Dresden University of Technology
Institute for Theoretical Computer Science
Chair for Automata Theory

LTCS–Report

Adding Causal Relationships to DL-based Action Formalisms

Franz Baader Marcel Lippmann Hongkai Liu

LTCS-Report 10 – 01

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

Adding Causal Relationships to DL-based Action Formalisms

Franz Baader, Marcel Lippmann, and Hongkai Liu

Institute of Theoretical Computer Science
Technische Universität Dresden
Germany
lastname@tcs.inf.tu-dresden.de

Abstract

In the reasoning about actions community, causal relationships have been proposed as a possible approach for solving the ramification problem, i. e., the problem of how to deal with indirect effects of actions. In this paper, we show that causal relationships can be added to action formalisms based on Description Logics without destroying the decidability of the consistency and the projection problem.

1 Introduction

For action theories represented in the situation or fluent calculus [13, 17], the important inference problems are in general undecidable since these calculi encompass full first-order logic (FOL). One possibility for avoiding this source of undecidability is to restrict the underlying logic from FOL to a decidable Description Logic (DL). The main argument for using DLs in this setting is that they offer considerable expressive power, going far beyond propositional logic, while reasoning is still decidable. An action formalism based on DLs was first introduced in [3], and it was shown that important reasoning problems such as the projection problem become decidable in this restricted formalism.

An action theory basically consists of three components: (i) a description of the *initial state*; (ii) a description of the possible *actions*, which specifies the *pre-conditions* that need to be satisfied for an action to be applicable as well as the *post-conditions*, i. e., the changes to the current state that its application causes; and (iii) *domain constraints*, which formulate general knowledge about the functioning of the domain in which the actions are executed, and thus restrict the possible states. In a DL-based action formalism, the initial state is (incompletely) described by an ABox, pre-conditions are ABox assertions that must hold, post-conditions are ABox assertions that are added or removed, and domain constraints are specified using TBox axioms.

The *ramification problem* is caused by the interaction of the post-conditions of an action with the domain constraints. To be more precise, when applying an action, it may not be enough to make only those changes to the current state that are explicitly required by its post-conditions (direct effects) since it might happen that the resulting state does not satisfy the domain constraints, in which case one needs to make additional changes in order to satisfy these constraints (indirect effects). For example, assume that we have a hiring action, which has the direct effect that the person that is hired is then an employee, and that we have a domain constraint that says that any employee must have a health insurance. If John does not have health insurance, then just applying the hiring action for John would result in a state that violates the health insurance domain constraint.

One approach for solving the ramification problem is trying to find a semantics for action theories that automatically deals with such indirect effects, i. e., somehow makes additional changes to the state in order to satisfy the domain constraints, while taking care that only "necessary" changes are made. An example of such an attempt is the possible models approach (PMA) [21, 7]. However, without additional restrictions, the PMA and all the other approaches in this direction can lead to unintuitive results. It is not clear how to construct a general semantics that does not suffer from this problem. In our example, assume that there are only two insurance companies that offer health insurance: AOK and TK. In order to satisfy the health insurance domain constraint, John must get insured by one of them, but how should a general semantic framework be able to decide which one to pick.

A *second approach* is to avoid rather than solve the issues raised by the ramification problem. This is actually what is done in [3]: the domain constraints are given by an acyclic TBox and post-conditions of actions are restricted such that only primitive concepts and roles are changed. Since, w. r. t. an acyclic TBox, the interpretations of the primitive concepts and roles uniquely determine the interpretations of the defined concepts, it is then clear what indirect effects such a change has. The semantics obtained this way can be seen as an instance of the PMA. It is shown in [3] that the use of the PMA in a less restrictive setting (use of more general TBoxes as domain constraints or of non-primitive concepts in post-conditions) leads to unintuitive results.

A *third approach* is to let the user rather than a general semantic machinery decide which are the implicit effects of an action. In our example, assume that employers actually are required to enroll new employees with AOK in case they do not already have a health insurance. One can now try to extend the action formalism such that it allows the user to add such information to the action theory. For DL-based action formalisms, this approach was first used in [11], where the formalism for describing the actions is extended such that the user can make complex statements about the changes to the interpretations of concepts and roles that can be caused by a given action. It is shown in [11] that important inference problems such as the projection problem stay decidable in this setting, but that the consistency[1] problem for actions becomes undecidable. In the present paper, we realize this third approach in a different way, by adapting a method for addressing the ramification problem that has already been employed in the reasoning about actions community [10, 16, 20, 5]. Instead of changing the formalism for defining actions, we introduce so-called *causal relationships* as an additional component of action theories. In our example, such a causal relationship would state that, whenever someone becomes a new employee, this person is then insured by AOK, unless (s)he already had a health insurance.

In this paper, we formally introduce DL-based action theories with causal relationships and show that important inference problems such as the projection problem are decidable in such theories. Our new formalism has two advantages over the one introduced in [11]. First, the formalism in [11] requires the user to deal with the ramification problem within every action description. In our formalism, causal relationships are defined independently of a specific action, stating general facts about causation. The semantics then takes care of how these relationships are translated into indirect effects of actions. A second, and more tangible, advantage is that, in our formalism, *consistency of actions* is decidable. Basically, an action is consistent if, whenever it is applicable in a state, there is a well-defined successor state that can be obtained by applying it. We believe that, in the context of the third approach, where the user is supposed to deal with the ramification problem (in our formalism by defining appropriate causal relationships), testing consistency helps the user to check whether (s)he got it right. For instance, consider our health insurance example. If the user does not specify any causal relationships, then the hiring action is inconsistent since its application may result in a state that does not satisfy the domain constraints, and thus is not well-defined. If (s)he adds the causal relationship mentioned above, then the action becomes consistent.

2 Description Logics

In principle, our action formalism can be parameterized with any DL. In this paper, we focus on DLs betweeen $\mathcal{ALC}$ and $\mathcal{ALCQIO}$.

The syntax of the DL $\mathcal{ALCQIO}$ is defined using three non-empty sets: a set $\mathsf{N_C}$ of *concept names*, a set $\mathsf{N_R}$ of *role names* and a set $\mathsf{N_I}$ of *individual names*. $\mathcal{ALCQIO}$-concept descriptions (or *concepts* for short) are inductively defined

[1] In [11], this is actually called strong consistency.

Name	Syntax	Semantics
negation	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
disjunction	$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
nominal	$\{a\}$	$\{a^{\mathcal{I}}\}$
inverse role	r^-	$(r^{\mathcal{I}})^{-1}$
at-least number restriction	$(\geqslant n\ r\ C)$	$\{x \mid \#\{y \mid (x,y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\} \geq n\}$
at-most number restriction	$(\leqslant n\ r\ C)$	$\{x \mid \#\{y \mid (x,y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\} \leq n\}$

Table 1: Syntax and semantics of $\mathcal{ALCQIO}$

with the help of the *constructors*, which are shown in Table 1. In Table 1 and in what follows, we use A and B to denote concept names, r and s to denote roles (i. e. role names and inverse roles), a and b to denote individual names, C and D to denote (possibly complex) concepts, n to denote some natural number, and $\#S$ to denote the cardinality of the set S. As usual, we use $\top$ as abbreviation for some arbitrary (but fixed) tautology (e. g. $A \sqcup \neg A$), $\bot$ for $\neg\top$, $\rightarrow$ and $\leftrightarrow$ for the usual Boolean abbreviations, $\exists r.C$ (*existential restriction*) for $(\geqslant 1\ r\ C)$, and $\forall r.C$ (*universal restriction*) for $(\leqslant 0\ r\ \neg C)$.

The DL allowing only for negation, conjunction, disjunction, existential restrictions and universal restrictions is called $\mathcal{ALC}$. If additional concept constructors are available, this is denoted by concatenating a corresponding letter: $\mathcal{Q}$ means (qualified) number restrictions, $\mathcal{I}$ means inverse roles, and $\mathcal{O}$ means nominals. For instance, the DL which is an extension of $\mathcal{ALC}$ and allows for inverse roles is called $\mathcal{ALCI}$.

The semantics of $\mathcal{ALCQIO}$-concepts is defined in terms of an *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$. The set $\Delta^{\mathcal{I}}$, the *domain*, is a non-empty set of individuals. The *interpretation function* $\cdot^{\mathcal{I}}$ maps each concept name A to a subset $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$, each role name r to a subset $r^{\mathcal{I}}$ of $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$, and each individual name a to an individual $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$. The extension of $\cdot^{\mathcal{I}}$ to complex concepts is defined inductively as indicated in the third column of Table 1.

An *ABox* is a finite set of *concept assertions* $C(a)$ and *role assertions* $r(a,b)$ and $\neg r(a,b)$, where C is a concept description, r is a role name, and a, b are individual names. An ABox is *simple* if all its assertions are of the form $A(a)$, $\neg A(a)$, $r(a,b)$, or $\neg r(a,b)$, where A is a concept name, r is a role name, and a, b are individual names. We will call the concept and role assertions that may occur in simple ABoxes *literals*. Literals of the form $A(a)$ and $r(a,b)$ ($\neg A(a)$ and $\neg r(a,b)$) are called *positive (negative)*. Given a literal L, its negation $\dot\neg L$ is $\neg L$ if L is a positive literal, and it is L' if $L = \neg L'$ is negative.

A *concept definition* is of the form $A \equiv C$ and a *general concept inclusion (GCI)* is of the form $C \sqsubseteq D$. An acyclic *terminological box (TBox)* is a finite set of concept definitions with unique left-hand sides. Additionally, we require that there are no cyclic dependencies between the definitions [2]. Concept names occuring on the left-hand side of some concept definition in the TBox are *defined concepts* whereas the others are called *primitive concepts*. A general TBox is a finite set of GCIs.

Example 1. *Coming back to the health insurance example from the introduction, the following GCIs express that all employees must be insured by a health insurance company, and that AOK and TK are health insurance companies:*

$$\begin{aligned} \mathsf{Employee} &\sqsubseteq \exists \mathsf{insuredBy}.\mathsf{HealthInsuranceCompany} \\ \{\mathsf{AOK}\} \sqcup \{\mathsf{TK}\} &\sqsubseteq \mathsf{HealthInsuranceCompany} \end{aligned}$$

The assertion $\neg\mathsf{Employee}(\mathsf{JOHN})$ *says that John is not an employee.*

An interpretation $\mathcal{I}$ *satisfies* an ABox assertion φ (written $\mathcal{I} \models \varphi$) if we have for φ being a concept assertion $C(a)$ that $a^{\mathcal{I}} \in C^{\mathcal{I}}$, for φ being a role assertion $r(a,b)$ that $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$, and for $\neg r(a,b)$ that $(a^{\mathcal{I}}, b^{\mathcal{I}}) \notin r^{\mathcal{I}}$, respectively. $\mathcal{I}$ *is a model* of $\mathcal{A}$ (written $\mathcal{I} \models \mathcal{A}$) iff we have for all $\varphi \in \mathcal{A}$ that $\mathcal{I} \models \varphi$.

The semantics of TBoxes is defined in the obvious way. $\mathcal{I}$ *is a model* of some TBox $\mathcal{T}$ (written $\mathcal{I} \models \mathcal{T}$) iff it *satisfies* each definition (and GCI, respectively) in $\mathcal{T}$. $\mathcal{I}$ satisfies some concept definition $A \equiv C$ (written $\mathcal{I} \models A \equiv C$) iff $A^{\mathcal{I}} = C^{\mathcal{I}}$. Analogously, $\mathcal{I}$ satisfies some GCI $C \sqsubseteq D$ (written $\mathcal{I} \models C \sqsubseteq D$) iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$.

