

# Using Causal Relationships to Deal with the Ramification Problem in Action Formalisms Based on Description Logics

Franz Baader, Marcel Lippmann, and Hongkai Liu\*

Theoretical Computer Science, TU Dresden  
*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 (DLs) without destroying the decidability of the consistency and the projection problem. We investigate the complexity of these decision problems based on which DL is used as base logic for the action formalism.

## 1 Introduction

For action theories represented in the situation or fluent calculus [13, 16], important inference problems such as the projection problem 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 [1]. 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 (possibly incomplete) 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. Given a finite sequence of actions  $\alpha_1 \dots \alpha_n$ , an incomplete description  $\mathcal{A}_0$  of the initial state,

---

\* supported by DFG under grant BA 1122/13-1.

and a formula  $\varphi$  specifying a (desired or unwanted) property of states, *projection* [13] is the inference problem that asks whether  $\varphi$  holds in all states that can be reached from a possible initial state (i.e., a state satisfying  $\mathcal{A}_0$ ) by applying this sequence of actions. The formula  $\varphi$  may, for example, be the prerequisite of an action  $\alpha$  to be applied after the last action of the sequence, or a condition used in the control structure of an agent’s program. In [3], it was shown that the projection problem is decidable in action theories based on DLs between  $\mathcal{ALC}$  and  $\mathcal{ALCQIO}$ . However, this paper did not deal with the so-called ramification problem [8, 15].

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) [18, 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 [9], 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 [9] that important inference problems such as the projection problem stay decidable in this setting, but that the consistency<sup>1</sup> 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 [8, 15, 17, 6]. 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. The semantics we define for such theories is an adaptation of the one introduced in [17, 6] in a more general setting, and it inherits the advantages and disadvantages of this approach. The main thrust of this work is not to invent a new solution of the ramification problem and discuss its appropriateness, but to show that adding a well-accepted existing solution from the reasoning about actions community [17, 6] to DL-based action theories leaves important inference problems such as the consistency problem and the projection problem decidable. More precisely, we provide not only decidability results, but detailed results on the complexity of these two problems depending on which DL is used as base logic. With a few exceptions, these results show that adding causal relationships to DL-based action formalisms does not increase the complexity of these inference problems.

Using causal relationships has two advantages over the formalism for handling the ramification problem introduced in [9]. First, the formalism in [9] 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,

---

<sup>1</sup> In [9], this is actually called strong consistency.

**Table 1.** Syntax and semantics of  $\mathcal{ALCO}$ 

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	$\{a\}$	$\{a^{\mathcal{I}}\}$
value restriction	$\forall r.C$	$\{x \mid \forall y. ((x, y) \in r^{\mathcal{I}} \rightarrow y \in C^{\mathcal{I}})\}$
existential restriction	$\exists r.C$	$\{x \mid \exists y. ((x, y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}})\}$
general concept inclusion	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
concept assertion	$C(a)$	$a^{\mathcal{I}} \in C^{\mathcal{I}}$
role assertion	$r(a, b)$	$(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$
negated role assertion	$\neg r(a, b)$	$(a^{\mathcal{I}}, b^{\mathcal{I}}) \notin r^{\mathcal{I}}$

and thus is not well-defined. If (s)he adds the causal relationship mentioned above, then the action becomes consistent.

Due to space constraints, we cannot give complete proofs of all our results here. They can be found in the accompanying technical report [2].

## 2 DL-based Action Formalisms and Causal Relationships

We assume that the reader is familiar with the basic notions from Description Logics, which can, for example, be found in [1]. In principle, our action formalism can be parameterized with any DL. In this paper, we restrict the detailed presentation to the DL  $\mathcal{ALCO}$ , and only list the results that we have obtained for other DLs in Section 5. The DL  $\mathcal{ALCO}$  extends the smallest propositionally closed DL  $\mathcal{ALC}$  with so-called nominals. In  $\mathcal{ALC}$ , one can build complex concept descriptions from atomic concepts (concept names) using the Boolean constructors ( $\sqcap$ ,  $\sqcup$ ,  $\neg$ ) as well as value restrictions ( $\forall r.C$ ) and existential restrictions ( $\exists r.C$ ), where  $r$  is a role name. In  $\mathcal{ALCO}$ , one can additionally use individual names  $a$  to build nominal concepts  $\{a\}$ , which are interpreted as singleton sets.

