a normalized acyclic $\tau\mathcal{EL}(\text{deg})$ TBox in reduced form.*

4.3 A PSpace upper bound

We now present a PSpace procedure that decides satisfiability of concepts of the form $\alpha_1 \sqcap \neg\alpha_2$ w.r.t. an acyclic $\tau\mathcal{EL}(\text{deg})$ TBox $\widehat{\mathcal{T}}$, where $\alpha_1, \alpha_2 \in \mathbf{N}_d^{\widehat{\mathcal{T}}}$. The restriction to defined concepts is without loss of generality, since every $\tau\mathcal{EL}(m)$ concept \widehat{C} correctly defined w.r.t. $\widehat{\mathcal{T}}$ can be equivalently replaced with a fresh concept name $\alpha_{\widehat{C}}$, by adding $\alpha_{\widehat{C}} \doteq \widehat{C}$ to \mathcal{T}_{τ} .

As mentioned earlier, by using unfolding, we can reduce our problem to satisfiability of the concept $u_{\widehat{\mathcal{T}}}(\alpha_1) \sqcap \neg u_{\widehat{\mathcal{T}}}(\alpha_2)$. Therefore, the application of Lemma 16

yields that $\alpha_1 \sqcap \neg\alpha_2$ is satisfiable in $\widehat{\mathcal{T}}$ iff there exists an interpretation \mathcal{I} over $\mathbf{N}_{pr}^{\widehat{\mathcal{T}}} \cup \mathbf{N}_R$ such that:

$$[u_{\widehat{\mathcal{T}}}(\alpha_1)]^{\mathcal{I}} \setminus [u_{\widehat{\mathcal{T}}}(\alpha_2)]^{\mathcal{I}} \neq \emptyset \quad \text{and} \quad |\Delta^{\mathcal{I}}| \leq \mathbf{s}(u_{\widehat{\mathcal{T}}}(\alpha_1)) \cdot \mathbf{s}(u_{\widehat{\mathcal{T}}}(\alpha_2))$$

Since $u_{\widehat{\mathcal{T}}}(\alpha_1)$ or $u_{\widehat{\mathcal{T}}}(\alpha_2)$ may be concepts of size exponential in $\mathbf{s}(\widehat{\mathcal{T}})$, this gives an exponential bounded model property, and hence the mentioned NExpTime upper bound. However, the construction used to prove Lemma 16 in [BBG15] provides additional information about \mathcal{I} , which allows us to improve on this upper bound:

- \mathcal{I} is *tree-shaped*,
- the depth of its description tree $T_{\mathcal{I}}$ is bounded by:

$$\text{rd}(u_{\widehat{\mathcal{T}}}(\alpha_1)) + \text{rd}(u_{\widehat{\mathcal{T}}}(\alpha_2)) \tag{22}$$

- the element $d_0 \in \Delta^{\mathcal{I}}$ corresponding to the root of $T_{\mathcal{I}}$ satisfies

$$d_0 \in [u_{\widehat{\mathcal{T}}}(\alpha_1) \sqcap \neg u_{\widehat{\mathcal{T}}}(\alpha_2)]^{\mathcal{I}}$$

Fortunately, the depth (22) of $T_{\mathcal{I}}$ is always polynomial in $\mathbf{s}(\widehat{\mathcal{T}})$. Thus, despite its size, one can non-deterministically generate \mathcal{I} in a top-down fashion, while keeping the used space polynomial in $\mathbf{s}(\widehat{\mathcal{T}})$. Let $\mathfrak{d} \geq 0$ and $\mathfrak{b} > 0$ be natural numbers. Then, each run ρ of the procedure *Gen* described below generates a *tree-shaped* interpretation \mathcal{I}_{ρ} over $\mathbf{N}_{pr}^{\widehat{\mathcal{T}}} \cup \mathbf{N}_R$, such that $|\Delta^{\mathcal{I}_{\rho}}| \leq \mathfrak{b}$ and the depth of $T_{\mathcal{I}_{\rho}}$ is not greater than \mathfrak{d} :

```

1: procedure Gen( $\mathfrak{d} : \mathbb{N}$ ,  $\mathfrak{b} : \text{binary}$ )
2:    $\mathfrak{b} := \mathfrak{b} - 1$ 
3:   non-deterministically choose a subset  $\mathcal{P}$  of  $\mathbf{N}_{pr}^{\widehat{\mathcal{T}}}$ 
4:   if ( $\mathfrak{d} \neq 0$ ) and ( $\mathfrak{b} \neq 0$ ) then
5:     for all  $r \in \mathbf{N}_R$  do
6:       non-deterministically choose  $0 \leq \mathfrak{b}_r \leq \mathfrak{b}$ 
7:        $\mathfrak{b} := \mathfrak{b} - \mathfrak{b}_r$ 
8:       for all  $1 \leq i \leq \mathfrak{b}_r$  do
9:         non-deterministically choose  $0 \leq \mathfrak{b}_r^i \leq \mathfrak{b}$ 
10:         $\mathfrak{b} := \mathfrak{b} - \mathfrak{b}_r^i$ 
11:        Gen( $\mathfrak{d} - 1$ ,  $\mathfrak{b}_r^i + 1$ )
12:       end for
13:     end for
14:   end if
15: end procedure

```

Note that each recursive call decreases the value of \mathfrak{d} , and therefore it is a *terminating* procedure executing at most \mathfrak{d} nested recursive calls. Moreover, as

evidenced by the parameter declaration $\mathfrak{b} : \text{binary}$, Gen works with the binary representation of the value \mathfrak{b} (similarly for variables \mathfrak{b}_r and \mathfrak{b}_r^i). Finally, the set of variables \mathfrak{b}_r and \mathfrak{b}_r^i can be reduced to two variables since they are only used within the scope of the **for** loops. Therefore, each run of Gen uses space polynomial on \mathfrak{d} and the number of bits needed to represent \mathfrak{b} .

The general idea of the procedure is as follows: each recursive call represents an individual of $\Delta^{\mathcal{I}_\rho}$ and the recursion tree lays out the tree-shaped form of \mathcal{I}_ρ . The set \mathcal{P} contains the primitive concept names that a domain element is an instance of, the number \mathfrak{b}_r stands for the number of r -successors, and \mathfrak{b}_r^i means that the interpretation rooted at the i -th r -successor has at most $\mathfrak{b}_r^i + 1$ elements. To formalize this intuition we define the notion of a run of Gen .

Definition 39. A run ρ of Gen on $(\mathfrak{d}, \mathfrak{b})$ is a tree of recursive calls $T_{(\mathfrak{d}, \mathfrak{b})}$ such that:

- its root v_0 is labeled by the non-deterministic choices \mathcal{P} , \mathfrak{b}_r for all $r \in \mathbf{N}_R$, and \mathfrak{b}_r^i for all $1 \leq i \leq \mathfrak{b}_r$.
- for all $r \in \mathbf{N}_R$, there are exactly \mathfrak{b}_r successors $v_{r1}, \dots, v_{r\mathfrak{b}_r}$ of v_0 such that, the tree rooted at v_{ri} is a run of Gen on $(\mathfrak{d} - 1, \mathfrak{b}_r^i + 1)$.

Figure 3 depicts a run ρ of Gen (left-hand side). Such a run induces the \mathcal{EL} description tree T_ρ (right-hand side) with the same structure, where its nodes are labeled with the corresponding sets \mathcal{P} chosen by ρ and the edges with the role names generating the corresponding recursive call (line 5 in Gen). Therefore, we say that ρ induces the interpretation \mathcal{I}_ρ that has the description tree T_ρ .

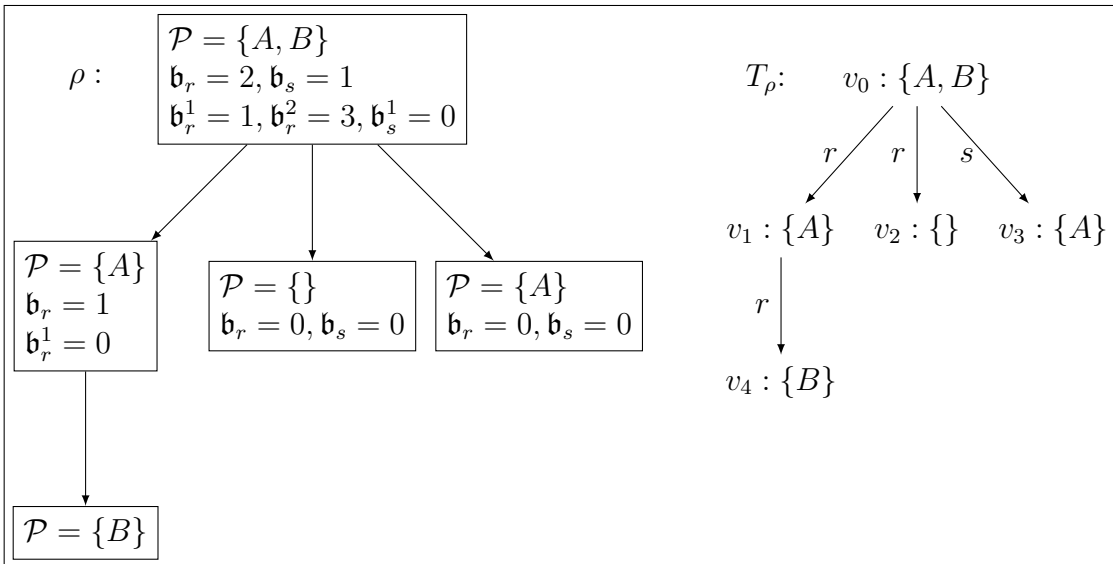


Figure 3: A run ρ of Gen on $(2, 5)$ and its induced \mathcal{EL} description tree T_ρ .

Conversely, for all tree-shaped interpretations \mathcal{I} of size at most \mathfrak{b} and depth not greater than \mathfrak{d} , there is always a run of *Gen* describing \mathcal{I} .