The ABox $\mathcal{A}$ is *consistent* w. r. t. $\mathcal{T}$ if there exists an interpretation that is a model of $\mathcal{A}$ and $\mathcal{T}$. We say that the assertion φ (the TBox $\mathcal{T}'$) is a *logical consequence* of the ABox $\mathcal{A}$ and the TBox $\mathcal{T}$, denoted with $\mathcal{A} \cup \mathcal{T} \models \varphi$ ($\mathcal{A} \cup \mathcal{T} \models \mathcal{T}'$) iff every interpretation that is a model of $\mathcal{A}$ and $\mathcal{T}$ is also a model of φ ($\mathcal{T}'$).

3 DL-based Action Formalisms and Causal Relationships

In our DL-based action formalisms, actions are described by ABox assertions. A TBox is used for describing the domain constraints, an ABox gives us an incomplete knowledge about the application domain, and an interpretation gives a complete description of the application domain.

The following definition recalls the notion of a DL action without occlusions, which has first been introduced in [3]. At the moment, we do not allow for occlusions in our framework since it is not yet clear how to handle them algorithmically in the presence of causal relationships.

Definition 2. An *action* is a pair $\alpha = (\mathsf{pre}, \mathsf{post})$, where pre is a finite set of assertions, the *pre-conditions*, and post is a finite set of *conditional post-conditions* of the form φ/ψ, where φ is an assertion and ψ is a *literal.* Such an action is called *unconditional* if all its post-conditions are of the form $true/\psi$, where "*true*" stands for an assertion that is satisfied in every interpretation. We write such unconditional post-conditions simply as ψ rather than $true/\psi$. △

Basically, an action is applicable in an interpretation if its pre-conditions are satisfied. The conditional post-condition φ/ψ requires that ψ must hold after the application of the action if φ was satisfied before the application. According to the semantics of DL actions defined in [3], nothing should change that is not explicitly required to change by some post-condition. As already discussed in the introduction, this semantics is not appropriate if the domain constraints are given by a TBox containing arbitrary GCIs.

For examples, consider the TBox $\mathcal{T}$ consisting of the GCIs of Example 1 and the action $\mathsf{HireJohn} = (\emptyset, \{\mathsf{Employee}(\mathsf{JOHN})\})$, which has no pre-conditions and a single unconditional post-condition. Assume that $\mathcal{I}$ is a model of $\mathcal{T}$ with $\mathcal{I} \not\models \mathsf{Employee}(\mathsf{JOHN})$ and $\mathcal{I} \not\models \exists\mathsf{insuredBy}.\mathsf{HealthInsuranceCompany}(\mathsf{JOHN})$ (obviously, such models exist). If we apply the semantics of DL actions introduced in [3], then $\mathcal{I}$ is transformed into an interpretation $\mathcal{I}'$, whose only difference to $\mathcal{I}$ is that now John is an employee, i. e., $\mathcal{I}' \models \mathsf{Employee}(\mathsf{JOHN})$. Since nothing else changes, we still have $\mathcal{I}' \not\models \exists\mathsf{insuredBy}.\mathsf{HealthInsuranceCompany}(\mathsf{JOHN})$, which shows that $\mathcal{I}'$ is not a model of $\mathcal{T}$. Consequently, although the action $\mathsf{HireJohn}$ is applicable to $\mathcal{I}$ (since the empty set of pre-conditions does not impose any applicability condition), its application does not result in an interpretation satisfying the domain constraints in $\mathcal{T}$. We will call an action where this kind of problem can occur an *inconsistent action.* In our example, consistency can be achieved by complementing the action $\mathsf{HireJohn}$ with an appropriate causal relationship.

Definition 3. A *causal relationship* is of the form $\mathcal{A}_1 \longrightarrow_{\mathcal{B}} \mathcal{A}_2$, where $\mathcal{A}_1, \mathcal{A}_2$ are simple ABoxes and $\mathcal{B}$ is an ABox. △

Such a causal relationship can be read as "$\mathcal{A}_1$ causes $\mathcal{A}_2$ if $\mathcal{B}$ holds." To be more precise, it says the following:[2] if $\mathcal{B}$ is satisfied *before*[3] the application of an action and $\mathcal{A}_1$ is *newly* satisfied by its application (i. e., was not satisfied before, but is satisfied after the application), then $\mathcal{A}_2$ must also be satisfied *after* the application. In our health insurance example, the causal relationship

$$\{\mathsf{Employee}(\mathsf{JOHN})\} \longrightarrow_{\{\neg\exists\mathsf{insuredBy}.\mathsf{HealthInsuranceCompany}(\mathsf{JOHN})\}} \{\mathsf{insuredBy}(\mathsf{JOHN}, \mathsf{AOK})\}$$

adds the following indirect effect to the direct effect of the $\mathsf{HireJohn}$ action: (i) if John becomes newly employed (i. e., was not an employee before) and did not

[2]Actually, there are different ways of defining the meaning of causal relationships. Here, we follow the approach used in [20, 5] rather than the one employed by [10, 16].

[3]In the semantics of causal relationship introduced in [10, 16], this "before" would need to be replaced by "after."

have a health insurance before the application of the action, then he is newly insured with AOK after its application; (ii) if he becomes newly employed, but already has a health insurance, then he keeps his old health insurance and is *not* newly insured with AOK. In both cases, the GCIs of Example 1 stay satisfied.

Note that causal relationships are not logical formulas in the form of implication. They are rules to deal with the ramification problem in action theory. More sophisticated examples of application domains and causal relationships are presented in [20] when the latter were first introduced to action theory.

In order to define the semantics of DL actions in the presence of causal relationships formally, we consider an action $\alpha = (\mathsf{pre}, \mathsf{post})$, a finite set of causal relationships CR, and an interpretation $\mathcal{I}$ to which the action is supposed to be applied. The actions and causal relationships introduced above can only effect changes to the membership of named individuals (pairs of named individuals) in atomic concepts (roles). Consequently, such *effects* can be described in an obvious way using literals. For this reason, we will sometimes call a simple ABox a *set of effects*.

Using the semantics of actions introduced in [3], the *set of direct effects* of α given $\mathcal{I}$ is defined as $\mathsf{Dir}(\alpha, \mathcal{I}) := \{\psi \mid \varphi/\psi \in \mathsf{post} \wedge \mathcal{I} \models \varphi\}$.

Direct effects of an action may cause indirect effects specified by causal relationships, and these indirect effects may again cause indirect effects, etc. Thus, the overall effects of an action are obtained by iteratively adding indirect effects to the direct ones until no new indirect effects can be added.

To be more precise, we start the iteration by defining $\mathcal{E}_0 := \mathsf{Dir}(\alpha, \mathcal{I})$. Assuming that $\mathcal{E}_i$ ($i \geq 0$) is already defined, we define $\mathcal{E}_{i+1} := \mathcal{E}_i \cup \mathsf{Ind}_{i+1}$, where

$$\begin{aligned}\mathsf{Ind}_{i+1} := \{\psi \mid \;& \exists \mathcal{A}_1 \longrightarrow_{\mathcal{B}} \mathcal{A}_2 \in \mathsf{CR} \text{ such that} \\ & (i)\ \psi \in \mathcal{A}_2,\quad (ii)\ \mathcal{I} \models \mathcal{B},\quad (iii)\ \mathcal{I} \not\models \mathcal{A}_1,\quad \textit{and} \\ & (iv)\ \forall \varphi \in \mathcal{A}_1.\ (\varphi \in \mathcal{E}_i \vee (\mathcal{I} \models \varphi \wedge \dot\neg\varphi \notin \mathcal{E}_i))\}\ .\end{aligned}$$

Thus, we add the indirect effect ψ to our set of effects if (i) it is in the consequence set $\mathcal{A}_2$ of a causal relationship $\mathcal{A}_1 \longrightarrow_{\mathcal{B}} \mathcal{A}_2$ for which (ii) the condition $\mathcal{B}$ is satisfied in $\mathcal{I}$ (i. e., before applying the action), and (iii)+(iv) the trigger $\mathcal{A}_1$ is newly satisfied, i. e., (iii) $\mathcal{A}_1$ is not satisfied in $\mathcal{I}$, but (iv) it is satisfied according to the current effect set, i. e., every assertion $\varphi \in \mathcal{A}_1$ is a (direct or indirect) effect, or it is satisfied in $\mathcal{I}$ and this is not changed by an effect.

By definition, we have $\mathcal{E}_0 \subseteq \mathcal{E}_1 \subseteq \mathcal{E}_2 \cdots$. Since we only add literals that belong to the consequence set of a causal relationship in the finite set CR, there is an n such that $\mathcal{E}_n = \mathcal{E}_{n+1} = \mathcal{E}_{n+2} = \cdots$. We define $\mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR}) := \mathcal{E}_n$. This set of literals represents the effects of applying the action α to the interpretation $\mathcal{I}$ w. r. t. the causal relationships in CR. It could happen, however, that this set is contradictory, and thus cannot lead to a well-defined successor interpretation: we say that $\mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})$ is *contradictory* if there is a literal L such that $\{L, \dot{\neg} L\} \subseteq \mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})$.

Now, we are ready to introduce our semantics of actions in the presence of causal relationships.

Definition 4. Let α be an action, CR a finite set of causal relationships, $\mathcal{T}$ a TBox, and $\mathcal{I}, \mathcal{I}'$ two interpretations. We say that α *may transform* $\mathcal{I}$ *to* $\mathcal{I}'$ *w. r. t.* $\mathcal{T}$ *and* CR (denoted by $\mathcal{I} \Longrightarrow_{\alpha}^{\mathcal{T},\mathsf{CR}} \mathcal{I}'$) if

- $\Delta^{\mathcal{I}} = \Delta^{\mathcal{I}'}$ and $a^{\mathcal{I}} = a^{\mathcal{I}'}$ for every individual name a,
- $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I}' \models \mathcal{T}$,
- $\mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})$ is not contradictory,
- for all concept names A we have $A^{\mathcal{I}'} = (A^{\mathcal{I}} \cup \{a^{\mathcal{I}} \mid A(a) \in \mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})\}) \setminus \{a^{\mathcal{I}} \mid \neg A(a) \in \mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})\}$, and
- for all role names r we have $r^{\mathcal{I}'} = (r^{\mathcal{I}} \cup \{(a^{\mathcal{I}}, b^{\mathcal{I}}) \mid r(a, b) \in \mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})\}) \setminus \{(a^{\mathcal{I}}, b^{\mathcal{I}}) \mid \neg r(a, b) \in \mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})\}$.

The sequence of actions $\alpha_1, \ldots, \alpha_n$ *may transform* $\mathcal{I}$ *to* $\mathcal{I}'$ *w. r. t.* $\mathcal{T}$ *and* CR (denoted by $\mathcal{I} \Longrightarrow_{\alpha_1,\ldots,\alpha_n}^{\mathcal{T},\mathsf{CR}} \mathcal{I}'$) iff there are interpretations $\mathcal{I}_0, \ldots, \mathcal{I}_n$ such that $\mathcal{I} = \mathcal{I}_0$, $\mathcal{I}_n = \mathcal{I}'$, and $\mathcal{I}_{i-1} \Longrightarrow_{\alpha_i}^{\mathcal{T},\mathsf{CR}} \mathcal{I}_i$ for all $i, 1 \leq i \leq n$. △

If $\mathcal{T}$ and CR are empty, then this semantics coincides with the one given in [3] for actions without occlusions. Note that our actions are *deterministic* in the sense that, for every model $\mathcal{I}$ of $\mathcal{T}$, there exists at most one interpretation $\mathcal{I}'$ such that $\mathcal{I} \Longrightarrow_{\alpha}^{\mathcal{T},\mathsf{CR}} \mathcal{I}'$. However, sometimes there may not exist any such interpretation $\mathcal{I}'$, either because $\mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})$ is contradictory, or because the new interpretation induced by $\mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})$ is not a model of $\mathcal{T}$. If this happens in the case where $\alpha = (\mathsf{pre}, \mathsf{post})$ is actually *applicable* to $\mathcal{I}$ (i. e., $\mathcal{I} \models \mathsf{pre}$), then this indicates a modeling error. In fact, the correct modeling of an action theory should ensure that, whenever an action is applicable, there is a well-defined successor state.