An *ABox* is a finite set of *concept assertions*  $C(a)$  and *role assertions*  $r(a, b)$ , and *negated role assertions*  $\neg r(a, b)$ , where  $C$  is a concept description,  $r$  is a role name, and  $a, b$  are individual names. An ABox is *simple* if all its concept assertions are of the form  $A(a)$  or  $\neg A(a)$ , where  $A$  is a concept name. We will call the concept and (negated) 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  $\neg L$  is  $\neg L$  if  $L$  is positive, and it is  $L'$  if  $L = \neg L'$  is negative. A *TBox* is a finite set of general concept inclusions (GCIs) of the form  $C \sqsubseteq D$ , where  $C, D$  are concept descriptions.

*Example 2.1.* 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:

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

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

The semantics of  $\mathcal{ALCO}$  is defined in terms of an *interpretation*  $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ , where  $\Delta^{\mathcal{I}}$ , the *domain*, is a non-empty set, and  $\cdot^{\mathcal{I}}$ , the *interpretation function*, maps each concept name  $A$  to a subset  $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ , each role name  $r$  to a binary relation  $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ , and each individual name  $a$  to an element  $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ . The extension of  $\cdot^{\mathcal{I}}$  to complex concept descriptions is defined inductively, as shown in the semantics column of Table 1. The interpretation  $\mathcal{I}$  *satisfies* a GCI or assertion  $\varphi$  (written  $\mathcal{I} \models \varphi$ ) if the condition in the semantics column of Table 1 is satisfied. If  $\mathcal{I}$  satisfies all the assertions of the ABox  $\mathcal{A}$  (GCIs of the TBox  $\mathcal{T}$ ), then we say that  $\mathcal{I}$  is a *model* of  $\mathcal{A}$  ( $\mathcal{T}$ ), and write  $\mathcal{I} \models \mathcal{A}$  ( $\mathcal{I} \models \mathcal{T}$ ). 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  $\varphi$  (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  $\varphi$  ( $\mathcal{T}'$ ).

The following definition recalls the notion of a DL action without occlusions, which has first been introduced in [3].<sup>2</sup> 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 2.2.** *An action is a pair  $\alpha = (\text{pre}, \text{post})$ , where **pre** is a finite set of assertions, the pre-conditions, and **post** is a finite set of conditional post-conditions of the form  $\varphi/\psi$ , where  $\varphi$  is an assertion and  $\psi$  is a literal. Such an action is called *unconditional* if all its post-conditions are of the form  $\text{true}/\psi$ , where “true” stands for an assertion that is satisfied in every interpretation. We write such unconditional post-conditions simply as  $\psi$  rather than  $\text{true}/\psi$ .*

Basically, an action is applicable in an interpretation if its pre-conditions are satisfied. The conditional post-condition  $\varphi/\psi$  requires that  $\psi$  must hold after the application of the action if  $\varphi$  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 2.1 and the action  $\text{HireJohn} = (\emptyset, \{\text{Employee}(\text{JOHN})\})$ , which has no pre-conditions and a single unconditional post-condition. Assume that  $\mathcal{I}$  is a model of  $\mathcal{T}$  with  $\mathcal{I} \not\models \text{Employee}(\text{JOHN})$  and  $\mathcal{I} \not\models \exists \text{insuredBy. HealthInsuranceCompany}(\text{JOHN})$  (obviously, such models exist). If we apply the semantics of DL actions introduced

<sup>2</sup> Intuitively, occlusions describe parts of the domain that can change arbitrarily when the action is applied. More details about occlusions can be found in [3] and in Section 7 of [2].

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.,  $\mathcal{I}' \models \text{Employee}(\text{JOHN})$ . Since nothing else changes, we still have  $\mathcal{I}' \not\models \exists \text{insuredBy. HealthInsuranceCompany}(\text{JOHN})$ , which shows that  $\mathcal{I}'$  is not a model of  $\mathcal{T}$ . Consequently, although the action `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 `HireJohn` with an appropriate causal relationship.

**Definition 2.3.** A causal relationship is of the form  $\mathcal{A}_1 \xrightarrow{\mathcal{B}} \mathcal{A}_2$ , 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:<sup>3</sup> if  $\mathcal{B}$  is satisfied *before*<sup>4</sup> 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

$$\{\text{Employee}(\text{JOHN})\} \xrightarrow{\{\neg \exists \text{insuredBy. HealthInsuranceCompany}(\text{JOHN})\}} \{\text{insuredBy}(\text{JOHN}, \text{AOK})\}$$

adds the following indirect effect to the direct effect of the `HireJohn` action: (i) if John becomes newly employed (i.e., was not an employee before) and did not 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 2.1 stay satisfied.