Lemma 40. *Let $\mathfrak{d} \geq 0$ and $\mathfrak{b} > 0$ be two natural numbers. For all tree-shaped interpretations \mathcal{I} over $\mathbf{N}_{pr}^{\widehat{\mathcal{T}}} \cup \mathbf{N}_R$ with $|\Delta^{\mathcal{I}}| \leq \mathfrak{b}$ and depth not greater than \mathfrak{d} , there is a run ρ of *Gen* on $(\mathfrak{d}, \mathfrak{b})$ such that $\mathcal{I} = \mathcal{I}_\rho$.*

Proof. Let \mathcal{I} be a tree-shaped interpretation of depth $d(\mathcal{I})$ such that $|\Delta^{\mathcal{I}}| \leq \mathfrak{b}$ and $d(\mathcal{I}) \leq \mathfrak{d}$. We show how to guide a run ρ of *Gen* such that $\mathcal{I}_\rho = \mathcal{I}$. The proof goes by induction on the number $d(\mathcal{I})$.

Let $d_0 \in \Delta^{\mathcal{I}}$ be the root of $T_{\mathcal{I}}$. For all $r \in \mathbf{N}_R$ we denote as $r(d_0) = \{e_1, \dots, e_n\}$ ($n \geq 0$) the set of r -successors of d_0 in \mathcal{I} . In addition, for an r -successor e_i of d_0 , $T_{\mathcal{I}}[e_i]$ denotes the subtree of $T_{\mathcal{I}}$ rooted at e_i , and \mathcal{I}_{e_i} the associated interpretation. Then, when *Gen* is invoked on $(\mathfrak{d}, \mathfrak{b})$ it makes the following non-deterministic choices:

- $\mathcal{P} = \ell_{\mathcal{I}}(d_0)$,
- for all $r \in \mathbf{N}_R$: $b_r = |r(d_0)|$,
- for all $r \in \mathbf{N}_R$ and $e_i \in r(d_0)$: $\mathfrak{b}_r^i = |\Delta^{\mathcal{I}_{e_i}}| - 1$,
- for all $r \in \mathbf{N}_R$ and $1 \leq i \leq \mathfrak{b}_r$, the recursive call *Gen*($\mathfrak{d} - 1, \mathfrak{b}_r^i + 1$) follows a run ρ_r^i such that $\mathcal{I}_{\rho_r^i} = \mathcal{I}_{e_i}$.

Since $|\Delta^{\mathcal{I}}| \leq \mathfrak{b}$ and $d(\mathcal{I}) \leq \mathfrak{d}$, the first three choices are consistent with the execution of *Gen*. Regarding the last choice, since $T_{\mathcal{I}}$ is a tree we know that $d(\mathcal{I}_{e_i}) < d(\mathcal{I})$. Consequently, $d(\mathcal{I}_{e_i}) \leq \mathfrak{d} - 1$ and the induction hypothesis can be applied to obtain the proper run ρ_r^i . Therefore, ρ induces an \mathcal{EL} description tree T_ρ such that:

- its root v_0 is labeled with $\ell_{\mathcal{I}}(d_0)$,
- for all $r \in \mathbf{N}_R$: v_0 has exactly $|r(d_0)|$ children $v_1, \dots, v_{|r(d_0)|}$, each edge (v_0, v_i) ($1 \leq i \leq |r(d_0)|$) is labeled with r , and the subtree $T_\rho[v_i]$ rooted at v_i in T_ρ is equal to $T_{\mathcal{I}}[e_i]$.

Thus, we can conclude that $\mathcal{I}_\rho = \mathcal{I}$. □

Lemma 40 ensures that, by choosing \mathfrak{d} as in (22) and \mathfrak{b} as $\mathfrak{s}(u_{\widehat{\mathcal{T}}}(\alpha_1)) \cdot \mathfrak{s}(u_{\widehat{\mathcal{T}}}(\alpha_2))$, the set of runs of *Gen* on $(\mathfrak{d}, \mathfrak{b})$ covers a set of interpretations that suffices to find out if $u_{\widehat{\mathcal{T}}}(\alpha_1) \sqcap \neg u_{\widehat{\mathcal{T}}}(\alpha_2)$ is satisfiable. Hence, it remains to see how to check for a run ρ of *Gen*, whether $d_0 \in [u_{\widehat{\mathcal{T}}}(\alpha_1) \sqcap \neg u_{\widehat{\mathcal{T}}}(\alpha_2)]^{\mathcal{I}_\rho}$. It turns out, however, that the unique extension of \mathcal{I}_ρ satisfying $\widehat{\mathcal{T}}$ (recall Proposition 21) can actually

be computed along the run ρ . The idea is that being \mathcal{I}_ρ tree-shaped, such an extension will be computed in a *bottom-up* manner. Therefore, by doing that we simply need to check whether $d_0 \in (\alpha_1 \sqcap \neg\alpha_2)^{\mathcal{I}_\rho}$.

To this end, we transform the *procedure* Gen into a function $Gen+$ such that each run ρ additionally computes a set $\text{Ex} \subseteq \mathbf{N}_d^{\widehat{\mathcal{T}}}$ with the following meaning:

$$\text{Ex} := \{\alpha \mid \alpha \in \mathbf{N}_d^{\widehat{\mathcal{T}}} \text{ and } d_0 \in \alpha^{\mathcal{I}_\rho}\}$$

The special forms introduced in Section 4.2 for acyclic TBoxes are of great help in computing Ex . In particular, the normal form of $\widehat{\mathcal{T}}$ provides the following shape for \widehat{C}_α :

$$\widehat{P}_1 \sqcap \dots \sqcap \widehat{P}_k \sqcap \exists r_1.\alpha_1 \sqcap \dots \sqcap \exists r_n.\alpha_n$$

Consequently for all $d \in \Delta^{\mathcal{I}_\rho}$, $d \in \alpha^{\mathcal{I}_\rho}$ iff:

1. $d \in (\widehat{P}_i)^{\mathcal{I}_\rho}$ for all $1 \leq i \leq k$, and
2. for all $1 \leq i \leq n$, there exists $d_i \in \Delta^{\mathcal{I}_\rho}$ such that $(d, d_i) \in (r_i)^{\mathcal{I}_\rho}$ and $d_i \in (\alpha_i)^{\mathcal{I}_\rho}$.

The computation of Ex will be based on checking these two conditions for d_0 . If \widehat{P}_i is of the form $A \in \mathbf{N}_{pr}^{\widehat{\mathcal{T}}}$, verifying whether $d_0 \in A^{\mathcal{I}_\rho}$ is simple since \mathcal{I}_ρ already contains that information (the non-deterministic choice in line 3). To check whether $d_0 \in (E_{\sim t})^{\mathcal{I}_\rho}$, we further extend $Gen+$ to compute for all runs ρ an assignment $D : \mathbf{N}_d^{\widehat{\mathcal{T}}} \rightarrow [0, 1]$ such that:

$$D(E) := \text{deg}^{\mathcal{I}_\rho}(d_0, u_{\mathcal{T}}(E))$$

Once D is computed for d_0 , it is immediate to verify whether $d_0 \in (E_{\sim t})^{\mathcal{I}_\rho}$. Regarding Condition 2, as explained before the successors e of d_0 in \mathcal{I}_ρ are the roots of the interpretations induced by runs corresponding to the recursive calls triggered by ρ . Hence, the sets Ex_e computed by such calls provide the necessary information to determine whether $d_0 \in (\exists r_i.\alpha_i)^{\mathcal{I}_\rho}$ for all $1 \leq i \leq n$. However, since d_0 may have exponentially many *direct* successors in \mathcal{I}_ρ , a PSpace procedure cannot store all the corresponding sets Ex_e . To deal with this, $Gen+$ will compute a relation of the form $z \subseteq (\mathbf{N}_R \times \mathbf{N}_d^{\widehat{\mathcal{T}}}) \cup (\epsilon \times \mathbf{N}_{pr}^{\widehat{\mathcal{T}}})$ such that: $(r, \alpha) \in z$ iff there is $e \in \Delta^{\mathcal{I}_\rho}$ satisfying $(d_0, e) \in r^{\mathcal{I}_\rho}$ and $\alpha \in \text{Ex}_e$. In this way we can keep the relevant information needed to verify whether $d_0 \in (\exists r_i.\alpha_i)^{\mathcal{I}_\rho}$, while using polynomial space.

Putting all these ideas together, we transform Gen into $Gen+$ as follows:

- 1: **function** $Gen+(\mathfrak{d} : \text{integer}, \mathfrak{b} : \text{binary})$
- 2: $\mathfrak{b} := \mathfrak{b} - 1$
- 3: *non-deterministically* choose a subset \mathcal{P} of $\mathbf{N}_{pr}^{\widehat{\mathcal{T}}}$

```

4:   initialize  $v$  and  $z$ 
5:   if ( $\mathfrak{d} \neq 0$ ) and( $\mathfrak{b} \neq 0$ ) then
6:     for all  $r \in \mathbb{N}_R$  do
7:       non-deterministically choose  $0 \leq \mathfrak{b}_r \leq \mathfrak{b}$ 
8:        $\mathfrak{b} := \mathfrak{b} - \mathfrak{b}_r$ 
9:       for all  $1 \leq i \leq \mathfrak{b}_r$  do
10:        non-deterministically choose  $0 \leq \mathfrak{b}_r^i \leq \mathfrak{b}$ 
11:         $\mathfrak{b} := \mathfrak{b} - \mathfrak{b}_r^i$ 
12:         $(\text{Ex}_r^i, D_r^i) := \text{Gen}+(\mathfrak{d} - 1, \mathfrak{b}_r^i + 1)$ 
13:        update  $v$ 
14:        update  $z$ 
15:      end for
16:    end for
17:  end if
18:   $D := \text{SUBDEG}(v)$ 
19:   $\text{Ex} := \text{SUBEX}(D, z)$ 
20:  return  $(\text{Ex}, D)$ 
21: end function

```

The subroutines SUBDEG and SUBEX invoked in lines 18 and 19 correspond to the computation of the assignment D and the set Ex , respectively. The execution of line 14 updates the relation z using the content of Ex_r^i after each recursive call has been executed. Regarding the symbol v in line 13, as we explain below it represents a table used to help the computation of D .

Let us now move on to the details of the computation of Ex and D . We start with the computation of D , and afterwards explain how to compute Ex .