Definition 5. The action α is *consistent* w. r. t. the TBox $\mathcal{T}$ and the finite set CR of causal relationships iff, for every model $\mathcal{I}$ of $\mathcal{T}$ with $\mathcal{I} \models \mathsf{pre}$, there exists an interpretation $\mathcal{I}'$ with $\mathcal{I} \Longrightarrow_{\alpha}^{\mathcal{T},\mathsf{CR}} \mathcal{I}'$. △

As argued above, the action $\mathsf{HireJohn}$ is not consistent w. r. t. the TBox consisting of the GCIs of Example 1 and the empty set of causal relationships, but it becomes consistent if we add the causal relationship introduced below Definition 3.

The projection problem is one of the most basic reasoning problems for action theories [13]. Given a (possibly incomplete) description of the initial world (interpretation), it asks whether a certain property is guaranteed to hold after the execution of a sequence of actions. Our formal definition of this problems is taken from [3], with the only difference that we use the “may transform” relation introduced in Definition 4, which takes causal relationships into account, instead of the one employed in [3].

Definition 6 (Projection problem). Let $\alpha_1, \ldots, \alpha_n$ be a sequence of actions such that, for all $i, 1 \leq i \leq n$, the action α_i is consistent w. r. t. $\mathcal{T}$ and CR. The assertion φ is *a consequence of applying* $\alpha_1, \ldots, \alpha_n$ *to* $\mathcal{A}$ *w. r. t.* $\mathcal{T}$ *and* CR iff, for all $\mathcal{I}$ and $\mathcal{I}'$, if $\mathcal{I} \models \mathcal{A}$ and $\mathcal{I} \Longrightarrow^{\mathcal{T},\mathsf{CR}}_{\alpha_1,\ldots,\alpha_n} \mathcal{I}'$, then $\mathcal{I}' \models \varphi$. $\triangle$

Note that we consider only consistent actions in our definition of the projection problem. In fact, if an action is inconsistent, then there is something wrong with the action theory, and this problem should be solved before starting to ask projection questions. Another interesting inference problem for action theories is *executability*: Are all pre-conditions guaranteed to be satisfied during the execution of a sequence of actions? As shown in [3], the projection and the executability problem can be reduced to each other in polynomial time. For this reason, we restrict our attention to the consistency and the projection problem.

4 Deciding Consistency

First, we develop a solution for the restricted case where the TBox is empty, and then we show how this solution can be extended to the general case.

4.1 Consistency w. r. t. the Empty TBox

We will show that, in this case, testing consistency of an action w. r. t. a set of causal relationships has the same complexity as the (in)consistency problem of an ABox. Given an action α and a finite set of causal relationships CR, we basically consider all the possible situations that the action could encounter when it is applied to an interpretation.

Definition 7. Let $\alpha = (\mathsf{pre}, \mathsf{post})$ be an action and CR a finite set of causal relationships. The ABox $\mathcal{A}(\alpha, \mathsf{CR})$ is defined as follows:

$$\mathcal{A}(\alpha, \mathsf{CR}) := \{\varphi, \neg\varphi \mid \varphi/\psi \in \mathsf{post} \text{ or } \varphi \in \mathcal{A}_1 \cup \mathcal{B} \text{ for some } \mathcal{A}_1 \longrightarrow_{\mathcal{B}} \mathcal{A}_2 \in \mathsf{CR}\}.$$

A *diagram* $\mathcal{D}$ *for* α *and* CR is a maximal, consistent subset of $\mathcal{A}(\alpha, \mathsf{CR})$, i. e., there is no consistent subset $\mathcal{D}'$ of $\mathcal{A}(\alpha, \mathsf{CR})$ such that $\mathcal{D} \subseteq \mathcal{D}'$ and $\mathcal{D} \neq \mathcal{D}'$. We denote the set of all diagrams for α and CR by $\mathfrak{D}(\alpha, \mathsf{CR})$. $\triangle$

As a direct consequence of maximality of a diagram $\mathcal{D}$, we have for all assertions $\varphi \in \mathcal{A}(\alpha, \mathsf{CR})$, either $\varphi \in \mathcal{D}$ or $\neg\varphi \in \mathcal{D}$. For a given interpretation $\mathcal{I}$, there is exactly one diagram $\mathcal{D}$ such that $\mathcal{I} \models \mathcal{D}$. It is sufficient to know this diagram to determine what are the direct and indirect effects of applying α to $\mathcal{I}$ w. r. t. CR.

Lemma 8. *Let $\alpha = (\mathsf{pre}, \mathsf{post})$ be an action and CR a finite set of causal relationships. For a given interpretation $\mathcal{I}$, there is exactly one diagram $\mathcal{D}$ for α and CR such that $\mathcal{I} \models \mathcal{D}$.*

Proof. Given an interpretation $\mathcal{I}$, we define $\mathcal{D} := \{\varphi \in \mathcal{A}(\alpha, \mathsf{CR}) \mid \mathcal{I} \models \varphi\}$. Obviously, $\mathcal{D}$ is a non-empty subset of $\mathcal{A}(\alpha, \mathsf{CR})$. $\mathcal{D}$ is consistent (since $\mathcal{I} \models \mathcal{D}$) and maximal since for all $\varphi \in \mathcal{A}(\alpha, \mathsf{CR})$ either $\varphi \in \mathcal{D}$ or $\neg\varphi \in \mathcal{D}$.

$\mathcal{D}$ is the only diagram with $\mathcal{I} \models \mathcal{D}$, which we will prove by contraposition. Assume there exists some $\mathcal{D}' \in \mathcal{A}(\alpha, \mathsf{CR})$ such that $\mathcal{I} \models \mathcal{D}'$ and $\mathcal{D} \neq \mathcal{D}'$. Since $\mathcal{D}$ and $\mathcal{D}'$ are non-equal diagrams, and thus non-equal maximal subsets of $\mathcal{A}(\alpha, \mathsf{CR})$, there exists some $\varphi \in \mathcal{A}(\alpha, \mathsf{CR})$ such that $\varphi \in \mathcal{D}$ and $\neg\varphi \in \mathcal{D}'$. Since $\mathcal{I} \models \mathcal{D}$ and $\mathcal{I} \models \mathcal{D}'$, we have that $\mathcal{I} \models \varphi$ and $\mathcal{I} \models \neg\varphi$, i. e. $\mathcal{I} \not\models \varphi$, which is a contradiction. ❏

Given a diagram $\mathcal{D}$, we will now define a set $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$ such that $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR}) = \mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})$ for every interpretation $\mathcal{I}$ with $\mathcal{I} \models \mathcal{D}$. The definition of the direct effects of an action can easily be adapted to the diagram case: $\widehat{\mathsf{Dir}}(\alpha, \mathcal{D}) := \{\psi \mid \varphi/\psi \in \mathsf{post} \wedge \varphi \in \mathcal{D}\}$.

The same is true for the sets $\mathcal{E}_i$. We start the iteration by defining $\widehat{\mathcal{E}}_0 := \widehat{\mathsf{Dir}}(\alpha, \mathcal{D})$. Assuming that $\widehat{\mathcal{E}}_i$ $(i \geq 0)$ is already defined, we define $\widehat{\mathcal{E}}_{i+1} := \widehat{\mathcal{E}}_i \cup \widehat{\mathsf{Ind}}_{i+1}$, where

$$\begin{aligned}\widehat{\mathsf{Ind}}_{i+1} := \{\psi \mid\ & \exists \mathcal{A}_1 \longrightarrow_{\mathcal{B}} \mathcal{A}_2 \in \mathsf{CR} \text{ such that} \\ & (i)\ \psi \in \mathcal{A}_2,\quad (ii)\ \mathcal{B} \subseteq \mathcal{D},\quad (iii)\ \mathcal{A}_1 \not\subseteq \mathcal{D},\quad \textit{and} \\ & (iv)\ \forall \varphi \in \mathcal{A}_1.\ (\varphi \in \widehat{\mathcal{E}}_i \vee (\varphi \in \mathcal{D} \wedge \dot{\neg}\varphi \notin \widehat{\mathcal{E}}_i))\}\ .\end{aligned}$$

Again, there exists an $n \geq 0$ such that $\widehat{\mathcal{E}}_n = \widehat{\mathcal{E}}_{n+1} = \widehat{\mathcal{E}}_{n+2} = \cdots$, and we define $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR}) := \widehat{\mathcal{E}}_n$. This set is *contradictory* if there is a literal L such that $\{L, \dot{\neg}L\} \subseteq \widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$.

Lemma 9. *Let α be an action, CR be a finite set of causal relationships, and $\mathcal{D}$ a diagram for α and CR. Then, for every interpretation $\mathcal{I}$ with $\mathcal{I} \models \mathcal{D}$, we have that $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR}) = \mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})$.*

Proof. Let $\mathcal{D}$ be a diagram for α and CR and let $\mathcal{I}$ be an interpretation such that $\mathcal{I} \models \mathcal{D}$. We prove the lemma by proving $\widehat{\mathcal{E}}_i = \mathcal{E}_i$ by induction on $i \geq 0$.

Claim 1. Let φ be an assertion in $\mathcal{A}(\alpha, \mathsf{CR})$ and $\mathcal{A} \subseteq \mathcal{A}(\alpha, \mathsf{CR})$ be an ABox. Then,

(i) $\mathcal{I} \models \varphi$ iff $\varphi \in \mathcal{D}$; and

(ii) $\mathcal{I} \models \mathcal{A}$ iff $\mathcal{A} \subseteq \mathcal{D}$.

Proof of Claim 1:

(i) "$\Rightarrow$": Assume that $\mathcal{I} \models \varphi$ and $\varphi \notin \mathcal{D}$. Since $\mathcal{D}$ is maximal, $\neg\varphi \in \mathcal{D}$. $\mathcal{I} \models \mathcal{D}$ implies that $\mathcal{I} \not\models \varphi$, which is a contradiction. "$\Leftarrow$": Since $\mathcal{I} \models \mathcal{D}$ and $\varphi \in \mathcal{D}$, we have $\mathcal{I} \models \varphi$.

(ii) "$\Rightarrow$": Let φ be an assertion in $\mathcal{A}$. Since $\mathcal{I} \models \mathcal{A}$, we have $\mathcal{I} \models \varphi$, which, by (i), implies that $\varphi \in \mathcal{D}$. "$\Leftarrow$": Let φ be an assertion in $\mathcal{A}$. Since $\mathcal{A} \subseteq \mathcal{D}$, we have $\varphi \in \mathcal{D}$, which, by (i), implies that $\mathcal{I} \models \varphi$.

This finishes the proof of Claim 1.