In order to define the semantics of DL actions in the presence of causal relationships formally, we consider an action  $\alpha = (\text{pre}, \text{post})$ , 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  $\alpha$  given  $\mathcal{I}$  is defined as

$$\text{Dir}(\alpha, \mathcal{I}) := \{\psi \mid \varphi/\psi \in \text{post} \wedge \mathcal{I} \models \varphi\}.$$

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.

<sup>3</sup> Actually, there are different ways of defining the meaning of causal relationships. Here, we follow the approach used in [17, 6] rather than the one employed by [8, 15].

<sup>4</sup> In the semantics of causal relationship introduced in [8, 15], this “before” would need to be replaced by “after.”

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

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

Thus, we add the indirect effect  $\psi$  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

$$\mathcal{E}(\alpha, \mathcal{I}, \text{CR}) := \mathcal{E}_n .$$

This set of literals represents the effects of applying the action  $\alpha$  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  $\mathcal{E}(\alpha, \mathcal{I}, \text{CR})$  is *contradictory* if there is a literal  $L$  such that  $\{L, \dot{\neg}L\} \subseteq \mathcal{E}(\alpha, \mathcal{I}, \text{CR})$ .

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

**Definition 2.4.** *Let  $\alpha$  be an action, CR a finite set of causal relationships,  $\mathcal{T}$  a TBox, and  $\mathcal{I}, \mathcal{I}'$  two interpretations. We say that  $\alpha$  may transform  $\mathcal{I}$  to  $\mathcal{I}'$  w.r.t.  $\mathcal{T}$  and CR (denoted by  $\mathcal{I} \Longrightarrow_{\alpha}^{\mathcal{T}, \text{CR}} \mathcal{I}'$ ) 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  $\mathcal{I}' \models \mathcal{T}$ ,
- $\mathcal{E}(\alpha, \mathcal{I}, \text{CR})$  is not contradictory,
- for all concept names  $A$  we have  $A^{\mathcal{I}'} = (A^{\mathcal{I}} \cup \{a^{\mathcal{I}} \mid A(a) \in \mathcal{E}(\alpha, \mathcal{I}, \text{CR})\}) \setminus \{a^{\mathcal{I}} \mid \neg A(a) \in \mathcal{E}(\alpha, \mathcal{I}, \text{CR})\}$ , and
- for all role names  $r$  we have  $r^{\mathcal{I}'} = (r^{\mathcal{I}} \cup \{(a^{\mathcal{I}}, b^{\mathcal{I}}) \mid r(a, b) \in \mathcal{E}(\alpha, \mathcal{I}, \text{CR})\}) \setminus \{(a^{\mathcal{I}}, b^{\mathcal{I}}) \mid \neg r(a, b) \in \mathcal{E}(\alpha, \mathcal{I}, \text{CR})\}$ .

The sequence of actions  $\alpha_1, \dots, \alpha_n$  may transform  $\mathcal{I}$  to  $\mathcal{I}'$  w.r.t.  $\mathcal{T}$  and CR (denoted by  $\mathcal{I} \Longrightarrow_{\alpha_1, \dots, \alpha_n}^{\mathcal{T}, \text{CR}} \mathcal{I}'$ ) iff there are interpretations  $\mathcal{I}_0, \dots, \mathcal{I}_n$  such that  $\mathcal{I} = \mathcal{I}_0$ ,  $\mathcal{I}_n = \mathcal{I}'$ , and  $\mathcal{I}_{i-1} \Longrightarrow_{\alpha_i}^{\mathcal{T}, \text{CR}} \mathcal{I}_i$  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  $\mathcal{I} \Longrightarrow_{\alpha}^{\mathcal{T}, \text{CR}} \mathcal{I}'$ . However, sometimes there may not exist any such interpretation  $\mathcal{I}'$ , either because  $\mathcal{E}(\alpha, \mathcal{I}, \text{CR})$  is contradictory, or because the new interpretation

induced by  $\mathcal{E}(\alpha, \mathcal{I}, \text{CR})$  is not a model of  $\mathcal{T}$ . If this happens in the case where  $\alpha = (\text{pre}, \text{post})$  is actually *applicable* to  $\mathcal{I}$  (i.e.,  $\mathcal{I} \models \text{pre}$ ), 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 2.5.** *The action  $\alpha$  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  $\mathcal{I} \models \text{pre}$ , there exists an interpretation  $\mathcal{I}'$  with  $\mathcal{I} \xRightarrow{\alpha}^{\mathcal{T}, \text{CR}} \mathcal{I}'$ .*

As argued above, the action `HireJohn` is not consistent w.r.t. the TBox consisting of the GCIs of Example 2.1 and the empty set of causal relationships, but it becomes consistent if we add the causal relationship introduced below Definition 2.3.