Due to the normal form of $\widehat{\mathcal{T}}$, the \mathcal{EL} concept E in $E_{\sim t}$ is a defined concept in \mathcal{T} . Therefore, by Definition 20 for all $d \in \Delta^{\mathcal{I}\rho}$:

$$d \in (E_{\sim t})^{\mathcal{I}\rho} \text{ iff } \text{deg}^{\mathcal{I}\rho}(d, u_{\mathcal{T}}(E)) \sim t$$

Coming back to Chapter 2.3 we know that $\text{deg}^{\mathcal{I}\rho}(d, u_{\mathcal{T}}(E))$ is the maximal value of $h_w(v_0)$ among all *ptghs* $h \in \mathcal{H}(T_{u_{\mathcal{T}}(E)}, G_{\mathcal{I}\rho}, d)$, where v_0 is the root of the description tree $T_{u_{\mathcal{T}}(E)}$. Note that we use directly $T_{u_{\mathcal{T}}(E)}$, since being \mathcal{T} in reduced form implies that $u_{\mathcal{T}}(E)$ is reduced (see Lemma 36). Now, E is defined in \mathcal{T} as follows:

$$E \doteq P_1 \sqcap \dots \sqcap P_q \sqcap \exists r_1. E_1 \sqcap \dots \sqcap \exists r_n. E_n$$

This gives us the following information regarding $T_{u_{\mathcal{T}}(E)}$:

- the label of v_0 in $T_{u_{\mathcal{T}}(E)}$ is the set $\{P_1, \dots, P_q\}$,
- v_0 has exactly n ($n \geq 0$) successors v_1, \dots, v_n in $T_{u_{\mathcal{T}}(E)}$,
- for all $1 \leq i \leq n$, the subtree $T_{u_{\mathcal{T}}(E)}[v_i]$ of $T_{u_{\mathcal{T}}(E)}$ rooted at v_i is exactly the description tree associated to $u_{\mathcal{T}}(E_i)$.

Additionally, the computation of $h_w(v_0)$ is based on the following expression:

$$h_w(v_0) = \begin{cases} 1 & \text{if } q + n = 0 \\ \frac{|\{P_1, \dots, P_q\} \cap \ell_{\mathcal{I}_\rho}(d)| + \sum_{1 \leq i \leq k} h_w(w_i)}{q+n} & \text{otherwise.} \end{cases}$$

where w_1, \dots, w_k are the children of v_0 in $T_{u_{\mathcal{T}}(E)}$ mapped by h . Now, regarding a *ptgh* h yielding a maximal value for $h_w(v_0)$ we observe the following:

- if $(d, e) \in (r_i)^{\mathcal{I}_\rho}$ for some $e \in \Delta^{\mathcal{I}_\rho}$, then we can assume that $v_i \in \text{dom}(h)$.
- Let $h(v_i) = e_i$ where $e_i \in \Delta^{\mathcal{I}_\rho}$. Then, $h_w(v_i) = \text{deg}^{\mathcal{I}_\rho}(e_i, u_{\mathcal{T}}(E_i))$. This is a consequence of v_i being the root of the description tree corresponding to $u_{\mathcal{T}}(E_i)$, and the fact that $h_w(v_0)$ is maximal.

Therefore, $\text{deg}^{\mathcal{I}_\rho}(d, u_{\mathcal{T}}(E))$ can be expressed as:

$$\frac{|\{P_1, \dots, P_q\} \cap \ell_{\mathcal{I}_\rho}(d)| + \sum_{i=1}^n \max\{\text{deg}^{\mathcal{I}_\rho}(e, u_{\mathcal{T}}(E_i)) \mid (d, e) \in (r_i)^{\mathcal{I}_\rho}\}}{q+n} \quad (23)$$

Thus, knowing the values $\text{deg}^{\mathcal{I}_\rho}(e, u_{\mathcal{T}}(F))$ for all successors e of d in \mathcal{I}_ρ and all $F \in \{E_1, \dots, E_n\}$, it is straightforward to compute $\text{deg}^{\mathcal{I}_\rho}(d, u_{\mathcal{T}}(E))$. Therefore, similar to the computation of **Ex** the assignment D for d_0 can be computed by using all the assignments D recursively computed for all successors of d_0 in \mathcal{I}_ρ . Once more, the problem related to the possible exponentially many successors of d_0 needs to be addressed. Here is where the aforementioned table v comes into play. It is defined as $v : (\mathbf{N}_R \times \mathbf{N}_d^{\mathcal{T}}) \cup (\epsilon \times \mathbf{N}_{pr}^{\widehat{\mathcal{T}}}) \rightarrow [0, 1]$ and each entry $v[r, E]$ stores the value $\max\{D_e(E) \mid (d_0, e) \in r^{\mathcal{I}_\rho}\}$, where D_e is the assignment D for e , and $v[\epsilon, P] = 1$ iff $P \in \mathcal{P}$ (0 otherwise). The following fragment of *pseudo-code* updates v within a run of *Gen+*:

- 1: $v[r, E] = 0$ for all $(r, E) \in (\mathbf{N}_R \times \mathbf{N}_d^{\mathcal{T}}) \cup (\epsilon \times \mathbf{N}_{pr}^{\widehat{\mathcal{T}}})$ // Initialization
- 2: $v[\epsilon, P] = 1$ iff $P \in \mathcal{P}$
- 3: \vdots
- 4: $D_r^i := \text{Gen+}(\mathfrak{d} - 1, \mathfrak{b}_r^i + 1)$
- 5: **for all** $(E \doteq C_E \in \mathcal{T})$ **do**
- 6: **if** $D_r^i(E) > v[r, E]$ **then**
- 7: $v[r, E] := D_r^i(E)$
- 8: **end if**
- 9: **end for**

Here, D_r^i stands for the assignment D corresponding to the root element of the interpretation induced by the recursive call. In other words, the i -th r -successor of d_0 in \mathcal{I}_ρ . After all the recursive calls have been executed, v is used to compute D as described in the following subroutine:

```

procedure SUBDEG( $v : (\mathbf{N}_R \times \mathbf{N}_d^T) \cup (\epsilon \times \mathbf{N}_{pr}^{\widehat{T}}) \rightarrow [0, 1]$ )
  for all ( $E \doteq C_E \in \mathcal{T}$ ) do
     $c := |\{P \mid P \in \text{tp}(C_E) \text{ and } v[\epsilon, P] = 1\}|$ 
    for all  $\exists r. E' \in \text{tp}(C_E)$  do
       $c := c + v[r, E']$ 
    end for
     $D(E) := \frac{c}{|\text{tp}(C_E)|}$ 
  end for
  return  $D$ 
end procedure

```

It remains to see the details of the computation of Ex. The updating of the relation z in *Gen+* is carried out as follows:

```

1:  $z := \{(\epsilon, P) \mid P \in \mathcal{P}\}$  // Initialization
2:  $\vdots$ 
3:  $\text{Ex}_r^i := \text{Gen+}(\mathfrak{d} - 1, \mathfrak{b}_r^i + 1)$ 
4: for all ( $\alpha \doteq \widehat{C}_\alpha \in \mathcal{T}_\tau \cup \mathcal{T}$ ) do
5:   if  $\alpha \in \text{Ex}_r^i$  then
6:      $z := z \cup \{(r, \alpha)\}$ 
7:   end if
8: end for

```

Then, using D and z Conditions 1 and 2 can be verified, and Ex can be computed in the following way:

```

procedure SUBEX( $D : \mathbf{N}_d^T \rightarrow [0, 1]$ ,  $z \subseteq (\mathbf{N}_R \times \mathbf{N}_d^T) \cup (\epsilon \times \mathbf{N}_{pr}^{\widehat{T}})$ )
   $s := \emptyset$ 
  for all ( $\alpha \doteq \widehat{C}_\alpha \in \mathcal{T}_\tau \cup \mathcal{T}$ ) do
    if ( $[P \in \text{tp}(\widehat{C}_\alpha)] \Rightarrow (\epsilon, P) \in z$ ) and ( $[E \sim t \in \text{tp}(\widehat{C}_\alpha)] \Rightarrow D(E) \sim t$ ) and
      ( $[\exists r. \beta \in \text{tp}(\widehat{C}_\alpha)] \Rightarrow (r, \beta) \in z$ ) then
         $s := s \cup \{\alpha\}$ 
      end if
    end for
  return  $s$ 
end procedure

```

Thus, using the function *Gen+* we define our non-deterministic algorithm to decide satisfiability of concepts of the form $\alpha_1 \sqcap \neg \alpha_2$ with respect to acyclic $\tau\mathcal{EL}(\text{deg})$ TBoxes.

Since *Gen+* terminates, this implies that Algorithm 1 terminates as well. In the following, we show that Algorithm 1 is sound and complete. Let us start by showing that *Gen+* computes the right values for D and Ex.

Lemma 41. *Let $\mathfrak{d} \geq 0$ and $\mathfrak{b} > 0$ be two natural numbers, and ρ be a run of *Gen+* on $(\mathfrak{d}, \mathfrak{b})$. Then,*

Algorithm 1 Satisf. of $\alpha_1 \sqcap \neg\alpha_2$ w.r.t. acyclic $\tau\mathcal{EL}(deg)$ TBoxes.

Input: An acyclic $\tau\mathcal{EL}(deg)$ TBox $\widehat{\mathcal{T}}$ and $\alpha_1, \alpha_2 \in \mathbf{N}_d^{\widehat{\mathcal{T}}}$.

Output: “yes”, if $\alpha_1 \sqcap \neg\alpha_2$ is satisfiable in $\widehat{\mathcal{T}}$, “no” otherwise.

```

1:  $\mathbf{b} := \mathbf{s}(u_{\widehat{\mathcal{T}}}(\alpha_1)) \cdot \mathbf{s}(u_{\widehat{\mathcal{T}}}(\alpha_2))$  //  $\mathbf{b}$  is stored in binary
2:  $\mathbf{d} := \mathbf{rd}(u_{\widehat{\mathcal{T}}}(\alpha_1)) + \mathbf{rd}(u_{\widehat{\mathcal{T}}}(\alpha_2))$ 
3:  $(\mathbf{Ex}, D) := \mathit{Gen}+(\mathbf{d}, \mathbf{b})$ 
4: if  $\alpha_1 \in \mathbf{Ex}$  and  $\alpha_2 \notin \mathbf{Ex}$  then
5:   return “yes”
6: end if
7: return “no”

```

1. $D(E) = deg^{\mathcal{I}_\rho}(d_0, u_{\mathcal{T}}(E))$, for all $E \doteq C_E \in \mathcal{T}$.