For the *induction start*, we prove $\widehat{\mathcal{E}}_0 = \mathcal{E}_0$, i. e. $\widehat{\mathsf{Dir}}(\alpha, \mathcal{D}) = \mathsf{Dir}(\alpha, \mathcal{I})$. We have:

$$\begin{array}{lll} & \chi \in \widehat{\mathsf{Dir}}(\alpha, \mathcal{D}) & \\ \text{iff} & \chi \in \{\psi \mid \varphi/\psi \in \mathsf{post} \wedge \varphi \in \mathcal{D}\} & \text{(by the definition of } \widehat{\mathsf{Dir}}(\alpha, \mathcal{D})) \\ \text{iff} & \chi \in \{\psi \mid \varphi/\psi \in \mathsf{post} \wedge \mathcal{I} \models \varphi\} & \text{(by Claim 1 (i))} \\ \text{iff} & \chi \in \mathsf{Dir}(\alpha, \mathcal{I}) & \text{(by the definition of } \mathsf{Dir}(\alpha, \mathcal{I}))\ . \end{array}$$

For the *induction step*, suppose we have $\widehat{\mathcal{E}}_n = \mathcal{E}_n$ for some $n \geq 0$. We show $\widehat{\mathcal{E}}_{n+1} = \mathcal{E}_{n+1}$. We have:

$$\begin{array}{ll} & \chi \in \widehat{\mathcal{E}}_{n+1} \\ \text{iff} & \chi \in \widehat{\mathcal{E}}_n \cup \widehat{\mathsf{Ind}}_{n+1} \ \ \text{(by the definition of } \widehat{\mathcal{E}}_{n+1}) \\ \text{iff} & \chi \in \widehat{\mathcal{E}}_n \cup \{\psi \mid \exists \mathcal{A}_1 \longrightarrow_{\mathcal{B}} \mathcal{A}_2 \in \mathsf{CR} \text{ such that} \\ & \qquad (i)\ \psi \in \mathcal{A}_2, \quad (ii)\ \mathcal{B} \in \mathcal{D}, \quad (iii)\ \mathcal{A}_1 \not\subseteq \mathcal{D}, \quad \text{and} \\ & \qquad (iv)\ \forall \varphi \in \mathcal{A}_1.\ (\varphi \in \widehat{\mathcal{E}}_i \vee (\varphi \in \mathcal{D} \wedge \dot{\neg}\varphi \notin \widehat{\mathcal{E}}_i))\} \\ & \text{(by the definition of } \widehat{\mathsf{Ind}}_{n+1}) \\ \text{iff} & \chi \in \mathcal{E}_n \cup \{\psi \mid \exists \mathcal{A}_1 \longrightarrow_{\mathcal{B}} \mathcal{A}_2 \in \mathsf{CR} \text{ such that} \\ & \qquad (i)\ \psi \in \mathcal{A}_2, \quad (ii)\ \mathcal{B} \in \mathcal{D}, \quad (iii)\ \mathcal{A}_1 \not\subseteq \mathcal{D}, \quad \text{and} \\ & \qquad (iv)\ \forall \varphi \in \mathcal{A}_1.\ (\varphi \in \mathcal{E}_i \vee (\varphi \in \mathcal{D} \wedge \dot{\neg}\varphi \notin \mathcal{E}_i))\} \\ & \text{(by the induction hypothesis)} \\ \text{iff} & \chi \in \mathcal{E}_n \cup \{\psi \mid \exists \mathcal{A}_1 \longrightarrow_{\mathcal{B}} \mathcal{A}_2 \in \mathsf{CR} \text{ such that} \\ & \qquad (i)\ \psi \in \mathcal{A}_2, \quad (ii)\ \mathcal{I} \models \mathcal{B}, \quad (iii)\ \mathcal{I} \not\models \mathcal{A}_1, \quad \text{and} \\ & \qquad (iv)\ \forall \varphi \in \mathcal{A}_1.\ (\varphi \in \mathcal{E}_i \vee (\mathcal{I} \models \varphi \wedge \dot{\neg}\varphi \notin \mathcal{E}_i))\} \\ & \text{(by Claim1 (ii))} \\ \text{iff} & \chi \in \mathcal{E}_n \cup \mathsf{Ind}_{n+1} \ \ \text{(by the definition of } \mathsf{Ind}_{n+1}) \\ \text{iff} & \chi \in \mathcal{E}_{n+1} \qquad\qquad \text{(by the definition of } \mathcal{E}_{n+1})\ . \end{array}$$

❑

Checking which of the sets $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$ for $\mathcal{D} \in \mathfrak{D}(\alpha, \mathsf{CR})$ are contradictory is sufficient for deciding the consistency problem in the case where the TBox is assumed to be empty. In fact, in this case the only reason for an interpretation not to have a successor interpretation w. r. t. α is that the set of effects is contradictory. Since we require the existence of a successor interpretation only for interpretations that satisfy the precondition set pre of α, it is enough to consider diagrams $\mathcal{D}$ that are consistent with pre.

Lemma 10. *The action* $\alpha = (\mathsf{pre}, \mathsf{post})$ *is consistent w. r. t.* CR iff $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$ *is not contradictory for all* $\mathcal{D} \in \mathfrak{D}(\alpha, \mathsf{CR})$ *for which* $\mathcal{D} \cup \mathsf{pre}$ *is consistent.*

Proof. "$\Rightarrow$": Assume that there exists some $\mathcal{D} \in \mathfrak{D}(\alpha, \mathsf{CR})$ such that $\mathcal{D} \cup \mathsf{pre}$ is consistent and $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$ is contradictory. The former implies that there exists an interpretation $\mathcal{I}$ such that $\mathcal{I} \models \mathcal{D} \cup \mathsf{pre}$. By Lemma 9, the latter implies $\mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})$ is contradictory. Thus, $\mathcal{I}$ has no successor w. r. t. $\Longrightarrow_{\alpha}^{\emptyset, \mathsf{CR}} \mathcal{I}'$, which is a contradiction.

"$\Leftarrow$": Assume α is not consistent w. r. t. CR, i. e. there exists an interpretation $\mathcal{I}$ such that $\mathcal{I} \models \mathsf{pre}$ and $\mathcal{I}$ has no successor w. r. t. $\Longrightarrow_{\alpha}^{\emptyset, \mathsf{CR}}$, which implies that $\mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})$ is contradictory. We define $\mathcal{D}$ as in Lemma 8. Thus, we know that $\mathcal{I} \models \mathcal{D}$ and that $\mathcal{D}$ is a diagram for α and CR. By Lemma 9, $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$ is contradictory. We know that $\mathcal{D} \cup \mathsf{pre}$ is consistent since $\mathcal{I} \models \mathcal{D}$ and that $\mathcal{I} \models \mathsf{pre}$. ❑

This lemma yields a decision procedure for deciding consistency of an action w. r. t. a finite set of causal relationships. For DLs between $\mathcal{ALC}$ [15] and $\mathcal{ALCQO}$ [4] (unary coding for qualified number restrictions) and DLs between $\mathcal{ALC}$ and $\mathcal{ALCQI}$ [19] (even binary coding for qualified number restrictions) for which deciding the consistency problem is PSpace-complete, in order to check whether α is *inconsistent*, we first guess[4] a diagram $\mathcal{D} \in \mathfrak{D}(\alpha, \mathsf{CR})$, and then check whether $\mathcal{D} \cup \mathsf{pre}$ is consistent using the PSpace decision procedure for ABox consistency. If $\mathcal{D} \cup \mathsf{pre}$ is consistent, we compute the set $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$. This can be realized in polynomial time by performing the iteration used in the definition of $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$. Checking whether this set is contradictory is obviously also possible in polynomial time. Overall, it is in PSpace.

For the DL $\mathcal{ALCIO}$ for which deciding ABox consistency is ExpTime-complete [1], in order to check whether α is consistent, we perform exponentially many ExpTime tests (each test is for a diagram), which yields overall an ExpTime decision procedure.

If the underlying DL is $\mathcal{ALCQIO}$, for which the ABox consistency is NExpTime-complete [18, 12] (even binary coding for qualified number restrictions), then we

[4] Recall that PSpace = NPSpace according to Savitch's theorem.

can use the above procedure by guessing a diagram and thus obtain a NExpTime procedure for deciding an action's *inconsistency*.

Those upper bounds are optimal since the ABox inconsistency problem can be reduced to our action consistency problem: for every ABox $\mathcal{A}$, we have that $\mathcal{A}$ is inconsistent iff $(\mathcal{A}, \{A(a), \neg A(a)\})$ is consistent w. r. t. the empty set of causal relationships, where A is an arbitrary concept name and a is an arbitrary individual name.

Theorem 11. *The problem of deciding consistency of an action w. r. t. a finite set of causal relationships is*

- PSpace-*complete for DLs between $\mathcal{ALC}$ and $\mathcal{ALCQO}$ if the numbers in qualified number restrictions are coded in unary;*
- PSpace-*complete for DLs between $\mathcal{ALC}$ and $\mathcal{ALCQI}$;*
- ExpTime-*complete for $\mathcal{ALCIO}$;*
- *co*-NExpTime-*complete for $\mathcal{ALCQIO}$.*

4.2 The General Case

If $\mathcal{T}$ is not empty, then there is an additional possible reason for an action to be inconsistent: the successor interpretation induced by a non-contradictory set of effects may not be a model of $\mathcal{T}$. Thus, given a non-contradictory set of effects $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$, we must check whether, for any model $\mathcal{I}$ of $\mathcal{T}$ and $\mathcal{D}$ that satisfies the preconditions of α, the interpretation $\mathcal{I}'$ obtained from $\mathcal{I}$ by applying the effects in $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$ (see Definition 4) is a model of $\mathcal{T}$. To this purpose, we first define an unconditional action $\beta_{\alpha,\mathsf{CR},\mathcal{D}}$ that, applied to models of $\mathcal{D}$, has the same effect as α w. r. t. CR. Then, we adapt the approach for solving the projection problem introduced in [3] to the problem of checking whether $\beta_{\alpha,\mathsf{CR},\mathcal{D}}$ transforms models of $\mathcal{T}$ into models of $\mathcal{T}$.

Definition 12. Let $\alpha = (\mathsf{pre}, \mathsf{post})$ be an action, CR a finite set of causal relationships, and $\mathcal{D} \in \mathfrak{D}(\alpha, \mathsf{CR})$. The action $\beta_{\alpha,\mathsf{CR},\mathcal{D}}$ has $\mathsf{pre} \cup \mathcal{D}$ as set of pre-conditions and $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$ as set of (unconditional) post-conditions. △

The following lemma is an easy consequence of the definition of $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$ and the semantics of actions (Definition 4).

Lemma 13. *For all $\mathcal{D} \in \mathfrak{D}(\alpha, \mathsf{CR})$, all models $\mathcal{I}$ of $\mathcal{D}$, and all interpretations $\mathcal{I}'$, we have $\mathcal{I} \Longrightarrow_{\alpha}^{\emptyset,\mathsf{CR}} \mathcal{I}'$ iff $\mathcal{I} \Longrightarrow_{\beta_{\alpha,\mathsf{CR},\mathcal{D}}}^{\emptyset,\emptyset} \mathcal{I}'$.*

Proof. Let $\mathcal{D}$ be a diagram for α and CR. Let $\mathcal{I}$ be an interpretation such that $\mathcal{I} \models \mathcal{D}$. By the construction of $\beta_{\alpha,\mathsf{CR},\mathcal{D}}$, we have that the direct effects of $\beta_{\alpha,\mathsf{CR},\mathcal{D}}$ given $\mathcal{D}$ are identical to the accumulated effects of α given $\mathcal{D}$, i. e., $\widehat{\mathsf{Dir}}(\beta_{\alpha,\mathsf{CR},\mathcal{D}}, \mathcal{D}) = \widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$. Thus, $\widehat{\mathcal{E}}(\beta_{\alpha,\mathsf{CR},\mathcal{D}}, \mathcal{D}, \emptyset) = \widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$. This implies, by Lemma 9, that we have $\mathcal{E}(\beta_{\alpha,\mathsf{CR},\mathcal{D}}, \mathcal{I}, \emptyset) = \mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})$, which, together with Definition 4, yields that $\mathcal{I} \Longrightarrow_{\alpha}^{\emptyset,\mathsf{CR}} \mathcal{I}'$ iff $\mathcal{I} \Longrightarrow_{\beta_{\alpha,\mathsf{CR},\mathcal{D}}}^{\emptyset,\emptyset} \mathcal{I}'$. ❑

The approach for solving the projection problem introduced in [3] considers a finite sequence of actions $\beta_1, \ldots, \beta_n$. In the present section, we are only interested in the special case where $n = 1$. However, since we will adopt this approach also in Section 5, where we consider the case $n \geq 1$, we recall the relevant notions and results for the general case. In this approach, time-stamped copies $r^{(i)}$ $(0 \leq i \leq n)$ for all relevant role names and new time-stamped concept names $T_C^{(i)}$ $(0 \leq i \leq n)$ for all relevant concept descriptions are introduced.