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 problem is taken from [3], with the only difference that we use the “may transform” relation introduced in Definition 2.4, which takes causal relationships into account, instead of the one employed in [3].

**Definition 2.6 (Projection problem).** *Let  $\alpha_1, \dots, \alpha_n$  be a sequence of actions such that, for all  $i, 1 \leq i \leq n$ , the action  $\alpha_i$  is consistent w.r.t.  $\mathcal{T}$  and CR. The assertion  $\varphi$  is a consequence of applying  $\alpha_1, \dots, \alpha_n$  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  $\mathcal{I} \xRightarrow{\alpha_1, \dots, \alpha_n}^{\mathcal{T}, \text{CR}} \mathcal{I}'$ , then  $\mathcal{I}' \models \varphi$ .*

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.

### 3 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.

#### 3.1 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  $\alpha$  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 3.1.** Let  $\alpha = (\text{pre}, \text{post})$  be an action and CR a finite set of causal relationships. The ABox  $\mathcal{A}(\alpha, \text{CR})$  is defined as follows:

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

A diagram  $\mathcal{D}$  for  $\alpha$  and CR is a maximal, consistent subset of  $\mathcal{A}(\alpha, \text{CR})$ . We denote the set of all diagrams for  $\alpha$  and CR by  $\mathfrak{D}(\alpha, \text{CR})$ .

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  $\alpha$  to  $\mathcal{I}$  w.r.t. CR. Given a diagram  $\mathcal{D}$ , we will now define a set  $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \text{CR})$  such that  $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \text{CR}) = \mathcal{E}(\alpha, \mathcal{I}, \text{CR})$  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:  $\widehat{\text{Dir}}(\alpha, \mathcal{D}) := \{\psi \mid \varphi/\psi \in \text{post} \wedge \varphi \in \mathcal{D}\}$ .

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

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

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

Checking which of the sets  $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \text{CR})$  for  $\mathcal{D} \in \mathfrak{D}(\alpha, \text{CR})$  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.  $\alpha$  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  $\text{pre}$  of  $\alpha$ , it is enough to consider diagrams  $\mathcal{D}$  that are consistent with  $\text{pre}$ .

**Lemma 3.2.** The action  $\alpha = (\text{pre}, \text{post})$  is consistent w.r.t. CR iff  $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \text{CR})$  is not contradictory for all  $\mathcal{D} \in \mathfrak{D}(\alpha, \text{CR})$  for which  $\mathcal{D} \cup \text{pre}$  is consistent.

This lemma yields a PSPACE decision procedure for deciding consistency of an action w.r.t. a finite set of causal relationships. In order to check whether  $\alpha$  is inconsistent, we first guess<sup>5</sup> a diagram  $\mathcal{D} \in \mathfrak{D}(\alpha, \text{CR})$ , and then check whether  $\mathcal{D} \cup \text{pre}$  is consistent using the PSPACE decision procedure for ABox consistency in  $\mathcal{ALCO}$  [14]. If  $\mathcal{D} \cup \text{pre}$  is consistent, we compute the set  $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \text{CR})$ . This can be realized in polynomial time by performing the iteration used in the definition of  $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \text{CR})$ . Checking whether this set is contradictory is obviously also possible in polynomial time.

<sup>5</sup> Recall that PSPACE = NPSpace according to Savitch's theorem.

This PSPACE upper bound is optimal since the ABox inconsistency problem in  $\mathcal{ALCO}$ , which is known to be PSPACE-complete [14], 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 3.3.** *The problem of deciding consistency of an action w.r.t. a finite set of causal relationships is PSPACE-complete for  $\mathcal{ALCO}$ .*

### 3.2 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  $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \text{CR})$ , we must check whether, for any model  $\mathcal{I}$  of  $\mathcal{T}$  and  $\mathcal{D}$  that satisfies the preconditions of  $\alpha$ , the interpretation  $\mathcal{I}'$  obtained from  $\mathcal{I}$  by applying the effects in  $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \text{CR})$  (see Definition 2.4) is a model of  $\mathcal{T}$ . To this purpose, we first define an unconditional action  $\beta_{\alpha, \text{CR}, \mathcal{D}}$  that, applied to models of  $\mathcal{D}$ , has the same effect as  $\alpha$  w.r.t.  $\text{CR}$ . Then, we adapt the approach for solving the projection problem introduced in [3] to the problem of checking whether  $\beta_{\alpha, \text{CR}, \mathcal{D}}$  transforms models of  $\mathcal{T}$  into models of  $\mathcal{T}$ .