2. $\mathbf{Ex} = \{\alpha \mid \alpha \in \mathbf{N}_d^{\widehat{\mathcal{T}}} \text{ and } d_0 \in \alpha^{\mathcal{I}_\rho}\}$

Proof. Let $\mathbf{d}(\mathcal{I}_\rho)$ denote the depth of $T_{\mathcal{I}_\rho}$. We prove our claims by induction on $\mathbf{d}(\mathcal{I}_\rho)$. To start, we fix a role name $r \in \mathbf{N}_R$ and define $r(d_0) = \{e_1, \dots, e_n\}$ to be the set of r -successors of d_0 in \mathcal{I}_ρ (with $n \geq 0$). By construction of $T_{\mathcal{I}_\rho}$, ρ does exactly n recursive calls $\mathit{Gen}+(\mathbf{d} - 1, \mathbf{b}_r^i)$ ($1 \leq i \leq n$). Let ρ_r^i denote the run corresponding to the i -th call. Then, the interpretation $\mathcal{I}_{\rho_r^i}$ induced by ρ_r^i is the one having the description tree $T_{\mathcal{I}_{\rho_r^i}}[e_i]$, i.e., the subtree of $T_{\mathcal{I}_\rho}$ rooted at e_i .

The tree shape of \mathcal{I}_ρ implies that $\mathbf{d}(\mathcal{I}_{\rho_r^i}) < \mathbf{d}(\mathcal{I}_\rho)$. Therefore, induction hypothesis can be applied to obtain:

$$D_r^i(E) = deg^{\mathcal{I}_{\rho_r^i}}(e_i, u_{\mathcal{T}}(E))$$

$$\mathbf{Ex}_r^i = \{\alpha \mid \alpha \in \mathbf{N}_d^{\widehat{\mathcal{T}}} \text{ and } e_i \in \alpha^{\mathcal{I}_{\rho_r^i}}\}$$

The same reasoning applies for all the other role names $s \in \mathbf{N}_R$. Note that since $\mathcal{I}_{\rho_r^i}$ is a subtree of \mathcal{I}_ρ , those two equalities are also valid for \mathcal{I}_ρ , i.e.:

$$D_r^i(E) = deg^{\mathcal{I}_\rho}(e_i, u_{\mathcal{T}}(E))$$

$$\mathbf{Ex}_r^i = \{\alpha \mid \alpha \in \mathbf{N}_d^{\widehat{\mathcal{T}}} \text{ and } e_i \in \alpha^{\mathcal{I}_\rho}\}$$

Therefore, after all the recursive calls have been executed and the values in table v and relation z have been fully updated, we have for all $(r, E) \in \mathbf{N}_R \times \mathbf{N}_d^{\mathcal{T}}$:

$$v[r, E] = \max\{deg^{\mathcal{I}_\rho}(e, u_{\mathcal{T}}(E)) \mid (d_0, e) \in r^{\mathcal{I}_\rho}\} \quad (24)$$

and,

$$z = \{(r, \alpha) \mid e \in \Delta^{\mathcal{I}_\rho}, (d_0, e) \in r^{\mathcal{I}_\rho} \text{ and } e \in \alpha^{\mathcal{I}_\rho}\} \quad (25)$$

Looking at the subroutine SUBDEG, for all $E \doteq C_E \in \mathcal{T}$ the value $D(E)$ is computed by the following expression:

$$D(E) = \frac{|\mathbf{tp}(C_E) \cap \mathcal{P}| + \sum_{\exists r. E' \in \mathbf{tp}(C_E)} v[r, E']}{\mathbf{tp}(C_E)}$$

Now, by construction of \mathcal{I}_ρ we have that $\ell_{\mathcal{I}_\rho}(d_0) = \mathcal{P}$. Hence, replacing $v[r, E']$ by the right-hand side of the equality in (24) we obtain the expression in (23). Consequently, we have shown that:

$$D(E) = \text{deg}^{\mathcal{I}_\rho}(d_0, u_{\mathcal{T}}(E))$$

Last, let $\alpha \doteq \widehat{C}_\alpha \in \mathcal{T}_\tau \cup \mathcal{T}$ with \widehat{C}_α of the form:

$$\widehat{P}_1 \sqcap \dots \sqcap \widehat{P}_q \sqcap \exists r_1.\alpha_1 \sqcap \dots \sqcap \exists r_n.\alpha_n$$

According to SUBEX, $\alpha \in \text{Ex}$ iff:

- for all $1 \leq i \leq q$: if \widehat{P}_i is of the form $E \sim t$ then $D(E) \sim t$, otherwise $\widehat{P}_i \in \mathcal{P}$, and
- $(r_j, \alpha_j) \in z$, for all $1 \leq j \leq n$.

Since $\ell_{\mathcal{I}_\rho}(d_0) = \mathcal{P}$ and $D(E) = \text{deg}^{\mathcal{I}_\rho}(d_0, u_{\mathcal{T}}(E))$, the first statement is equivalent to have $d_0 \in (\widehat{P}_i)^{\mathcal{I}_\rho}$ ($1 \leq i \leq q$). Furthermore, (25) makes the second statement equivalent to having $d_0 \in (\exists r_j.\alpha_j)^{\mathcal{I}_\rho}$ ($1 \leq j \leq n$). Thus, $\alpha \in \text{Ex}$ iff $d_0 \in \alpha^{\mathcal{I}_\rho}$.

Note that the base case for the induction is already contained in the proof. \square

Using Lemma 41 we now prove that Algorithm 1 is sound and complete.

Lemma 42. *Let $\widehat{\mathcal{T}}$ be an acyclic $\tau\mathcal{EL}(\text{deg})$ TBox and α_1, α_2 two defined concepts in $\widehat{\mathcal{T}}$. Then,*

Algorithm 1 answers “yes” iff $\alpha_1 \sqcap \neg\alpha_2$ is satisfiable in $\widehat{\mathcal{T}}$.

Proof. (\Rightarrow) Suppose that the algorithm gives a positive answer and let ρ be the run of function *Gen+* that leads to it. Then, we can talk about the interpretation \mathcal{I}_ρ induced by ρ . The “yes” answer means that for ρ , $\alpha_1 \in \text{Ex}$ and $\alpha_2 \notin \text{Ex}$. Then, the application of Lemma 41 yields:

$$d_0 \in (\alpha_1 \sqcap \neg\alpha_2)^{\mathcal{I}_\rho}$$

with $d_0 \in \Delta^{\mathcal{I}_\rho}$. Hence, $\alpha_1 \sqcap \neg\alpha_2$ is satisfiable with respect to $\widehat{\mathcal{T}}$.

(\Leftarrow) Assume that $\alpha_1 \sqcap \neg\alpha_2$ is satisfiable with respect to $\widehat{\mathcal{T}}$. This means that there exists an interpretation \mathcal{I} such that $\mathcal{I} \models \widehat{\mathcal{T}}$ and $(\alpha_1 \sqcap \neg\alpha_2)^{\mathcal{I}} \neq \emptyset$. By the bounded model property discussed at the beginning of this section and the subsequent remarks, one can assume that \mathcal{I} is tree-shaped and satisfies the following properties:

1. $\Delta^{\mathcal{I}}$ has at most $\mathfrak{s}(u_{\widehat{\mathcal{T}}}(\alpha_1)) \cdot \mathfrak{s}(u_{\widehat{\mathcal{T}}}(\alpha_2))$ elements,

2. the depth of $T_{\mathcal{I}}$ is not greater than $\text{rd}(u_{\widehat{\mathcal{T}}}(\alpha_1)) + \text{rd}(u_{\widehat{\mathcal{T}}}(\alpha_2))$, and
3. its root element d_0 satisfies: $d_0 \in (\alpha_1 \sqcap \neg\alpha_2)^{\mathcal{I}}$.

The selection of \mathfrak{d} and \mathfrak{b} in Algorithm 1 and the application of Lemma 40 guarantee the existence of a run ρ of $Gen+$ on $(\mathfrak{d}, \mathfrak{b})$ generating the restriction of \mathcal{I} to $\mathbf{N}_{pr}^{\widehat{\mathcal{T}}} \cup \mathbf{N}_{\mathbf{R}}$. Hence, the application of Lemma 41 implies that the conditional in line 4 must evaluate to *true* for such a run ρ . Thus, Algorithm 1 answers “yes”. \square

Algorithm 1 uses space polynomial in the size of $\widehat{\mathcal{T}}$ to store the binary representation of \mathfrak{b} . Furthermore, z and v are also stored within polynomial space, and the two subroutines run in polynomial time. Therefore, since each run ρ of $Gen+$ on $(\mathfrak{d}, \mathfrak{b})$ does at most \mathfrak{d} many nested recursive calls, ρ uses space polynomial in $s(\widehat{\mathcal{T}})$. In addition, it is easy to see that both \mathfrak{b} and \mathfrak{d} can be computed in time polynomial in $s(\widehat{\mathcal{T}})$. Thus, Algorithm 1 is a non-deterministic polynomial space decision procedure for satisfiability of concepts of the form $\alpha_1 \sqcap \neg\alpha_2$ with respect to acyclic $\tau\mathcal{EL}(deg)$ TBoxes. This means that satisfiability and non-subsumption are in NPSpace. Then, by Savitch’s theorem [Sav70] and since PSpace is closed under complement, we obtain the following results.

Theorem 43. *In $\tau\mathcal{EL}(deg)$, satisfiability and subsumption are in PSpace, with respect to acyclic $\tau\mathcal{EL}(deg)$ TBoxes.*

4.4 Reasoning with acyclic knowledge bases

We show in this section that satisfiability and subsumption are still decidable in PSpace with respect to acyclic knowledge bases. Furthermore, we also consider the *consistency* and the *instance* problem. Let $\mathcal{K} = (\widehat{\mathcal{T}}, \mathcal{A})$ be an acyclic $\tau\mathcal{EL}(deg)$ knowledge base:

- \mathcal{K} is *consistent* iff there is an interpretation \mathcal{I} such that $\mathcal{I} \models \mathcal{K}$.

Additionally, let $a \in \mathbf{N}_1$ be an individual name and α a defined concept in $\widehat{\mathcal{T}}$:

- a is an instance of α with respect to \mathcal{K} iff for all models \mathcal{I} of \mathcal{K} it holds that $a^{\mathcal{I}} \in \alpha^{\mathcal{I}}$.