For the projection problem in [4], the ABox $\mathcal{A}_{\mathsf{red}}$ and TBox $\mathcal{T}_{\mathsf{red}}$ in the reduction are constructed based on the role names and concept descriptions in its input $\mathcal{A}$, $\mathcal{T}$, $\beta_1, \ldots, \beta_n$, and φ. On the one hand, we construct $\mathcal{A}_{\mathsf{red}}$ and $\mathcal{T}_{\mathsf{red}}$ based only on a sequence of actions, i. e., without $\mathcal{A}$, $\mathcal{T}$, and φ, to ensure the semantics of the action, as we will see in Lemma 14. On the other hand, we need to refer to role names and concept descriptions with a time stamp, which occur in the input of the consistency or projection problem in this paper (for the projection problem, see Section 5), but not in the sequence of actions. Thus, we need to take those role names and concept descriptions into account when $\mathcal{A}_{\mathsf{red}}$ and $\mathcal{T}_{\mathsf{red}}$ are constructed. Since the number of role names and concept descriptions are polynomially bounded by the size of the input, this does not spoil the polynomial size of $\mathcal{A}_{\mathsf{red}}$ and $\mathcal{T}_{\mathsf{red}}$. Moreover, the desired property (1) which is described before Lemma 14 still holds with this addition of roles and concept descriptions.

In our setting, the *relevant* role names (concept descriptions) will be the ones occurring in the input of the consistency algorithm. For every ABox assertion φ built using a relevant concept description C or a relevant role name r (called relevant assertion in the following) and every $i, 0 \leq i \leq n$, we can then define a time-stamped variant $\varphi^{(i)}$ as follows:

$$C(a)^{(i)} := T_C^{(i)}(a), \quad r(a,b)^{(i)} := r^{(i)}(a,b), \quad \neg r(a,b)^{(i)} := \neg r^{(i)}(a,b).$$

Given a set of relevant assertions $\mathcal{A}$, we define its time-stamped copy as $\mathcal{A}^{(i)} := \{\varphi^{(i)} \mid \varphi \in \mathcal{A}\}$. Given a set of GCIs $\mathcal{T}$ built from relevant concept descriptions, we define its time-stamped copy as $\mathcal{T}^{(i)} := \{T_C^{(i)} \sqsubseteq T_D^{(i)} \mid C \sqsubseteq D \in \mathcal{T}\}$.

Intuitively, given an initial interpretation $\mathcal{I}_0$, the application of β_1 to $\mathcal{I}_0$ yields a successor interpretation $\mathcal{I}_1$, the application of β_2 to $\mathcal{I}_1$ yields a successor interpretation $\mathcal{I}_2$, etc. Using the time-stamped copies of the relevant role names and concept descriptions, we can encode the sequence of interpretations $\mathcal{I}_0, \mathcal{I}_1, \ldots, \mathcal{I}_n$

into a single interpretation $\mathcal{J}$ such that

$$\text{the relevant assertion } \varphi \text{ holds in } \mathcal{I}_i \text{ iff} \qquad \text{its time-stamped variant } \varphi^{(i)} \text{ holds in } \mathcal{J}. \tag{1}$$

In order to enforce that $\mathcal{J}$ really encodes a sequence of interpretations induced by the application of the action sequence $\beta_1, \ldots, \beta_n$, we require it to be a model of the (acyclic) TBox $\mathcal{T}_{\mathsf{red}}$ and the ABox $\mathcal{A}_{\mathsf{red}}$. The construction of $\mathcal{T}_{\mathsf{red}}$ and $\mathcal{A}_{\mathsf{red}}$ is very similar to the one introduced in [4] with the only difference that we use more relevant role names and concept descriptions as we explained above.[5]

Now we recall the construction of $\mathcal{A}_{\mathsf{red}}$ and $\mathcal{T}_{\mathsf{red}}$ in [4].[6] Let $\beta_1, \ldots, \beta_n$ be a sequence of (unconditional) actions and $\mathcal{R}$ a set of relevant concept descriptions and rrelevant role names. We define Obj as the set of the individual names occurring in the input of the consistency problem.

$$\mathcal{T}_N = \{N \equiv \bigsqcup_{a \in \mathsf{Obj}} \{a\}\}.$$

The TBox $\mathcal{T}_{\mathsf{Sub}}$ consists of a concept definition of $T_C^{(i)}$ for every $C \in \mathcal{R}$ and for every $i \leq n$. The concept definition of $T_C^{(i)}$ is defined inductively on the structure of C as described in Figure 1. We are now ready to assemble $\mathcal{T}_{\mathsf{red}}$:

$$\mathcal{T}_{\mathsf{red}} = \mathcal{T}_N \cup \mathcal{T}_{\mathsf{Sub}}.$$

Let post_i be the post-conditions of β_i for all i with $1 \leq i \leq n$. We define $\mathcal{A}_{\mathsf{post}}^{(i)} = \{\gamma^{(i)} \mid \gamma \in \mathsf{post}_i\}$. For $1 \leq i \leq n$ the ABox $\mathcal{A}_{\mathsf{min}}^{(i)}$ consists of

1. the following assertions for every $a \in \mathsf{Obj}$ and every concept name in the input:

$$\begin{array}{lcl} a & : & (A^{(i-1)} \to A^{(i)}) \text{ if } \neg A(a) \notin \mathsf{post}_{i-1} \\ a & : & (\neg A^{(i-1)} \to \neg A^{(i)}) \text{ if } A(a) \notin \mathsf{post}_{i-1} \end{array}$$

2. the following assertions for all $a, b \in \mathsf{Obj}$ and every role name r in the input:

$$\begin{array}{lcl} a & : & (\exists r^{(i-1)}.\{b\} \to \exists r^{(i)}.\{b\}) \text{ if } \neg r(a,b) \notin \mathsf{post}_{i-1} \\ a & : & (\forall r^{(i-1)}.\neg\{b\} \to \forall r^{(i)}.\neg\{b\}) \text{ if } r(a,b) \notin \mathsf{post}_{i-1}. \end{array}$$

[5] Note that this construction makes use of nominals.

[6] Compared to the original construction of $\mathcal{A}_{\mathsf{red}}$ and $\mathcal{T}_{\mathsf{red}}$, the present construction is simplifed since it does not need to consider the acyclic TBox as domain constraints and it deal only with unconditional actions. Moreover, it corrects a small error in the concept definition for at-most number restriction $(\geqslant n\ r\ C)$ in $\mathcal{T}_{\mathsf{Sub}}$.

$$
\begin{array}{rcl}
T^{(i)}_{A} & \equiv & (N \sqcap A^{(i)}) \sqcup (\neg N \sqcap A^{(0)}) \\
T^{(i)}_{\{a\}} & \equiv & \{a\} \\
T^{(i)}_{\neg C} & \equiv & \neg T^{(i)}_{C} \\
T^{(i)}_{C \sqcap D} & \equiv & T^{(i)}_{C} \sqcap T^{(i)}_{D} \\
T^{(i)}_{C \sqcup D} & \equiv & T^{(i)}_{C} \sqcup T^{(i)}_{D} \\
T^{(i)}_{\exists r.C} & \equiv & \Big(N \sqcap \big((\exists r^{(0)}.(\neg N \sqcap T^{(i)}_{C})) \sqcup (\exists r^{(i)}.(N \sqcap T^{(i)}_{C}))\big)\Big) \\
 & & \sqcup\, (\neg N \sqcap \exists r^{(0)}.T^{(i)}_{C})) \\
T^{(i)}_{\forall r.C} & \equiv & \Big(N \rightarrow \big((\forall r^{(0)}.(\neg N \rightarrow T^{(i)}_{C})) \sqcap (\forall r^{(i)}.(N \rightarrow T^{(i)}_{C}))\big)\Big) \\
 & & \sqcap\, (\neg N \rightarrow \forall r^{(0)}.T^{(i)}_{C}) \\
T^{(i)}_{(\geqslant n\; r\; C)} & \equiv & \Big(N \sqcap \bigsqcup_{0 \leq j \leq \min\{n, \#\mathsf{Obj}\}} \big((\geqslant j\; r^{(i)}\; (N \sqcap T^{(i)}_{C})) \sqcap \\
 & & \qquad (\geqslant (n-j)\; r^{(0)}\; (\neg N \sqcap T^{(i)}_{C}))\big)\Big) \\
 & & \sqcup\, (\neg N \sqcap (\geqslant n\; r^{(0)}\; T^{(i)}_{C})) \\
T^{(i)}_{(\leqslant n\; r\; C)} & \equiv & \Big(N \rightarrow \bigsqcap_{0 \leq j \leq \min\{n+1, \#\mathsf{Obj}\}} \big(\neg(\geqslant j\; r^{(i)}\; (N \sqcap T^{(i)}_{C})) \sqcup \\
 & & \qquad \neg(\geqslant (n-j)\; r^{(0)}\; (\neg N \sqcap T^{(i)}_{C}))\big)\Big) \\
 & & \sqcap\, (\neg N \rightarrow (\leqslant n\; r^{(0)}\; T^{(i)}_{C}))
\end{array}
$$

Figure 1: Concept definitions in $\mathcal{T}_{\mathsf{Sub}}$.

The ABox $\mathcal{A}_{\mathsf{ini}}$ is defined as follows:

$$\mathcal{A}_{\mathsf{ini}} = \{\varphi^{(0)} \mid \varphi \in \mathcal{A}\}.$$

Then, we construct $\mathcal{A}_{\mathsf{red}}$:

$$\mathcal{A}_{\mathsf{red}} = \mathcal{A}_{\mathsf{ini}} \cup \bigcup_{i=1}^{n} \mathcal{A}^{(i)}_{\mathsf{post}} \cup \bigcup_{i=1}^{n} \mathcal{A}^{(i)}_{\mathsf{min}} \ .$$

We recall the pertinent properties of $\mathcal{T}_{\mathsf{red}}$ and $\mathcal{A}_{\mathsf{red}}$ in the next lemma (whose proof is very similar to the one of Theorem 14 in [4]).