**Definition 3.4.** *Let  $\alpha = (\text{pre}, \text{post})$  be an action,  $\text{CR}$  a finite set of causal relationships, and  $\mathcal{D} \in \mathfrak{D}(\alpha, \text{CR})$ . The action  $\beta_{\alpha, \text{CR}, \mathcal{D}}$  has  $\text{pre} \cup \mathcal{D}$  as set of preconditions and  $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \text{CR})$  as set of (unconditional) post-conditions.*

The following lemma is an easy consequence of the definition of  $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \text{CR})$  and the semantics of actions (Definition 2.4).

**Lemma 3.5.** *For all  $\mathcal{D} \in \mathfrak{D}(\alpha, \text{CR})$ , all models  $\mathcal{I}$  of  $\mathcal{D}$ , and all interpretations  $\mathcal{I}'$ , we have  $\mathcal{I} \xrightarrow{\emptyset, \text{CR}}_{\alpha} \mathcal{I}'$  iff  $\mathcal{I} \xrightarrow{\emptyset, \emptyset}_{\beta_{\alpha, \text{CR}, \mathcal{D}}} \mathcal{I}'$ .*

The approach for solving the projection problem introduced in [3] considers a finite sequence of actions  $\beta_1, \dots, \beta_n$ . In the present section, we are only interested in the special case where  $n = 1$ . However, since we will adopt this approach also in the next section, 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. In our setting, the *relevant* role names (concept descriptions) will be the ones occurring in the input of the consistency or projection algorithm (see [2] for details). For every ABox assertion  $\varphi$  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  $\beta_1$  to  $\mathcal{I}_0$  yields a successor interpretation  $\mathcal{I}_1$ , the application of  $\beta_2$  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  $\mathcal{I}_0, \mathcal{I}_1, \dots, \mathcal{I}_n$  into a single interpretation  $\mathcal{J}$  such that the relevant assertion  $\varphi$  holds in  $\mathcal{I}_i$  iff its time-stamped variant  $\varphi^{(i)}$  holds in  $\mathcal{J}$ . In order to enforce that  $\mathcal{J}$  really encodes a sequence of interpretations induced by the application of the action sequence  $\beta_1, \dots, \beta_n$ , we require it to be a model of the (acyclic) TBox  $\mathcal{T}_{\text{red}}$  and the ABox  $\mathcal{A}_{\text{red}}$ . Due to the space constraints, we cannot describe the construction of  $\mathcal{T}_{\text{red}}$  and  $\mathcal{A}_{\text{red}}$  here. This construction is very similar to the one introduced in [4], and it is described in detail in [2].<sup>6</sup> Here, we only recall the pertinent properties of  $\mathcal{T}_{\text{red}}$  and  $\mathcal{A}_{\text{red}}$  in the next lemma (whose proof is very similar to the one of Theorem 14 in [4]). It should be noted that our results actually do not depend on how the TBox  $\mathcal{T}_{\text{red}}$  and the ABox  $\mathcal{A}_{\text{red}}$  are exactly constructed. Any TBox and ABox satisfying the properties stated in the lemma can be used in our approach.

**Lemma 3.6.** *Let  $\beta_1, \dots, \beta_n$  be a sequence of ALCO 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  $\beta_1, \dots, \beta_n$ . Then, there are an ALCO ABox  $\mathcal{A}_{\text{red}}$  and an (acyclic) ALCO TBox  $\mathcal{T}_{\text{red}}$  of size polynomial in the size of  $\beta_1, \dots, \beta_n$  and  $\mathcal{R}$ , such that the following properties (a) and (b) hold:*

- (a) *For all interpretations  $\mathcal{I}_0, \dots, \mathcal{I}_n$  such that  $\mathcal{I}_i \xrightarrow{\beta_i, \emptyset}^{\emptyset, \emptyset} \mathcal{I}_{i+1}$  for every  $i, 0 \leq i < n$ , there exists an interpretation  $\mathcal{J}$  such that  $\mathcal{J} \models \mathcal{A}_{\text{red}}$ ,  $\mathcal{J} \models \mathcal{T}_{\text{red}}$ , and*
  - (i) *for all  $i, 0 \leq i \leq n$  and for all relevant assertions  $\psi$ :  $\mathcal{I}_i \models \psi$  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  $\mathcal{J} \models \mathcal{A}_{\text{red}}$  and  $\mathcal{J} \models \mathcal{T}_{\text{red}}$ , there exist interpretations  $\mathcal{I}_0, \dots, \mathcal{I}_n$  such that  $\mathcal{I}_i \xrightarrow{\beta_i, \emptyset}^{\emptyset, \emptyset} \mathcal{I}_{i+1}$  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  $\alpha = (\text{pre}, \text{post})$  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  $\alpha$ , CR, or  $\mathcal{T}$ . Given a diagram  $\mathcal{D} \in \mathfrak{D}(\alpha, \text{CR})$ , we can compute the set  $\hat{\mathcal{E}}(\alpha, \mathcal{D}, \text{CR})$ , and check whether this set is non-contradictory. If this is the case, then we consider the action  $\beta_{\alpha, \text{CR}, \mathcal{D}}$ , and test whether an application of this action transforms models of  $\mathcal{T}$  satisfying  $\text{pre}$  and  $\mathcal{D}$  into models of  $\mathcal{T}$ . This test can be realized using the ABox  $\mathcal{A}_{\text{red}}$  and the (acyclic) TBox  $\mathcal{T}_{\text{red}}$  of Lemma 3.6.