Without loss of generality, we can restrict our attention to the consistency problem for KBs of the form $(\widehat{\mathcal{T}}, \mathcal{A} \cup \{\neg\alpha(a)\})$, since all the other problems can be reduced to it.

Proposition 44. *Let $\mathcal{K} = (\widehat{\mathcal{T}}, \mathcal{A})$ be an acyclic $\tau\mathcal{EL}(deg)$ KB, α, α_1 and α_2 defined concepts in $\widehat{\mathcal{T}}$ and $a \in \mathbf{N}_1$. Then,*

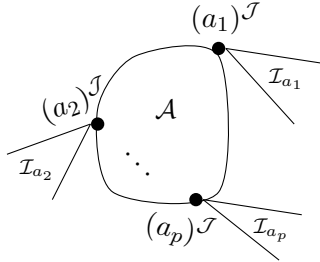
1. α is satisfiable with respect to \mathcal{K} iff $(\widehat{\mathcal{T}}, \mathcal{A} \cup \{\alpha(b)\})$ is consistent, where b is an individual name not occurring in \mathcal{A} .
2. α_1 is subsumed by α_2 with respect to \mathcal{K} (in symbols $\alpha_1 \sqsubseteq_{\mathcal{K}} \alpha_2$) iff the knowledge base $(\widehat{\mathcal{T}}, \mathcal{A} \cup \{\alpha_1(b), \neg\alpha_2(b)\})$ is inconsistent, where b is an individual name not occurring in \mathcal{A} .
3. a is an instance of α in \mathcal{K} (in symbols $\mathcal{K} \models \alpha(a)$) iff $(\widehat{\mathcal{T}}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is not consistent.

Further, since $\widehat{\mathcal{T}}$ is acyclic, by using unfolding we can again get rid of the TBox and reduce reasoning to consistency with respect to the empty terminology. The unfolding of a $\tau\mathcal{EL}(\text{deg})$ ABox \mathcal{A} with respect to $\widehat{\mathcal{T}}$ is defined as follows:

$$u_{\widehat{\mathcal{T}}}(\mathcal{A}) := \bigcup_{\substack{\widehat{C}(a) \in \mathcal{A} \\ a \in \text{Ind}(\mathcal{A})}} \{[u_{\widehat{\mathcal{T}}}(\widehat{C})](a)\} \cup \bigcup_{\substack{r(a,b) \in \mathcal{A} \\ a,b \in \text{Ind}(\mathcal{A})}} \{r(a,b)\}$$

Proposition 45. *Let $\mathcal{K} = (\widehat{\mathcal{T}}, \mathcal{A})$ be an acyclic $\tau\mathcal{EL}(\text{deg})$ KB, α a defined concept in $\widehat{\mathcal{T}}$ and $a \in \mathbf{N}_I$. $(\widehat{\mathcal{T}}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is consistent iff $u_{\widehat{\mathcal{T}}}(\mathcal{A}) \cup \{[\neg u_{\widehat{\mathcal{T}}}(\alpha)](a)\}$ is consistent.*

In what follows, we show how to extend the ideas used to design *Gen+* and Algorithm 1, to decide consistency of $u_{\widehat{\mathcal{T}}}(\mathcal{A}) \cup \{[\neg u_{\widehat{\mathcal{T}}}(\alpha)](a)\}$. As mentioned in Section 2.3, there is again a bound for the size of the interpretations that one needs to look at to decide consistency of $u_{\widehat{\mathcal{T}}}(\mathcal{A}) \cup \{[\neg u_{\widehat{\mathcal{T}}}(\alpha)](a)\}$. Moreover, if such an ABox is consistent, it has a model \mathcal{J} of the following form (see [BBF15]):



where $\text{Ind}(\mathcal{A}) = \{a_1, a_2, \dots, a_p\}$ and $\mathcal{I}_{a_1}, \mathcal{I}_{a_2}, \dots, \mathcal{I}_{a_p}$ are tree-shaped interpretations. The inner area of the diagram consists of the satisfaction of the role assertions in \mathcal{A} , i.e., $(a^{\mathcal{J}}, b^{\mathcal{J}}) \in r^{\mathcal{J}}$ iff $r(a,b) \in \mathcal{A}$. Additionally, an upper bound is provided for the size of these tree-shaped interpretations. We will later talk about how big this bound could be, but for the moment let us focus in how to reuse *Gen+* and Algorithm 1.

To start, it is clear that by choosing the appropriate values for \mathfrak{d} and \mathfrak{b} , the interpretations \mathcal{I}_a can be independently generated using the function *Gen+*. It

is important to keep in mind that $a^{\mathcal{I}_a}$ is the root of \mathcal{I}_a . Consequently, a run ρ_a of $Gen+$ inducing \mathcal{I}_a will compute two sets Ex_a and D_a with the following meaning:

$$D_a(E) = \text{deg}^{\mathcal{I}_a}(a^{\mathcal{I}_a}, u_{\mathcal{T}}(E)), \text{ for all } E \doteq C_E \in \mathcal{T}$$

$$\text{Ex}_a = \{\beta \mid \beta \doteq \widehat{C}_\beta \in \mathcal{T}_\tau \cup \mathcal{T} \text{ and } a^{\mathcal{I}_a} \in \beta^{\mathcal{I}_a}\}$$

Recall that technically \mathcal{I}_a (as generated by $Gen+$) only interprets symbols from $\mathbf{N}_{pr}^{\widehat{\mathcal{T}}} \cup \mathbf{N}_R$, but when writing $\beta^{\mathcal{I}_a}$ we meant its unique extension to a model of $\widehat{\mathcal{T}}$. The veracity of the previous two equalities has been shown in Lemma 41. Now, the construction of the model \mathcal{J} depicted above (Lemma 44 in [BBF15]) is done in two steps. First, the interpretations \mathcal{I}_a for all $a \in \text{Ind}(\mathcal{A})$ are obtained using Lemma 42 in [BBF15] and the combined (Lemma 43, [BBF15]) in the following way:

- $\Delta^{\mathcal{J}} = \bigcup_{a \in \text{Ind}(\mathcal{A})} \Delta^{\mathcal{I}_a}$,
- $A^{\mathcal{J}} = \bigcup_{a \in \text{Ind}(\mathcal{A})} A^{\mathcal{I}_a}$ for all $A \in \mathbf{N}_{pr}^{\widehat{\mathcal{T}}}$,
- $r^{\mathcal{J}} = \bigcup_{a \in \text{Ind}(\mathcal{A})} r^{\mathcal{I}_a} \cup \{(a^{\mathcal{I}_a}, b^{\mathcal{I}_b}) \mid r(a, b) \in \mathcal{A}\}$ for all $r \in \mathbf{N}_R$, and
- $a^{\mathcal{J}} = a^{\mathcal{I}_a}$, for all $a \in \text{Ind}(\mathcal{A})$.

This means that given an individual $a \in \text{Ind}(\mathcal{A})$, a defined concept β and an element $d \in \Delta^{\mathcal{I}_a}$, it is not necessarily the case that $d^{\mathcal{J}} \in \beta^{\mathcal{J}}$ iff $\beta \in \text{Ex}_a$ (similarly for the membership degrees and the assignment D_d). The reason is that the role assertions between individual names are used to build \mathcal{J} , but they are not taken into account by ρ_a to compute Ex_a and D_a . Fortunately, this could only be the case for the domain elements $a^{\mathcal{J}} = a^{\mathcal{I}_a}$ corresponding to the individual names of \mathcal{A} . This is a consequence of the following observation: for all $a \in \text{Ind}(\mathcal{A})$ and $d \in \Delta^{\mathcal{I}_a}$ such that $d \neq a^{\mathcal{I}_a}$, no path in $G_{\mathcal{J}}$ starting at d reaches a domain element $b^{\mathcal{J}}$ ($b \in \text{Ind}(\mathcal{A})$). As a result we obtain the following:

$$\text{deg}^{\mathcal{J}}(d, u_{\widehat{\mathcal{T}}}(E)) = \text{deg}^{\mathcal{I}_a}(d, u_{\widehat{\mathcal{T}}}(E)), \text{ for all } E \doteq C_E \in \mathcal{T}$$

$$d \in \beta^{\mathcal{J}} \text{ iff } d \in \beta^{\mathcal{I}_a}, \text{ for all } \beta \doteq \widehat{C}_\beta \in \mathcal{T}_\tau \cup \mathcal{T}$$

Therefore, if we can compute the correct content/values of Ex_a and D_a for the unique extension of \mathcal{J} satisfying $\widehat{\mathcal{T}}$, it will be possible to verify whether \mathcal{J} satisfies $u_{\widehat{\mathcal{T}}}(\mathcal{A}) \cup \{[-u_{\widehat{\mathcal{T}}}(\alpha)](a)\}$ (as it is done for subsumption in the previous section). There are two obstacles that we need to overcome. The first one is that Ex_a and D_a , as computed by ρ_a , do not contain enough information to obtain the ones corresponding to \mathcal{J} .

Example 46. Let $a_1, a_2 \in \text{Ind}(\mathcal{A})$ and $r(a_1, a_2) \in \mathcal{A}$. Suppose that a run ρ_{a_1} of $\text{Gen}+$ representing \mathcal{I}_{a_1} yields $D_{a_1}(E) = t_1$ for some $E \doteq C_E \in \mathcal{T}$. Likewise, $D_{a_2}(E') = t_2$ for some run ρ_{a_2} representing \mathcal{I}_{a_2} and $E' \doteq C_{E'} \in \mathcal{T}$. In addition, there is a top-level atom in C_E of the form $\exists r.E'$.

As explained above, the value of $D_{a_2}(E')$ has not been considered in the computation of $D_{a_1}(E)$, and it may well be the case that it actually affects $D_{a_1}(E)$ in the big model \mathcal{J} , i.e., $\text{deg}^{\mathcal{J}}((a_1)^{\mathcal{J}}, u_{\widehat{\mathcal{T}}}(E)) > t_1$. This could happen if for all r -successors d of a_1 in \mathcal{I}_{a_1} , we have that $\text{deg}^{\mathcal{I}_{a_1}}(d, u_{\widehat{\mathcal{T}}}(E')) < t_2$. Clearly, this is not something that can be inferred from D_{a_1} , but from the table v computed for a_1 by ρ_{a_1} .