Lemma 14. *Let $\mathcal{L}$ be a DL between $\mathcal{ALC}$ and $\mathcal{ALCQIO}$ and $\mathcal{LO}$ the DL which extends $\mathcal{L}$ with nominals. Let $\beta_1, \ldots, \beta_n$ be a sequence of $\mathcal{L}$ actions, and $\mathcal{R}$ a set of relevant role names and concept descriptions such that $\mathcal{R}$ contains all the role names and concept descriptions occurring in $\beta_1, \ldots, \beta_n$. Then, there are an $\mathcal{LO}$ ABox $\mathcal{A}_{\mathsf{red}}$ and an (acyclic) $\mathcal{LO}$ TBox $\mathcal{T}_{\mathsf{red}}$ of size polynomial in the size of $\beta_1, \ldots, \beta_n$ and $\mathcal{R}$, such that the following properties (a) and (b) hold:*

(a) For all interpretations $\mathcal{I}_0, \ldots, \mathcal{I}_n$ such that $\mathcal{I}_i \Longrightarrow^{\emptyset,\emptyset}_{\beta_i} \mathcal{I}_{i+1}$ for every $i, 0 \leq i < n$, there exists an interpretation $\mathcal{J}$ such that $\mathcal{J} \models \mathcal{A}_{\mathsf{red}}$, $\mathcal{J} \models \mathcal{T}_{\mathsf{red}}$, and

(i) for all $i, 0 \leq i \leq n$ and for all relevant assertions ψ: $\mathcal{I}_i \models \psi$ iff $\mathcal{J} \models \psi^{(i)}$;

(ii) for all $i, 0 \leq i \leq n$ and all relevant concept descriptions C, we have $C^{\mathcal{I}_i} = (T_C^{(i)})^{\mathcal{J}}$.

(b) For all interpretations $\mathcal{J}$ such that $\mathcal{J} \models \mathcal{A}_{\mathsf{red}}$ and $\mathcal{J} \models \mathcal{T}_{\mathsf{red}}$, there exist interpretations $\mathcal{I}_0, \ldots, \mathcal{I}_n$ such that $\mathcal{I}_i \Longrightarrow_{\beta_i}^{\emptyset,\emptyset} \mathcal{I}_{i+1}$ for every $i, 0 \leq i < n$, and (i) and (ii) of (a) hold.

Now, we can come back to the consistency problem for actions. Let $\alpha = (\mathsf{pre}, \mathsf{post})$ be an action, CR a finite set of causal relationships, and $\mathcal{T}$ a TBox. The set $\mathcal{R}$ of relevant role names and concept descriptions consists of the ones occurring in α, CR, or $\mathcal{T}$. Given a diagram $\mathcal{D} \in \mathfrak{D}(\alpha, \mathsf{CR})$, we can compute the set $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$, and check whether this set is non-contradictory. If this is the case, then we consider the action $\beta_{\alpha,\mathsf{CR},\mathcal{D}}$, and test whether an application of this action transforms models of $\mathcal{T}$ satisfying pre and $\mathcal{D}$ into models of $\mathcal{T}$. This test can be realized using the ABox $\mathcal{A}_{\mathsf{red}}$ and the (acyclic) TBox $\mathcal{T}_{\mathsf{red}}$ of Lemma 14.

Lemma 15. *The action α is consistent w. r. t. $\mathcal{T}$ and* CR iff *the following holds for all $\mathcal{D} \in \mathfrak{D}(\alpha, \mathsf{CR})$: if $\mathcal{D} \cup \mathsf{pre}$ is consistent w. r. t. $\mathcal{T}$, then*

- *$\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$ is non-contradictory, and*
- *$\mathcal{A}_{\mathsf{red}} \cup \mathcal{T}_{\mathsf{red}} \cup \mathcal{D}^{(0)} \cup \mathsf{pre}^{(0)} \cup \mathcal{T}^{(0)} \models \mathcal{T}^{(1)}$, where $\mathcal{A}_{\mathsf{red}}$ and $\mathcal{T}_{\mathsf{red}}$ are constructed using $\beta_{\alpha,\mathsf{CR},\mathcal{D}}$ and $\mathcal{R}$.*

Proof. ($\Rightarrow$) Let $\mathcal{D}$ be a diagram for α and CR such that $\mathcal{D} \cup \mathsf{pre}$ is consistent with $\mathcal{T}$. Then, there exists an interpretation $\mathcal{I}$ such that $\mathcal{I} \models \mathsf{pre}$, $\mathcal{I} \models \mathcal{D}$, and $\mathcal{I} \models \mathcal{T}$. Since α is consistent w. r. t. $\mathcal{T}$ and CR, there exists an interpretation $\mathcal{I}'$ such that $\mathcal{I} \Longrightarrow_{\alpha}^{\mathcal{T},\mathsf{CR}} \mathcal{I}'$. Thus, we know that $\mathcal{I} \models \mathcal{T}$, $\mathcal{I}' \models \mathcal{T}$, and $\mathcal{I} \Longrightarrow_{\alpha}^{\emptyset,\mathsf{CR}} \mathcal{I}'$. By Lemma 13, we have $\mathcal{I} \Longrightarrow_{\beta_{\alpha,\mathsf{CR},\mathcal{D}}}^{\emptyset,\emptyset} \mathcal{I}'$. Thus, it is clear that $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$ is non-contradictory.

Let $\mathcal{J}$ be a model of $\mathcal{A}_{\mathsf{red}}$, $\mathcal{D}^{(0)}$, $\mathsf{pre}^{(0)}$, $\mathcal{T}_{\mathsf{red}}$, and $\mathcal{T}^{(0)}$. We need to show that $\mathcal{J} \models \mathcal{T}^{(1)}$. By (b) of Lemma 14, there exist interpretations $\mathcal{I}_0$ and $\mathcal{I}_1$ such that $\mathcal{I}_0 \Longrightarrow_{\beta_{\alpha,\mathsf{CR},\mathcal{D}}}^{\emptyset,\emptyset} \mathcal{I}_1$. By (bi) of Lemma 14, $\mathcal{I}_0 \models \mathcal{D} \cup \mathsf{pre}$. By (bii) of Lemma 14, $\mathcal{I}_0 \models \mathcal{T}$. By Lemma 13, $\mathcal{I}_0 \Longrightarrow_{\beta_{\alpha,\mathsf{CR},\mathcal{D}}}^{\emptyset,\emptyset} \mathcal{I}_1$ implies $\mathcal{I}_0 \Longrightarrow_{\alpha}^{\emptyset,\mathsf{CR}} \mathcal{I}_1$. Assume that $\mathcal{I}_1 \not\models \mathcal{T}$. Then $\mathcal{I}_0$ has no successors w. r. t. $\Longrightarrow_{\alpha}^{\mathcal{T},\mathsf{CR}}$, which contradicts α's consistency w. r. t. $\mathcal{T}$ and CR.[7] Thus, we have $\mathcal{I}_1 \models \mathcal{T}$, which together with (bii) of Lemma 14, yields $\mathcal{J} \models \mathcal{T}^{(1)}$.

[7]This would not be true if we had allowed for occ since $\mathcal{I}_0$ can have more than more successors w. r. t. $\Longrightarrow_{\beta_{\alpha,\mathsf{CR},\mathcal{D}}}$. See Section 6 for the definition of the semantics of actions with occlusions.

($\Leftarrow$) Let $\mathcal{I}$ be an interpretation with $\mathcal{I} \models \mathsf{pre}$ and $\mathcal{I} \models \mathcal{T}$. We define $\mathcal{D}$ as in Lemma 8. Thus, we know that $\mathcal{I} \models \mathcal{D}$ and that $\mathcal{D}$ is a diagram for α and CR. Since $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$ is non-contradictory, there exists an interpretation $\mathcal{I}'$ such that $\mathcal{I} \Longrightarrow^{\emptyset,\emptyset}_{\beta_{\alpha,\mathsf{CR},\mathcal{D}}} \mathcal{I}'$. Thus, by Lemma 13, we have $\mathcal{I} \Longrightarrow^{\emptyset,\mathsf{CR}}_{\alpha} \mathcal{I}'$. Moreover, by Lemma 14 (a), there exists an interpretation $\mathcal{J}$ such that $\mathcal{J} \models \mathcal{A}_{\mathsf{red}}$, $\mathcal{J} \models \mathcal{T}_{\mathsf{red}}$, $\mathcal{J} \models \mathcal{D}^{(0)} \cup \mathsf{pre}^{(0)}$, and $\mathcal{J} \models \mathcal{T}^{(0)}$. Hence, $\mathcal{J} \models \mathcal{T}^{(1)}$. By Lemma 14 (aii), $\mathcal{I}' \models \mathcal{T}$, which, together with the facts that $\mathcal{I} \Longrightarrow^{\emptyset,\mathsf{CR}}_{\alpha} \mathcal{I}'$ and that $\mathcal{I} \models \mathcal{T}$, yields that $\mathcal{I}_0 \Longrightarrow^{\mathcal{T},\mathsf{CR}}_{\alpha} \mathcal{I}'$. ❏

This lemma shows that consistency of an action w. r. t. a TBox and a finite set of causal relationships can be tested by considering the exponentially many elements of $\mathfrak{D}(\alpha, \mathsf{CR})$. For each element $\mathcal{D} \in \mathfrak{D}(\alpha, \mathsf{CR})$, we

- check the consistency of $\mathcal{D} \cup \mathsf{pre}$ w. r. t. $\mathcal{T}$;
- construct the set $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$ as well as $\beta_{\alpha,\mathsf{CR},\mathcal{D}}$ and $\mathcal{R}$ (in PTime) and check whether $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \mathsf{CR})$ is contradictory or not (in PTime), and
- construct $\mathcal{A}_{\mathsf{red}}$ and $\mathcal{T}_{\mathsf{red}}$ using $\beta_{\alpha,\mathsf{CR},\mathcal{D}}$ and $\mathcal{R}$ (in PTime) and construct $\mathcal{D}^{(0)}$, $\mathsf{pre}^{(0)}$, $\mathcal{T}^{(0)}$, and $\mathcal{T}^{(1)}$ (in PTime) and check whether $\mathcal{A}_{\mathsf{red}} \cup \mathcal{T}_{\mathsf{red}} \cup \mathcal{D}^{(0)} \cup \mathsf{pre}^{(0)} \cup \mathcal{T}^{(0)} \models \mathcal{T}^{(1)}$, which can be reduced to inconsistency of ABoxes w. r. t. general TBoxes.

For $\mathcal{ALCIO}$ and $\mathcal{ALCQO}$, ABox consistency w. r. t. general TBoxes can be decided in ExpTime [8, 9], and thus, an action's consistency w. r. t. a general TBox and a set of causal relationships in $\mathcal{ALCIO}$ and $\mathcal{ALCQO}$ can be decided in ExpTime. For $\mathcal{ALCQIO}$, ABox consistency w. r. t. general TBoxes can be decided in NExpTime [18, 12], so an action's *inconsistency* can be decided in $\text{PTime}^{\text{NExpTime}}$. This is due to the fact that the above procedure employs a NExpTime check and a co-NExpTime check, respectively, using a guessed diagram, which is clearly in $\text{NP}^{\text{NExpTime}}$, i. e. this can be done in NP using a NExpTime oracle. Since the strong exponential hierachy collapses [6] and in particular we have that $\text{PTime}^{\text{NExpTime}} = \text{NP}^{\text{NExpTime}}$, it follows that an action's inconsistency can be decided in $\text{PTime}^{\text{NExpTime}}$ for $\mathcal{ALCQIO}$. An action's *consistency* can be decided in $\text{PTime}^{\text{NExpTime}}$ for $\mathcal{ALCQIO}$ since $\text{PTime}^{\text{NExpTime}}$ is a deterministic complexity class.

Let $\mathcal{A}$ be an ABox and $\mathcal{T}$ a general TBox. Then, $\mathcal{A}$ is inconsistent w. r. t. $\mathcal{T}$ iff $(\mathcal{A}, \{A(a), \neg A(a)\})$ is consistent w. r. t. $\mathcal{T}$ and $\emptyset$, where A is an arbitrary concept name and a is an arbitrary individual name. Since ABox consistency w. r. t. a general TBox is ExpTime-hard [14], it follows that an action's consistency is ExpTime-hard for $\mathcal{ALC}$.