**Lemma 3.7.** *The action  $\alpha$  is consistent w.r.t.  $\mathcal{T}$  and CR iff the following holds for all  $\mathcal{D} \in \mathfrak{D}(\alpha, \text{CR})$ : if  $\mathcal{D} \cup \text{pre}$  is consistent w.r.t.  $\mathcal{T}$ , then*

<sup>6</sup> Note that this construction makes use of nominals.

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

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  $\mathfrak{D}(\alpha, \text{CR})$ . For each element  $\mathcal{D} \in \mathfrak{D}(\alpha, \text{CR})$ , consistency of  $\mathcal{D} \cup \text{pre}$  w.r.t.  $\mathcal{T}$  can be tested in exponential time since reasoning in  $\mathcal{ALCO}$  w.r.t. (general) TBoxes is EXPTIME-complete [5]. For a given diagram  $\mathcal{D}$ , the set  $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \text{CR})$  as well as  $\beta_{\alpha, \text{CR}, \mathcal{D}}$  and  $\mathcal{R}$  can be computed in polynomial time, and the same is true for the construction of the ABox  $\mathcal{A}_{\text{red}}$  and the TBox  $\mathcal{T}_{\text{red}}$  using  $\beta_{\alpha, \text{CR}, \mathcal{D}}$  and  $\mathcal{R}$ . Checking whether  $\widehat{\mathcal{E}}(\alpha, \mathcal{D}, \text{CR})$  is contradictory or not can also be realized in polynomial time. Finally, testing whether  $\mathcal{A}_{\text{red}} \cup \mathcal{T}_{\text{red}} \cup \mathcal{D}^{(0)} \cup \text{pre}^{(0)} \cup \mathcal{T}^{(0)} \models \mathcal{T}^{(1)}$  is again a reasoning problem for  $\mathcal{ALCO}$ , which can be solved in exponential time. Overall, we have seen that Lemma 3.7 yields a consistency test that requires at most exponentially many calls to EXPTIME reasoning procedures. This yields an EXPTIME upper bound for the complexity of the consistency problem. EXPTIME-hardness can be shown similarly to our proof of PSPACE-hardness for the case with an empty TBox.

**Theorem 3.8.** *The problem of deciding consistency of an action w.r.t. a TBox and a finite set of causal relationships is EXPTIME-complete for  $\mathcal{ALCO}$ .*

## 4 Deciding Projection

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

Our solution of the projection problem w.r.t.  $\mathcal{T}$  and  $\text{CR}$  uses the same ideas as the solution of the consistency sketched in Section 3. *First*, instead of considering interpretations  $\mathcal{I}_0, \dots, \mathcal{I}_{n-1}$ , we consider diagrams  $\mathcal{D}_0, \dots, \mathcal{D}_{n-1}$ , where  $\mathcal{D}_i \in \mathfrak{D}(\alpha_{i+1}, \text{CR})$  for  $i = 0, \dots, n-1$ .<sup>7</sup> *Second*, we use the original sequence of actions  $\alpha_1, \dots, \alpha_n$  and the diagrams  $\mathcal{D}_0, \dots, \mathcal{D}_{n-1}$  to build the corresponding sequence of actions  $\beta_{\alpha_1, \text{CR}, \mathcal{D}_0}, \dots, \beta_{\alpha_n, \text{CR}, \mathcal{D}_{n-1}}$ . Lemma 3.5 then tells us that, for all models  $\mathcal{I}_{i-1}$  of  $\mathcal{D}_{i-1}$  and all interpretations  $\mathcal{I}_i$  we have  $\mathcal{I}_{i-1} \xRightarrow{\emptyset, \text{CR}}_{\alpha_i} \mathcal{I}_i$  iff  $\mathcal{I}_{i-1} \xRightarrow{\emptyset, \emptyset}_{\beta_{\alpha_i, \text{CR}, \mathcal{D}_{i-1}}} \mathcal{I}_i$ . *Third*, we use the sequence  $\beta_{\alpha_1, \text{CR}, \mathcal{D}_0}, \dots, \beta_{\alpha_n, \text{CR}, \mathcal{D}_{n-1}}$  and the set of relevant role names and concept descriptions  $\mathcal{R}$  to construct an ABox  $\mathcal{A}_{\text{red}}$  and (acyclic) a TBox  $\mathcal{T}_{\text{red}}$  such that the properties (a) and (b) of Lemma 3.6 hold. 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