Similarly, assume that $\beta \notin \text{Ex}_{a_1}$ for some $\beta \doteq \widehat{C}_\beta \in \mathcal{T}_\tau$. This means that $a^{\mathcal{I}_a} \notin \beta^{\mathcal{I}_a}$. It could happen that a_2 satisfies properties in \mathcal{J} that would make $(a_1)^{\mathcal{J}} \in \beta^{\mathcal{J}}$. Then, we would need to look into the relation z computed for a_1 by ρ_{a_1} , to discern such a change.

To deal with that, we rearrange the structure of function $\text{Gen}+$ such that it returns the pair (z, v) instead of (Ex, D) . The following sketches how to modify $\text{Gen}+$ accordingly.

```

1: function  $\text{Gen}+(\mathfrak{d} : \text{integer}, \mathfrak{b} : \text{binary})$ 
2:    $\vdots$ 
3:   initialize  $v$  and  $z$ 
4:    $\vdots$ 
5:    $(z_r^i, v_r^i) := \text{Gen}+(\mathfrak{d} - 1, \mathfrak{b}_r^i + 1)$ 
6:    $D_r^i := \text{SUBDEG}(v_r^i)$ 
7:    $\text{Ex}_r^i := \text{SUBEX}(D_r^i, z_r^i)$ 
8:   update  $v$ 
9:   update  $z$ 
10:   $\vdots$ 
11:  return  $(z, v)$ 
12: end function

```

Note that in the previous version of $\text{Gen}+$, the computation of D_r^i and Ex_r^i are the last operations executed inside the recursive call $\text{Gen}+(\mathfrak{d} - 1, \mathfrak{b}_r^i + 1)$, and v, z are updated right away after that. This order of actions is kept in the new definition given above. Since the computation of D_r^i and Ex_r^i only requires of v_r^i and z_r^i , and these are returned by $\text{Gen}+$, the new modifications preserve the properties of $\text{Gen}+$.

The next step is to recompute z_a and v_a for all $a \in \text{Ind}(\mathcal{A})$ using the information provided by the role assertions in \mathcal{A} . Following Example 46, since $b^{\mathcal{J}}$ is related to $a^{\mathcal{J}}$ by the role name r , this means that v_a and z_a must be updated with respect to r , Ex_b and D_b . Obviously, changes in v_a and z_a should be propagated to the

individuals that a is related to, and so on. The function $Gen+$ can cope with such propagation in a bottom-up form, because it is *computing* a tree-shaped structure. However, this is no longer the case for the individuals in \mathcal{A} , since role assertions can define cycles involving them.

To solve this we appeal to the acyclic nature of $\widehat{\mathcal{T}}_\tau$ and \mathcal{T} . It allows to traverse the structure of any defined concept (bottom-up) based on the partial order \preceq induced by \rightarrow^+ on $\mathbf{N}_d^{\widehat{\mathcal{T}}}$. Note that now we limit our attention to the fragment of \mathcal{J} corresponding to the role assertions in \mathcal{A} , which is part of the input. Therefore, provided that (z_a, v_a) has been computed for all $a \in \text{Ind}(\mathcal{A})$, the following subroutine updates all those pairs with respect to the combined interpretation \mathcal{J} .

```

1: procedure UPDATE( )
2:   compute  $D_a := \text{SUBDEG}(v_a)$  // for all  $a \in \text{Ind}(\mathcal{A})$ 
3:   let  $\{E_1, \dots, E_n\}$  be a post-order of  $\preceq$  (induced by  $\rightarrow^+$  on  $\mathbf{N}_d^{\widehat{\mathcal{T}}}$ )
4:   for all  $1 \leq i \leq n$  do
5:     for all  $r(a, b) \in \mathcal{A}$  do
6:       if  $D_b(E_i) > v_a[r, E_i]$  then
7:          $v_a[r, E_i] := D_b(E_i)$ 
8:       end if
9:     end for
10:    re-compute  $D_a$  // for all  $a \in \text{Ind}(\mathcal{A})$ 
11:  end for
12:  compute  $\text{Ex}_a := \text{SUBEX}(D_a, z_a)$  // for all  $a \in \text{Ind}(\mathcal{A})$ 
13:  let  $\{\beta_1, \dots, \beta_n\}$  be a post-order of  $\preceq$  on  $\mathbf{N}_d^{\widehat{\mathcal{T}}}$ 
14:  for all  $1 \leq i \leq n$  do
15:    for all  $r(a, b) \in \mathcal{A}$  do
16:      if  $\beta_i \in \text{Ex}_b$  then
17:         $z_a := z_a \cup \{(r, \beta_i)\}$ 
18:      end if
19:    end for
20:    re-compute  $\text{Ex}_a$  // for all  $a \in \text{Ind}(\mathcal{A})$ 
21:  end for
22: end procedure

```

Let us prove that UPDATE does what we have claimed.

Lemma 47. *For all $a \in \text{Ind}(\mathcal{A})$, let ρ_a be a run of A and \mathcal{I}_a its induced interpretation. Moreover, based on these interpretations let \mathcal{J} be the interpretation that results from the combination described above. Then,*

1. $D_a(E) = \text{deg}^{\mathcal{J}}(a^{\mathcal{J}}, u_{\mathcal{T}}(E))$, for all $E \doteq C_E \in \mathcal{T}$.
2. $\text{Ex}_a = \{\beta \mid \beta \doteq \widehat{C}_\beta \in \mathcal{T}_\tau \cup \mathcal{T} \text{ and } a^{\mathcal{J}} \in \beta^{\mathcal{J}}\}$

Proof. We give the proof for the assignments D_a . The case for Ex_a can be done using the same idea and Lemma 41. To differentiate the final assignment D_a from the initial one computed by ρ_a , we denote the latter as D_a^0 (likewise for v_a and v_a^0). We show the claim by well-founded induction on the partial order \preceq .

Let $a \in \text{Ind}(\mathcal{A})$ and $E \doteq C_E \in \mathcal{T}$. Since \mathcal{T} is normalized, the concept description C_E has the following structure:

$$P_1 \sqcap \dots \sqcap P_q \sqcap \exists r_1.E_1 \sqcap \dots \sqcap \exists r_n.E_n$$

Clearly, when $n = 0$ the value $\text{deg}^{\mathcal{J}}(a^{\mathcal{J}}, u_{\mathcal{T}}(E))$ does not depend on any successor of $a^{\mathcal{J}}$. Moreover, by construction of \mathcal{J} we know that $a^{\mathcal{J}} \in (P_i)^{\mathcal{J}}$ iff $a^{\mathcal{I}_a} \in (P_i)^{\mathcal{I}_a}$ for all $1 \leq i \leq q$. This implies that:

$$\text{deg}^{\mathcal{J}}(a^{\mathcal{J}}, u_{\mathcal{T}}(E)) = \text{deg}^{\mathcal{I}_a}(a^{\mathcal{I}_a}, u_{\mathcal{T}}(E))$$

Then, by Lemma 41 we obtain that:

$$D_a^0(E) = \text{deg}^{\mathcal{J}}(a^{\mathcal{J}}, u_{\mathcal{T}}(E))$$

Looking at SUBDEG one can see that the computation of $D_a^0(E)$ depends only on the values $v_a^0[\epsilon, P]$. Furthermore, it is easy to see that those values are never changed by a run of UPDATE. Hence, $v_a[\epsilon, P] = v_a^0[\epsilon, P]$ and $D_a(E) = D_a^0(E)$. Thus, $D_a(E)$ is the right number.

Now, to show the claim for $n > 0$ we start by making some observations for all $b \in \text{Ind}(\mathcal{A})$. Let F be a defined concept in \mathcal{T} :

- By Lemma 41, the initial table v_b^0 satisfies the following:

$$v_b^0[r, F] = \max\{\text{deg}^{\mathcal{I}_b}(d, u_{\mathcal{T}}(F)) \mid d \in \Delta^{\mathcal{I}_b} \text{ and } (b^{\mathcal{I}_b}, d) \in r^{\mathcal{I}_b}\}$$

As explained above, since $d \neq b^{\mathcal{I}_b}$ it further satisfies:

$$v_b^0[r, F] = \max\{\text{deg}^{\mathcal{J}}(d, u_{\mathcal{T}}(F)) \mid d \in \Delta^{\mathcal{I}_b} \text{ and } (b^{\mathcal{I}_b}, d) \in r^{\mathcal{I}_b}\} \quad (26)$$

Additionally, let j be the index of F in the post-order created in line 3. Then,

- the value of $v_b[r, F]$ only changes at the j^{th} iteration of the outer-loop in line 4.
- let k be the largest index of F' among all the top-level atoms of the form $\exists r.F'$ in the definition of F . Then, taking into account the previous statement, the value of $D_b(F)$ never changes after the k^{th} iteration of the outer-loop.
- since $F' \preceq F$, this means that $j > k$. Consequently, the final value of $D_b(F)$ is computed before the iteration corresponding to F .