Theorem 16. *The problem of deciding consistency of an action w. r. t. a TBox and a finite set of causal relationships is*

- ExpTime*-complete for the DLs* $\mathcal{ALC}$, $\mathcal{ALCO}$, $\mathcal{ALCQ}$, $\mathcal{ALCI}$, $\mathcal{ALCQO}$, *and* $\mathcal{ALCIO}$*;*
- *in* $\text{PTime}^{\text{NExpTime}}$ *for* $\mathcal{ALCQI}$ *and* $\mathcal{ALCQIO}$*.*

It is still open whether the upper bounds for $\mathcal{ALCQI}$ and $\mathcal{ALCQIO}$ are optimal.

5 Deciding Projection

The projection problem considers a sequence of actions $\alpha_1, \ldots, \alpha_n$, together with a TBox $\mathcal{T}$, a finite set of causal relationships CR, an initial ABox $\mathcal{A}$, and an assertion φ. By definition, φ is a consequence of applying $\alpha_1, \ldots, \alpha_n$ to $\mathcal{A}$ w. r. t. $\mathcal{T}$ and CR iff, for all interpretations $\mathcal{I}_0, \ldots, \mathcal{I}_{n-1}, \mathcal{I}_n$, if $\mathcal{I}_0 \models \mathcal{A}$ and $\mathcal{I}_0 \Longrightarrow_{\alpha_1}^{\mathcal{T},\mathsf{CR}} \mathcal{I}_1 \Longrightarrow_{\alpha_2}^{\mathcal{T},\mathsf{CR}} \cdots \mathcal{I}_{n-1} \Longrightarrow_{\alpha_n}^{\mathcal{T},\mathsf{CR}} \mathcal{I}_n$, then $\mathcal{I}_n \models \varphi$.

Our solution of the projection problem w. r. t. $\mathcal{T}$ and CR uses the same ideas as the solution of the consistency sketched in Section 4. *First*, instead of considering interpretations $\mathcal{I}_0, \ldots, \mathcal{I}_{n-1}$, we consider diagrams $\mathcal{D}_0, \ldots, \mathcal{D}_{n-1}$, where $\mathcal{D}_i \in \mathfrak{D}(\alpha_{i+1}, \mathsf{CR})$ for $i = 0, \ldots, n-1$.[8] *Second*, we use the original sequence of actions $\alpha_1, \ldots, \alpha_n$ and the diagrams $\mathcal{D}_0, \ldots, \mathcal{D}_{n-1}$ to build the corresponding sequence of actions $\beta_{\alpha_1,\mathsf{CR},\mathcal{D}_0}, \ldots, \beta_{\alpha_n,\mathsf{CR},\mathcal{D}_{n-1}}$. Lemma 13 then tells us that, for all models $\mathcal{I}_{i-1}$ of $\mathcal{D}_{i-1}$ and all interpretations $\mathcal{I}_i$ we have $\mathcal{I}_{i-1} \Longrightarrow_{\alpha_i}^{\emptyset,\mathsf{CR}} \mathcal{I}_i$ iff $\mathcal{I}_{i-1} \Longrightarrow_{\beta_{\alpha_i,\mathsf{CR},\mathcal{D}_{i-1}}}^{\emptyset,\emptyset} \mathcal{I}_i$. *Third*, we use the sequence $\beta_{\alpha_1,\mathsf{CR},\mathcal{D}_0}, \ldots, \beta_{\alpha_n,\mathsf{CR},\mathcal{D}_{n-1}}$ and the set of relevant role names and concept descriptions $\mathcal{R}$ to construct an ABox $\mathcal{A}_{\mathsf{red}}$ and (acyclic) a TBox $\mathcal{T}_{\mathsf{red}}$ such that the properties (a) and (b) of Lemma 14 hold. In this setting, the set $\mathcal{R}$ consists of the role names and concept descriptions occurring in $\mathcal{A}$, $\mathcal{T}$, $\alpha_1, \ldots, \alpha_n$, CR, and φ. These properties can be used to express that the initial interpretation $\mathcal{I}_0$ must be a model of $\mathcal{A}$ and that we only consider interpretations $\mathcal{I}_i$ that are models of $\mathcal{T}$. In addition, we can then check, whether all this implies that the final interpretation $\mathcal{I}_n$ is a model of φ. To be more precise, we can show that the following characterization of the projection problem holds:

Lemma 17. *Let* $\alpha_1, \ldots, \alpha_n$ *be a sequence of actions,* $\mathcal{T}$ *a TBox,* CR *a finite set of causal relationships,* $\mathcal{A}$ *an initial ABox, and* φ *an assertion. Then,* φ *is a consequence of applying* $\alpha_1, \ldots, \alpha_n$ *to* $\mathcal{A}$ *w. r. t.* $\mathcal{T}$ *and* CR *iff for all diagrams* $\mathcal{D}_0, \ldots, \mathcal{D}_{n-1}$ *such that* $\mathcal{D}_i \in \mathfrak{D}(\alpha_{i+1}, \mathsf{CR})$ *for* $i = 0, \ldots, n-1$*, we have*

$$\bigcup_{i=0}^{n-1} \mathcal{D}_i^{(i)} \cup \bigcup_{i=0}^{n} \mathcal{T}^{(i)} \cup \mathcal{A}^{(0)} \cup \mathcal{A}_{\mathsf{red}} \cup \mathcal{T}_{\mathsf{red}} \models \varphi^{(n)}, \qquad (2)$$

where $\mathcal{A}_{\mathsf{red}}$ *and* $\mathcal{T}_{\mathsf{red}}$ *are constructed from* $\beta_{\alpha_1,\mathsf{CR},\mathcal{D}_0}, \ldots, \beta_{\alpha_n,\mathsf{CR},\mathcal{D}_{n-1}}$ *and* $\mathcal{R}$*.*

[8]Note that it is enough to consider diagrams $\mathcal{D}_0, \ldots, \mathcal{D}_{n-1}$ for $\mathcal{I}_0, \ldots, \mathcal{I}_{n-1}$ since no action is applied to $\mathcal{I}_n$.

Proof. "$\Rightarrow$": Consider diagrams $\mathcal{D}_0, \ldots, \mathcal{D}_{n-1}$ such that $\mathcal{D}_i \in \mathfrak{D}(\alpha_{i+1}, \mathsf{CR})$ for $i = 0, \ldots, n-1$. Let $\mathcal{J}$ be a model of $\bigcup_{i=0}^{n-1} \mathcal{D}^{(i)} \cup \bigcup_{i=0}^{n} \mathcal{T}^{(i)} \cup \mathcal{A}^{(0)} \cup \mathcal{A}_{\mathsf{red}} \cup \mathcal{T}_{\mathsf{red}}$. By Lemma 14 (b), there are interpretations $\mathcal{I}_0, \ldots, \mathcal{I}_n$ such that $\mathcal{I}_i \Longrightarrow^{\emptyset,\emptyset}_{\beta_{\alpha_{i+1},\mathsf{CR},\mathcal{D}_i}} \mathcal{I}_{i+1}$ for every i with $0 \leq i < n$. By Lemma 14 (bi), we have that $\mathcal{I}_i \models \mathcal{D}_i$ for all i with $0 \leq i < n$ and $\mathcal{I}_0 \models \mathcal{A}$. By Lemma 13, we get $\mathcal{I}_i \Longrightarrow^{\emptyset,\mathsf{CR}}_{\alpha_{i+1}} \mathcal{I}_{i+1}$. Lemma 14 (bii) yields that $\mathcal{I}_i \models \mathcal{T}$ for all $i \leq n$. Thus, $\mathcal{I}_i \Longrightarrow^{\mathcal{T},\mathsf{CR}}_{\alpha_{i+1}} \mathcal{I}_{i+1}$. Since φ is a consequence of applying $\alpha_1, \ldots, \alpha_n$ to $\mathcal{A}$ w. r. t. $\mathcal{T}$ and CR, we know that $\mathcal{I}_n \models \varphi$, which implies by Lemma 14 (bi) that $\mathcal{J} \models \varphi^{(n)}$.

"$\Leftarrow$": Let $\mathcal{I}_0, \ldots, \mathcal{I}_n$ be interpretations such that $\mathcal{I}_0 \models \mathcal{A}$ and $\mathcal{I}_i \Longrightarrow^{\mathcal{T},\mathsf{CR}}_{\alpha_{i+1}} \mathcal{I}_{i+1}$ for all i with $0 \leq i < n$. Then, we know that

- $\mathcal{I}_i \Longrightarrow^{\emptyset,\mathsf{CR}}_{\alpha_{i+1}} \mathcal{I}_{i+1}$ for all $i < n$, and
- $\mathcal{I}_i \models \mathcal{T}$ for all $i \leq n$.

We define $\mathcal{D}_i$ with $\mathcal{A}(\alpha_{i+1}, \mathsf{CR})$ for all i with $i < n$ as in Lemma 8. Thus, we know that $\mathcal{I}_i \models \mathcal{D}_i$ and that $\mathcal{D}_i$ is a diagram for α_{i+1} and CR. Then, by Lemma 13, $\mathcal{I}_i \Longrightarrow^{\emptyset,\emptyset}_{\beta_{\alpha_{i+1},\mathsf{CR},\mathcal{D}_i}} \mathcal{I}_{i+1}$. Thus, by Lemma 14 (a), there exists an interpretation $\mathcal{J}$ such that $\mathcal{J} \models \mathcal{A}_{\mathsf{red}}$ and $\mathcal{J} \models \mathcal{T}_{\mathsf{red}}$. Lemma 14 (ai) yields that $\mathcal{J} \models \mathcal{D}_i^{(i)}$ for all $i < n$ and $\mathcal{J} \models \mathcal{A}^{(0)}$ and (aii) yields that $\mathcal{J} \models \mathcal{T}^{(i)}$ for all $i \leq n$. Thus, we have $\mathcal{J} \models \varphi^{(n)}$, which implies by Lemma 14 (ai) that $\mathcal{I}_n \models \varphi$. ❑

It is easy to see that this lemma yields a decision procedure for the projection problem. If the ABox consistency problem w. r. t. a general TBox is ExpTime-complete such as $\mathcal{ALCIO}$ [8] and $\mathcal{ALCQO}$ [9], one needs to consider exponentially many sequences of diagrams $\mathcal{D}_0, \ldots, \mathcal{D}_{n-1}$. For each such sequence, the actions $\beta_{\alpha_1,\mathsf{CR},\mathcal{D}_0}, \ldots, \beta_{\alpha_n,\mathsf{CR},\mathcal{D}_{n-1}}$, and thus also $\mathcal{A}_{\mathsf{red}}$ and $\mathcal{T}_{\mathsf{red}}$, can be constructed in polynomial time. Thus, the inference problem (2) is of polynomial size, and it can be solved in exponential time. If the underlying DL is $\mathcal{ALCQIO}$, then the ABox consistency problem w. r. t. a general TBox is in NExpTime [18, 12]. Then, the action's *inconsistency* can be decided by guessing a sequence of diagrams $\mathcal{D}_0, \ldots, \mathcal{D}_{n-1}$, constructing the actions $\beta_{\alpha_1,\mathsf{CR},\mathcal{D}_0}, \ldots, \beta_{\alpha_n,\mathsf{CR},\mathcal{D}_{n-1}}$ and thus also $\mathcal{A}_{\mathsf{red}}$ and $\mathcal{T}_{\mathsf{red}}$, and checking whether we have (2). Overall, it yields a NExpTime procedure.