<sup>7</sup> Note that it is enough to consider diagrams  $\mathcal{D}_0, \dots, \mathcal{D}_{n-1}$  for  $\mathcal{I}_0, \dots, \mathcal{I}_{n-1}$  since no action is applied to  $\mathcal{I}_n$ .

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  $\varphi$ . To be more precise, we can show that the following characterization of the projection problem holds:

**Lemma 4.1.** *Let  $\alpha_1, \dots, \alpha_n$  be a sequence of actions,  $\mathcal{T}$  a TBox, CR a finite set of causal relationships,  $\mathcal{A}$  an initial ABox, and  $\varphi$  an assertion. Then,  $\varphi$  is a consequence of applying  $\alpha_1, \dots, \alpha_n$  to  $\mathcal{A}$  w.r.t.  $\mathcal{T}$  and CR iff for all diagrams  $\mathcal{D}_0, \dots, \mathcal{D}_{n-1}$  such that  $\mathcal{D}_i \in \mathfrak{D}(\alpha_{i+1}, \text{CR})$  for  $i = 0, \dots, n-1$ , we have*

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

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

It is easy to see that this lemma yields an EXPTIME decision procedure for the projection problem in  $\mathcal{ALCO}$ . In fact, one needs to consider exponentially many sequences of diagrams  $\mathcal{D}_0, \dots, \mathcal{D}_{n-1}$ . For each such sequence, the actions  $\beta_{\alpha_1, \text{CR}, \mathcal{D}_0}, \dots, \beta_{\alpha_n, \text{CR}, \mathcal{D}_{n-1}}$ , and thus also  $\mathcal{A}_{\text{red}}$  and  $\mathcal{T}_{\text{red}}$ , can be constructed in polynomial time. Thus, the inference problem (1) is of polynomial size, and it can be solved in exponential time since reasoning in  $\mathcal{ALCO}$  w.r.t. a (general) TBox is EXPTIME-complete. EXPTIME-hardness of the projection problem can easily be shown by a reduction of concept satisfiability w.r.t. a TBox in  $\mathcal{ALCO}$ .

**Theorem 4.2.** *The projection problem w.r.t. a TBox and a finite set of causal relationships for  $\mathcal{ALCO}$  is EXPTIME-complete.*

For the special case of an empty TBox, the decision procedure derived from Lemma 4.1 actually needs only polynomial space. In fact, the exponentially many sequences of diagrams  $\mathcal{D}_0, \dots, \mathcal{D}_{n-1}$  can be enumerated within polynomial space, and for  $\mathcal{T} = \emptyset$ , the inference problem (1) contains no GCIs since TBox  $\mathcal{T}_{\text{red}}$  is acyclic. Thus, it can be solved within PSPACE. PSPACE-hardness of the projection problem is again easy to show.

**Corollary 4.3.** *The projection problem w.r.t. the empty TBox and a finite set of causal relationships is PSPACE-complete for  $\mathcal{ALCO}$ .*

## 5 Additional Results and Future Work

Our approach for deciding the consistency and the projection problem works not only for  $\mathcal{ALCO}$ , but also for all the other DLs considered in [3]. Basically, we can use the same algorithms. 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 (consistency of actions in  $\mathcal{ALCQI}$  and  $\mathcal{ALCQIO}$  in the case where the TBox is non-empty), we get the matching hardness results by a reduction from such a basic inference problem. The complexity results obtained this way are listed in Table 2. They are proved in detail in [2]. The table shows

**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* PTIME<sup>NEXPTIME</sup>.

	TBox?	ACC	ACC $\circ$	ACC $\circ$	ACCI	ACC $\circ\circ$	ACC $\circ\mathcal{I}$	ACC $\circ\mathcal{O}$	ACC $\circ\mathcal{I}\mathcal{O}$
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

that (with the two exceptions mentioned above) the projection problem and the consistency problem in DL-based action formalisms with causal relationships is not harder than reasoning in the underlying DL (extended with nominals).