Coming back to the defined concept E , we know that $E \preceq E_j$ for all $1 \leq j \leq n$. Then, the application of induction hypothesis yields:

$$D_a(E_j) = \text{deg}^{\mathcal{J}}(a^{\mathcal{J}}, u_{\mathcal{T}}(E_j)) \quad (27)$$

Moreover, since at the moment of updating $v_a[r, E_j]$ the value of $D_b(E_j)$ is the one in (27) for all $b \in \text{Ind}(\mathcal{A})$, using (26) we obtain:

$$v_a[r, E_j] = \max\{\text{deg}^{\mathcal{J}}(d, u_{\mathcal{T}}(E_j)) \mid d \in \Delta^{\mathcal{J}} \text{ and } (a^{\mathcal{J}}, d) \in r^{\mathcal{J}}\}$$

Thus, by the same arguments given in Lemma 41 it follows:

$$D_a(E) = \text{deg}^{\mathcal{J}}(a^{\mathcal{J}}, u_{\mathcal{T}}(E))$$

□

By the previous lemma, once (Ex_a, D_a) has been computed by UPDATE for all $a \in \text{Ind}(\mathcal{A})$, it is easy to verify whether \mathcal{J} satisfies $\mathcal{A} \cup \{\neg\alpha(a)\}$. Therefore, it remains to make sure that enough candidates \mathcal{J} are considered to decide the satisfiability status of $u_{\hat{\mathcal{T}}}(\mathcal{A}) \cup \{[\neg u_{\hat{\mathcal{T}}}(\alpha)](a)\}$. This relies on estimating the appropriate values for \mathfrak{d} and \mathfrak{b} . Given an ABox \mathcal{A} in $\tau\mathcal{EL}(\text{deg})$, the ABox $\mathcal{A}(a)$ consists of all the concept assertions $\widehat{D}(a)$ occurring in \mathcal{A} . Then,

- Let $m_{\text{rd}}(\mathcal{A})$ be the maximal role depth of a concept \widehat{D} occurring in an ABox \mathcal{A} , i.e.,

$$m_{\text{rd}}(\mathcal{A}) := \max\{\text{rd}(\widehat{D}) \mid \widehat{D}(a) \in \mathcal{A}\}$$

The interpretations \mathcal{I}_a used to compose \mathcal{J} are built in such a way that its depth $\text{d}(\mathcal{I}_a)$ can be bounded by:

$$\text{d}(\mathcal{I}_a) \leq m_{\text{rd}}(\mathcal{A}(a)) + \text{rd}(\widehat{C})$$

In the present context this means that \mathfrak{d}_a can be chosen as:

$$m_{\text{rd}}(u_{\hat{\mathcal{T}}}(\mathcal{A}(a))) + \text{rd}(u_{\hat{\mathcal{T}}}(\alpha))$$

- Moreover, we have an upper-bound for $|\Delta^{\mathcal{I}_a}|$, namely,

$$|\Delta^{\mathcal{I}_a}| \leq \mathfrak{s}(\mathcal{A}(a)) \cdot [\mathfrak{s}(\widehat{C})]^u$$

where $u = |\text{sub}(\widehat{C})|$. Translating this bound to our current setting, we obtain:

$$|\Delta^{\mathcal{I}_a}| \leq \mathfrak{s}(u_{\hat{\mathcal{T}}}(\mathcal{A}(a))) \cdot [\mathfrak{s}(u_{\hat{\mathcal{T}}}(\alpha))]^{u^*}$$

with u^* now being $|\text{sub}(u_{\hat{\mathcal{T}}}(\alpha))|$.

Putting all the given arguments together, we devise Algorithm 2 below as a non-deterministic procedure to decide consistency of $(\widehat{\mathcal{T}}, \mathcal{A} \cup \{\neg\alpha(a)\})$. The following lemma shows that it is correct.

Lemma 48. *Let $\mathcal{K} = (\widehat{\mathcal{T}}, \mathcal{A})$ be an acyclic $\tau\mathcal{EL}(\text{deg})$ KB, α a defined concept in $\widehat{\mathcal{T}}$ and $a \in \text{Ind}(\mathcal{A})$. Then,*

Algorithm 2 answers “yes” iff $(\widehat{\mathcal{T}}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is consistent.

Proof. (\Rightarrow) Suppose that the algorithm gives a positive answer, and for all $a \in \text{Ind}(\mathcal{A})$ let ρ_a be the run of *Gen+* that leads to it. Then, we can talk about the interpretation \mathcal{I}_a induced by ρ_a . Now, let \mathcal{J} be the interpretation that results from the combination of all the fragments \mathcal{I}_a and the role assertions occurring in \mathcal{A} . A “yes” answer implies that the **for** loop described between lines 7 and 13 never falsifies $\beta \in \text{Ex}_b$ for all concept assertions $\beta(b) \in \mathcal{A}$. By Lemma 47, this means that the extension of \mathcal{J} satisfying $\widehat{\mathcal{T}}$ is also a model of \mathcal{A} .

In addition, the conditional in line 14 must evaluate to *true*. Consequently, for the same reasons explained above, we obtain that $a^{\mathcal{J}} \notin \alpha^{\mathcal{J}}$. Thus, $(\widehat{\mathcal{T}}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is consistent.

(\Leftarrow) Conversely, assume that $(\widehat{\mathcal{T}}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is consistent. This means that there is an interpretation $\mathcal{J} \models \mathcal{K}$ such that $a^{\mathcal{J}} \notin \alpha^{\mathcal{J}}$. By Proposition 45 and the mentioned results from [BBF15], one can assume that \mathcal{J} is of the form described before. Therefore, for all $a \in \text{Ind}(\mathcal{A})$ the corresponding interpretation \mathcal{I}_a is tree-shaped and satisfies:

- $d(\mathcal{I}_a) \leq m_{\text{rd}}(u_{\widehat{\mathcal{T}}}(\mathcal{A})) + \text{rd}(u_{\widehat{\mathcal{T}}}(\alpha))$, and
- $|\Delta^{\mathcal{I}_a}| \leq s(u_{\widehat{\mathcal{T}}}(\mathcal{A})) \cdot [s(u_{\widehat{\mathcal{T}}}(\alpha))]^{u^*}$ (note that $s(u_{\widehat{\mathcal{T}}}(\mathcal{A}(a))) \leq s(u_{\widehat{\mathcal{T}}}(\mathcal{A}))$).

By the selection of \mathfrak{d} and \mathfrak{b} in Algorithm 2 and an application of Lemma 40, there is always a run ρ_a of *Gen+* generating \mathcal{I}_a for all $a \in \text{Ind}(\mathcal{A})$. Then, by Lemma 47, after executing UPDATE none of the subsequent conditionals could evaluate to *false*. Thus, the algorithm answers “yes”. \square

Regarding the computational complexity of Algorithm 2, one can see that the value of \mathfrak{d} is a polynomial in the size of \mathcal{K} . Furthermore, since there are polynomially many individual names, this means that any run of the algorithm uses polynomial space (including the execution of UPDATE), except maybe for the number of bits needed to represent \mathfrak{b} . Indeed, the expression that calculates \mathfrak{b} is exponential in u^* . To give a preliminary approximation of how big \mathfrak{b} could be, we observe that due to unfolding we may end up with the following worst-case lower bounds:

$$2^{s(\widehat{\mathcal{T}})} \leq s(u_{\widehat{\mathcal{T}}}(\mathcal{A})) \text{ and } 2^{s(\widehat{\mathcal{T}})} \leq s(u_{\widehat{\mathcal{T}}}(\alpha))$$

Algorithm 2 Consistency of $(\widehat{\mathcal{T}}, \mathcal{A} \cup \{\neg\alpha(a)\})$.

Input: An acyclic KB $(\widehat{\mathcal{T}}, \mathcal{A})$, a defined concept α in $\widehat{\mathcal{T}}$ and $a \in \mathbb{N}_1$.

Output: “yes”, if $(\widehat{\mathcal{T}}, \mathcal{A} \cup \{\neg\alpha(a)\})$ is consistent, “no” otherwise.

```

1:  $\mathfrak{b} := \mathfrak{s}(u_{\widehat{\mathcal{T}}}(\mathcal{A})) \cdot [\mathfrak{s}(u_{\widehat{\mathcal{T}}}(\alpha))]^{u^*}$  //  $\mathfrak{b}$  is represented in binary
2:  $\mathfrak{d} := m_{\text{rd}}(u_{\widehat{\mathcal{T}}}(\mathcal{A})) + \text{rd}(u_{\widehat{\mathcal{T}}}(\alpha))$ 
3: for all  $b \in \text{Ind}(\mathcal{A})$  do
4:    $(z_b, v_b) := \text{Gen}+(\mathfrak{d}, \mathfrak{b})$ 
5: end for
6: UPDATE( )
7: for all  $b \in \text{Ind}(\mathcal{A})$  do
8:   for all  $\beta(b) \in \mathcal{A}$  do
9:     if  $\beta \notin \text{Ex}_b$  then
10:      return “no”
11:     end if
12:   end for
13: end for
14: if  $\alpha \notin \text{Ex}_a$  then
15:   return “yes”
16: end if
17: return “no”

```

In particular, u^* corresponds to the number of sub-descriptions of $u_{\widehat{\mathcal{T}}}(\alpha)$. Hence, in view of the lower bound for the size of $u_{\widehat{\mathcal{T}}}(\alpha)$ one might think that the following lower bound also holds:

$$2^{2^{\mathfrak{s}(\widehat{\mathcal{T}})}} \leq [\mathfrak{s}(u_{\widehat{\mathcal{T}}}(\alpha))]^{u^*} \quad (28)$$

Therefore, in the worst-case we would end up with an ExpSpace non-deterministic procedure. However, on the one side, a closer look at the reductions in Proposition 44 reveals that there are better choices for \mathfrak{b} depending on the reasoning problem. On the other side, the statement in (28) is actually false.

- *Knowledge base consistency and satisfiability:* in these cases the problem reduces to consistency of a $\tau\mathcal{EL}(\text{deg})$ ABox. Consequently, such double exponential explosion does not exist. Thus, \mathfrak{b} simply becomes $\mathfrak{s}(u_{\widehat{\mathcal{T}}}(\mathcal{A}))$ or $\mathfrak{s}(u_{\widehat{\mathcal{T}}}(\mathcal{A} \cup \{\alpha(b)\}))$.
- *Subsumption:* the reduction produces an ABox of the form:

$$\mathcal{A} \cup \{\alpha_1(b), \neg\alpha_2(b)\}$$

The key aspect is that b does not occur in \mathcal{A} . This means that the pre-processing propagation of the negative assertions does not go through the

cycles that may occur in \mathcal{A} . This obviously avoids the exponential explosion and \mathbf{b} can be selected as:

$$\mathbf{s}(u_{\hat{\mathcal{T}}}(\mathcal{A})) + [\mathbf{s}(u_{\hat{\mathcal{T}}}(\alpha_1)) \cdot \mathbf{s}(u_{\hat{\mathcal{T}}}(\alpha_2))]$$

- *Instance checking:* According to (28), in this case the algorithm would need to store a value $\mathbf{b} \geq 2^{2^{\hat{\mathcal{T}}}}$. However, one can show that the number of sub-descriptions in $u_{\hat{\mathcal{T}}}(\alpha)$ is actually bounded by $\mathbf{s}(\hat{\mathcal{T}})$ (see Corollary 51 in the Appendix). Hence, the statement made in (28) is false and \mathbf{b} can be chosen as:

$$\mathbf{s}(u_{\hat{\mathcal{T}}}(\mathcal{A})) \cdot [\mathbf{s}(u_{\hat{\mathcal{T}}}(\alpha))]^{\mathbf{s}(\hat{\mathcal{T}})}$$

Consequently, the binary representation of \mathbf{b} needs only polynomially many bits in the size of $\hat{\mathcal{T}}$.