Now, we consider lower bounds: C is satisfiable w. r. t. $\mathcal{T}$ iff $\neg C(a)$ is not a consequence of applying $(\emptyset, \emptyset)$ to $\emptyset$ w. r. t. $\mathcal{T}$ and CR, where a is an individual name which does not occur in C or $\mathcal{T}$. Satisfiability of a concept w. r. t. a general TBox is ExpTime-complete [2]. Thus, the projection problem is ExpTime-hard for $\mathcal{ALC}$. The projection problem for $\mathcal{ALCQI}$ is co-NExpTime-hard even if $\mathcal{T}$ is empty [3]. Thus, we obtain the following theorem:

Theorem 18. *The projection problem w. r. t. a TBox and a finite set of causal relationships*

- ExpTime-*complete for* $\mathcal{ALC}$, $\mathcal{ALCO}$, $\mathcal{ALCI}$, $\mathcal{ALCIO}$, $\mathcal{ALCQ}$, *and* $\mathcal{ALCQO}$.
- *co*-NExpTime-*complete for* $\mathcal{ALCQI}$ *and* $\mathcal{ALCQIO}$.

For the special case of an empty TBox, we observe that in the construction of $\bigcup_{i=0}^{n-1} \mathcal{D}^{(i)} \cup \mathcal{A}^{(0)} \cup \mathcal{A}_{\mathsf{red}} \cup \mathcal{T}_{\mathsf{red}}$ we do *not* introduce GCIs explicitly (since $\mathcal{T}_{\mathsf{red}}$ is an acyclic TBox [3]). Thus, checking whether $\bigcup_{i=0}^{n-1} \mathcal{D}^{(i)} \cup \mathcal{A}^{(0)} \cup \mathcal{A}_{\mathsf{red}} \cup \mathcal{T}_{\mathsf{red}} \models \varphi^{(n)}$ can be done in PSpace for $\mathcal{ALCQO}$ [4] (unary coding for qualified number restrictions), ExpTime for $\mathcal{ALCIO}$ [1], and co-NExpTime for $\mathcal{ALCQIO}$ [18, 12] (even coding for qualified number restrictions). Moreover, the projection problem [3] is PSpace-hard for $\mathcal{ALC}$, ExpTime-hard for $\mathcal{ALCI}$, and co-NExpTime-hard for $\mathcal{ALCQI}$ even if $\mathcal{T}$ is empty. Thus, we obtain the following corollary:

Corollary 19. *The projection problem w. r. t. the empty TBox and a finite set of causal relationships is*

- PSpace-*complete for* $\mathcal{ALC}$, $\mathcal{ALCO}$, $\mathcal{ALCQ}$, *and* $\mathcal{ALCQO}$ *if the number restrictions are coded in unary.*
- ExpTime-*complete for* $\mathcal{ALCI}$ *and* $\mathcal{ALCIO}$.
- *co*-NExpTime-*complete for* $\mathcal{ALCQI}$ *and* $\mathcal{ALCQIO}$.

6 Additional Results and Future Work

In this paper, we have proposed to use causal relationships to deal with the ramification problem for DL-based action formalisms. We focused on deciding the consistency problem of an action and the projection problem in the setting with and without domain knowledge, which is described with a general TBox, for DLs considered in [3]. What differs from DL to DL is the complexity of the basic inference problems in the respective DL (extended with nominals). Except for two cases, we get the matching hardness results by a reduction from such a basic inference problem. The complexity results obtained in this paper are listed in Table 2.

Regarding future work, one interesting question is whether our approach for deciding the consistency and the projection problem can be extended to actions with occlusions [3]. Let $\alpha = (\mathsf{pre}, \mathsf{occ}, \mathsf{post})$ be an action with a finite set occ of occlusions of the form $A(a)$ or $r(a, b)$ where A is a concept name, r is a role name, and a, b are individual names. Let CR be a finite set of causal relationships, and

	TBox?	$\mathcal{ALC}$	$\mathcal{ALCO}$	$\mathcal{ALCQ}$	$\mathcal{ALCI}$	$\mathcal{ALCQO}$	$\mathcal{ALCQI}$	$\mathcal{ALCIO}$	$\mathcal{ALCQIO}$
Consistency	$\mathcal{T} = \emptyset$	PSp	PSp	PSp	PSp	PSp	PSp	Exp	cNE
	$\mathcal{T} \neq \emptyset$	Exp	Exp	Exp	Exp	Exp	PNE	Exp	PNE
Projection	$\mathcal{T} = \emptyset$	PSp	PSp	PSp	Exp	PSp	cNE	Exp	cNE
	$\mathcal{T} \neq \emptyset$	Exp	Exp	Exp	Exp	Exp	cNE	Exp	cNE

Table 2: The complexity of the consistency and the projection problem. PSp is short for PSpace-complete, Exp for ExpTime-complete, cNE for co-NExpTime-complete, PNE for *in* $\text{PTime}^{\text{NExpTime}}$.

$\mathcal{I}$ an interpretation. For each concept name A and role name r, set:

$$\begin{array}{rcl}
A^+ & := & \{b^\mathcal{I} \mid A(b) \in \mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})\} \\
A^- & := & \{b^\mathcal{I} \mid \neg A(b) \in \mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})\} \\
I_A & := & (\Delta^\mathcal{I} \setminus \{b^\mathcal{I} \mid A(b) \in \mathsf{occ}\}) \cup (A^+ \cup A^-) \\
r^+ & := & \{(a^\mathcal{I}, b^\mathcal{I}) \mid r(a,b) \in \mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})\} \\
r^- & := & \{(a^\mathcal{I}, b^\mathcal{I}) \mid \neg r(a,b) \in \mathcal{E}(\alpha, \mathcal{I}, \mathsf{CR})\} \\
I_r & := & ((\Delta^\mathcal{I} \times \Delta^\mathcal{I}) \setminus \{(a^\mathcal{I}, b^\mathcal{I}) \mid r(a,b) \in \mathsf{occ}\}) \cup (r^+ \cup r^-)
\end{array}$$

In order to give the semantics of actions with occlusions, we then revise the last two conditions in Definition 4 as follows:

- for all concept names A, we have $A^{\mathcal{I}'} \cap I_A = ((A^\mathcal{I} \cup A^+) \setminus A^-) \cap I_A$, and
- for all role names r, we have $r^{\mathcal{I}'} \cap I_r = ((r^\mathcal{I} \cup r^+) \setminus r^-) \cap I_r$.

Note that such actions are non-deterministic, i. e., their application to an interpretation may yield several possible successor interpretations. Consequently, such an action may still be consistent although some of the successors interpretations are not models of the TBox (see the proof of Lemma 15). Thus, consistency can no longer be characterized by an analog of Lemma 15.

When defining our semantics for actions in the presence of causal relationships, we followed the approach used in [20, 5] rather than the one employed by [10, 16]. In our health insurance example, this was actually the appropriate semantics, but there may also be examples where it would be better to use the other semantics. Thus, it would be interesting to see whether our approach for deciding the consistency and the projection problem can be adapted to deal with the semantics of [10, 16].

References

[1] Carlos Areces, Patrick Blackburn, and Maarten Marx. A road-map on complexity for hybrid logics. In Jörg Flum and Mario Rodríguez-Artalejo, editors, *Computer Science Logic, 13th International Workshop, CSL '99, 8th Annual Conference of the EACSL, Proceedings*, volume 1683 of *Lecture Notes in Computer Science*, pages 307–321, Madrid, Spain, September 1999. Springer-Verlag.

[2] 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, 2003.

[3] Franz Baader, Carsten Lutz, Maja Miličić, Ulrike Sattler, and Frank Wolter. Integrating description logics and action formalisms: First results. In Manuela M. Veloso and Subbarao Kambhampati, editors, *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05) and the Seventeenth Innovative Applications of Artificial Intelligence Conference (IAAI-05)*, pages 572–577, Pittsburgh, Pennsylvania, USA, July 2005. AAAI Press/The MIT Press.

[4] Franz Baader, Maja Miličić, Carsten Lutz, Ulrike Sattler, and Frank Wolter. Integrating description logics and action formalisms for reasoning about web services. LTCS-Report 05-02, Chair of Automata Theory, Institute of Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany, 2005. See http://lat.inf.tu-dresden.de/research/reports.html.

[5] Marc Denecker, Daniele Theseider-Dupré, and Kristof van Belleghem. An inductive definition approach to ramifications. *Linkoping Electronic Articles in Computer and Information Science*, 3(7):1–43, January 1998.

[6] Lane Adrian Hemachandra. The strong exponential hierarchy collapses. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing (STOC-1987)*, pages 110–122, New York City, New York, USA, May 1987. ACM.

[7] Andreas Herzig. The PMA revisited. In *Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning (KR-1996)*, pages 40–50, 1996.

[8] Jan Hladik. A tableau system for the description logic $\mathcal{SHIO}$. In Ulrike Sattler, editor, *Contributions to the Doctoral Programme of the 2nd International Joint Conference on Automated Reasoning (IJCAR-04)*, volume 106, Cork, Ireland, July 2004. CEUR-WS.org.

[9] Ian Horrocks and Ulrike Sattler. Ontology reasoning in the $\mathcal{SHOQ}$(D) description logic. In Bernhard Nebel, editor, *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-2001)*, pages 199–204, Seattle, Washington, USA, August 2001. Morgan Kaufmann, Los Altos.

[10] Fangzhen Lin. Embracing causality in specifying the indirect effects of actions. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, pages 1985–1993, 1995.

[11] Hongkai Liu, Carsten Lutz, Maja Miličić, and Frank Wolter. Reasoning about actions using description logics with general TBoxes. In Michael Fisher, Wiebe van der Hoek, Boris Konev, and Alexei Lisitsa, editors, *Proceedings of the 10th European Conference on Logics in Artificial Intelligence (JELIA-2006)*, volume 4160 of *Lecture Notes in Artifical Intelligence*, pages 266–279. Springer-Verlag, 2006.

[12] Ian Pratt-Hartmann. Complexity of the two-variable fragment with counting quantifiers. *Journal of Logic, Language and Information*, 14(3):369–395, 2005.

[13] Raymond Reiter. *Knowledge in Action: Logical Foundations for Describing and Implementing Dynamical Systems.* The MIT Press, Bradford Books, Cambridge, Massachusetts, USA, 2001.

[14] Klaus Schild. A correspondence theory for terminological logics: Preliminary report. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI-1991)*, pages 466–471, Sidney, Australia, 1991.

[15] Manfred Schmidt-Schauß and Gert Smolka. Attributive concept descriptions with complements. *Artificial Intelligence Journal*, 48(1):1–26, 1991.

[16] Michael Thielscher. Ramification and causality. *Artificial Intelligence Journal*, 89(1–2):317–364, 1997.

[17] Michael Thielscher. *Reasoning Robots: The Art and Science of Programming Robotic Agents*, volume 33 of *Applied Logic Series.* Springer Netherlands, 2005.

[18] Stephan Tobies. The complexity of reasoning with cardinality restrictions and nominals in expressive description logics. *Journal of Artificial Intelligence Research*, 12:199–217, 2000.

[19] Stephan Tobies. *Complexity results and practical algorithms for logics in Knowledge Representation.* PhD thesis, RWTH Aachen, 2001.

[20] Kristof van Belleghem, Marc Denecker, and Daniele Theseider-Dupré. A constructive approach to the ramification problem. In Guiseppe De Giacomo and Daniele Nardi, editors, *Proceedings of the ESSLLI-98 Workshop on Reasoning about Actions: Foundations and Applications*, pages 1–17, Saarbrücken, Germany, 1998.

[21] Marianne Winslett. Reasoning about action using a possible models approach. In *Proceedings of the Seventh National Conference on Artificial Intelligence (AAAI-1988)*, pages 89–93, 1988.