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. 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. Thus, consistency can no longer be characterized by an analog of Lemma 3.7.

When defining our semantics for actions in the presence of causal relationships, we followed the approach used in [17, 6] rather than the one employed by [8, 15]. 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 [8, 15].

Instead of trying to decide the projection problem directly, one can also follow the progression approach: given an action and a (possibly incomplete) description of the current state, this approach tries to compute a description of the possible successor states. Projection then boils down to computing consequences of this successor description. For DL-based action theories, progression has been investigated in [10]. It would be interesting to see whether the results obtained there can be extended to the DL-based action theories with causal relationships considered in the present paper.

This paper follows the approach for obtaining decidability results for action theories introduced in [3], which is based on the idea of restricting the base logic to a decidable DL. In the literature, other ways of restricting the base logic to achieve this goal have been considered. For example, in [11] the authors consider so-called local effect actions<sup>8</sup> and restrict the base logic to so-called “proper<sup>+</sup> knowledge bases” [12]. They show that, in this setting, progression is efficiently

<sup>8</sup> Note that our DL-based actions are local effect actions.

computable, which implies that the projection problem is efficiently decidable. It would be interesting to see whether this result can be extended to actions theories with causal relationships.

## References

1. F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.
2. F. Baader, M. Lippmann, and H. Liu. Adding causal relationships to DL-based action formalisms. LTCS-Report 10-01, Chair of Automata Theory, Technische Universität Dresden, 2010. See <http://lat.inf.tu-dresden.de/research/reports.html>.
3. F. Baader, C. Lutz, M. Miličić, U. Sattler, and F. Wolter. Integrating description logics and action formalisms: First results. In *Proceedings of AAAI-2005 and IAAI-2005*, pages 572–577. AAAI Press/The MIT Press, 2005. Short version of [4].
4. F. Baader, M. Miličić, C. Lutz, U. Sattler, and F. Wolter. Integrating description logics and action formalisms for reasoning about web services. LTCS-Report 05-02, Chair of Automata Theory, Technische Universität Dresden, 2005. Long version of [3]; available at <http://lat.inf.tu-dresden.de/research/reports.html>.
5. G. De Giacomo. *Decidability of Class-Based Knowledge Representation Formalisms*. PhD thesis, Università di Roma “La Sapienza”, 1995.
6. M. Denecker, D. Theseider-Dupré, and K. van Belleghem. An inductive definition approach to ramifications. *Linköping Electronic Articles in Computer and Information Science*, 3(7):1–43, 1998.
7. A. Herzig. The PMA revisited. In *Proceedings of KR-1996*, pages 40–50, 1996.
8. F. Lin. Embracing causality in specifying the indirect effects of actions. In *Proceedings of IJCAI-1995*, pages 1985–1993, 1995.
9. H. Liu, C. Lutz, M. Miličić, and F. Wolter. Reasoning about actions using description logics with general TBoxes. In *Proceedings of JELIA-2006*, volume 4160 of *LNAI*, pages 266–279. Springer-Verlag, 2006.
10. H. Liu, C. Lutz, M. Miličić, and F. Wolter. Updating description logic ABoxes. In *Proceedings of KR-2006*, pages 46–56. AAAI Press, 2006.
11. Y. Liu and G. Lakemeyer. On first-order definability and computability of progression for local-effect actions and beyond. In *Proceedings of IJCAI-2009*, pages 860–866, 2009.
12. Y. Liu and H. J. Levesque. Tractable reasoning in first-order knowledge bases with disjunctive information. In *Proceedings of AAAI-2005 and IAAI-2005*, pages 639–644. AAAI Press/The MIT Press, 2005.
13. R. Reiter. *Knowledge in Action: Logical Foundations for Describing and Implementing Dynamical Systems*. The MIT Press, Bradford Books, 2001.
14. A. Schaerf. Reasoning with individuals in concept languages. *Data Knowledge Engineering*, 13(2):141–176, 1994.
15. M. Thielscher. Ramification and causality. *AIJ*, 89(1–2):317–364, 1997.
16. M. Thielscher. *Reasoning Robots: The Art and Science of Programming Robotic Agents*, volume 33 of *Applied Logic Series*. Springer Netherlands, 2005.
17. K. van Belleghem, M. Denecker, and D. Theseider-Dupré. A constructive approach to the ramification problem. In *Proceedings of ESSLLI-1998*, pages 1–17, 1998.
18. M. Winslett. Reasoning about action using a possible models approach. In *Proceedings of AAAI-1988*, pages 89–93, 1988.