Thus, reasoning in $\tau\mathcal{EL}(deg)$ with respect to acyclic KBs is in PSpace.

Theorem 49. *In $\tau\mathcal{EL}(deg)$, consistency and instance checking w.r.t. acyclic $\tau\mathcal{EL}(deg)$ knowledge bases are in PSpace.*

5 Conclusion

We have introduced a notion of acyclic TBoxes for $\tau\mathcal{EL}(m)$ such that unfolding still works both from the syntactic and the semantic point of view. For the special case of $\tau\mathcal{EL}(deg)$, we have investigated the complexity of reasoning w.r.t. such acyclic TBoxes. In contrast to the case of \mathcal{EL} , in $\tau\mathcal{EL}(deg)$ the presence of acyclic TBoxes increases the complexity.

Regarding future research, we will try to close the gap between Π_2^P/Σ_2^P and PSpace. Unfortunately, it is not clear to us how the construction employed in the hardness proof could be extended to higher levels of the polynomial hierarchy, let alone to PSpace. Conversely, it is also not clear how to generate and test an exponentially large model on some fixed level of the polynomial hierarchy. Another interesting and non-trivial problem is to extend our approach to more general forms of TBoxes (e.g., GCIs). As demonstrated by the semantic problems for unrestricted sets of concept definitions shown in this paper, naive extensions will probably lead to unintuitive results. For example, we have seen that, embedded in a threshold concept, a concept name and its definition need not lead to the same result. We have overcome this problem by modifying the graded membership function using unfolding. For TBoxes that are not acyclic, or do not even consist of concept definitions, this simple solution is not possible. Other interesting open problems are, for instance, to provide an intuitive semantics for nested threshold operators, and to apply our approach of approximately defining concepts to other DLs.

6 Appendix

Lemma 50. *Let \mathcal{T} be an acyclic \mathcal{EL} TBox in normal form. Then, for all $\alpha \in \mathbf{N}_d^{\mathcal{T}}$ the number of sub-descriptions of $u_{\mathcal{T}}(\alpha)$ is at most $\mathfrak{s}(\mathcal{T})$.*

Proof. Recall the definition of $\mathbf{sub}(C)$ in Definition 3. Let $\mathbf{sub}^*(C) \subseteq \mathbf{sub}(C)$ be the following set:

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

Furthermore, for all $\alpha \doteq C_{\alpha} \in \mathcal{T}$, let $\rightarrow^+(\alpha)$ denotes the set of defined concepts in \mathcal{T} that α depends on, i.e.:

$$\rightarrow^+(\alpha) := \{\beta \mid \beta \in \mathbf{N}_d^{\mathcal{T}} \text{ and } \alpha \rightarrow^+ \beta\}$$

We prove the following claim about the set $\mathbf{sub}(u_{\mathcal{T}}(\alpha))$:

$$\mathbf{sub}(u_{\mathcal{T}}(\alpha)) = \mathbf{sub}^*(u_{\mathcal{T}}(C_{\alpha})) \cup \bigcup_{\substack{\beta \doteq C_{\beta} \in \mathcal{T} \\ \beta \in \rightarrow^+(\alpha)}} \mathbf{sub}^*(u_{\mathcal{T}}(C_{\beta})) \quad (29)$$

The proof is by well-founded induction on the partial order \preceq induced by \rightarrow^+ on $\mathbf{N}_d^{\mathcal{T}}$. Let $\alpha \doteq C_{\alpha} \in \mathcal{T}$, due to the normal form of \mathcal{T} the concept C_{α} has the following structure:

$$P_1 \sqcap \dots \sqcap P_q \sqcap \exists r_1.\beta_1 \sqcap \dots \sqcap \exists r_n.\beta_n$$

The unfolding of α with respect to \mathcal{T} is the following concept description:

$$u_{\mathcal{T}}(\alpha) = P_1 \sqcap \dots \sqcap P_q \sqcap \exists r_1.u_{\mathcal{T}}(\beta_1) \sqcap \dots \sqcap \exists r_n.u_{\mathcal{T}}(\beta_n)$$

By the definitions of \mathbf{sub} and \mathbf{sub}^* , we can express the set $\mathbf{sub}(u_{\mathcal{T}}(\alpha))$ as follows:

$$\mathbf{sub}(u_{\mathcal{T}}(\alpha)) = \mathbf{sub}^*(u_{\mathcal{T}}(C_{\alpha})) \cup \bigcup_{i=1}^n \mathbf{sub}(u_{\mathcal{T}}(\beta_i)) \quad (30)$$

Now, the application of the induction hypothesis to each β_i ($1 \leq i \leq n$) yields:

$$\mathbf{sub}(u_{\mathcal{T}}(\beta_i)) = \mathbf{sub}^*(u_{\mathcal{T}}(C_{\beta_i})) \cup \bigcup_{\substack{\beta \doteq C_{\beta} \in \mathcal{T} \\ \beta \in \rightarrow^+(\beta_i)}} \mathbf{sub}^*(u_{\mathcal{T}}(C_{\beta}))$$

Hence, substituting the previous equality in (30) we obtain the following one:

$$\text{sub}(u_{\mathcal{T}}(\alpha)) = \text{sub}^*(u_{\mathcal{T}}(C_{\alpha})) \cup \bigcup_{i=1}^n \left[\text{sub}^*(u_{\mathcal{T}}(C_{\beta_i})) \cup \bigcup_{\substack{\beta \doteq C_{\beta} \in \mathcal{T} \\ \beta \in \rightarrow^+(\beta_i)}} \text{sub}^*(u_{\mathcal{T}}(C_{\beta})) \right]$$

Finally, since $\rightarrow^+(\alpha) = \bigcup_{i=1}^n (\{\beta_i\} \cup \rightarrow^+(\beta_i))$, it is clear that the set defined by the big union in the previous equality is equal to the one represented by the big union in (29). Thus, our claim in (29) is true.

According to the definition of sub^* , for a top-level atom $\exists r_i.\beta_i$ of C_{α} the set of concepts $\text{sub}^*(\exists r_i.u_{\mathcal{T}}(\beta_i))$ corresponds to $\{\exists r_i.u_{\mathcal{T}}(\beta_i)\}$. Hence, it is not hard to see that for all $\alpha \doteq C_{\alpha} \in \mathcal{T}$ it holds:

$$|\text{sub}^*(u_{\mathcal{T}}(C_{\alpha}))| \leq \mathfrak{s}(C_{\alpha})$$

Thus, using (29) we can conclude that $|\text{sub}(u_{\mathcal{T}}(\alpha))| \leq \mathfrak{s}(\mathcal{T})$ for all $\alpha \in \mathbf{N}_d^{\mathcal{T}}$. \square

Now, since $\text{sub}(E_{\sim t})$ is equal to $\{E_{\sim t}\}$, the previous result also applies to acyclic $\tau\mathcal{EL}(\text{deg})$ TBoxes.

Corollary 51. *Let $\widehat{\mathcal{T}}$ be an acyclic $\tau\mathcal{EL}(\text{deg})$ TBox in normal form. Then, for all $\alpha \in \mathbf{N}_d^{\widehat{\mathcal{T}}}$ it holds:*

$$|\text{sub}(u_{\widehat{\mathcal{T}}}(\alpha))| \leq \mathfrak{s}(\widehat{\mathcal{T}})$$

References

- [Baa02] Franz 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>.
- [Baa03] Franz Baader. Terminological cycles in a description logic with existential restrictions. In Georg Gottlob and Toby Walsh, editors, *IJCAI*, pages 325–330. Morgan Kaufmann, 2003.
- [BBF15] Franz Baader, Gerhard Brewka, and Oliver Fernández Gil. Adding Threshold Concepts to the Description Logic \mathcal{EL} . LTCS-Report LTCS-15-09, LTCS-15-09, Chair for Automata Theory, Institute for Theoretical Computer Science, TU Dresden, Germany, Germany, 2015. See <http://lat.inf.tu-dresden.de/research/reports.html>.
- [BBG15] Franz Baader, Gerhard Brewka, and Oliver Fernandez Gil. Adding threshold concepts to the description logic \mathcal{EL} . In Carsten Lutz and Silvio Ranise, editors, *Frontiers of Combining Systems - 10th International Symposium, FroCoS 2015, Wroclaw, Poland, September 21-24, 2015. Proceedings*, volume 9322 of *Lecture Notes in Computer Science*, pages 33–48. Springer, 2015.
- [BCM⁺03] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, New York, NY, USA, 2003.
- [BKM99] Franz Baader, Ralf Küsters, and Ralf Molitor. Computing least common subsumers in description logics with existential restrictions. In Thomas Dean, editor, *IJCAI*, pages 96–103. Morgan Kaufmann, 1999.
- [Bra04] Sebastian Brandt. Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and - what else? In Ramon López de Mántaras and Lorenza Saitta, editors, *Proceedings of the 16th European Conference on Artificial Intelligence, ECAI'2004, including Prestigious Applicants of Intelligent Systems, PAIS 2004, Valencia, Spain, August 22-27, 2004*, pages 298–302. IOS Press, 2004.
- [DLNS94] Francesco M. Donini, Maurizio Lenzerini, Daniele Nardi, and Andrea Schaerf. Deduction in concept languages: From subsumption to instance checking. *J. Log. Comput.*, 4(4):423–452, 1994.
- [Küs01] Ralf Küsters. *Non-Standard Inferences in Description Logics*, volume 2100 of *Lecture Notes in Computer Science*. Springer, 2001.

- [Neb90] Bernhard Nebel. Terminological reasoning is inherently intractable. *Artif. Intell.*, 43(2):235–249, 1990.
- [Neb91] Bernhard Nebel. Terminological cycles: Semantics and computational properties, 1991.
- [Sav70] Walter J. Savitch. Relationships between nondeterministic and deterministic tape complexities. *J. Comput. Syst. Sci.*, 4(2):177–192, 1970.
- [Sto76] Larry J. Stockmeyer. The polynomial-time hierarchy. *Theor. Comput. Sci.*, 3(1):1–22, 1976.