Provably Correct Theories of Action[1]

Fangzhen Lin

Department of Computer Science

University of Toronto

Toronto, Canada M5S 1A4

Yoav Shoham

Department of Computer Science

Stanford University

Stanford, CA 94305

[1]Most of the research was done while the first author was with the Department of Computer Science at Stanford University, where both authors' research was supported in part by a grant from the Air Force Office of Scientific Research. The work of the first author was also supported in part by the Government of Canada Institute for Robotics and Intelligent Systems.

Abstract

We investigate logical formalization of the effects of actions in the situation calculus. We propose a formal criterion against which to evaluate theories of deterministic actions. We show how the criterion provides us a formal foundation upon which to tackle the frame problem, as well as its variant in the context of concurrent actions. Our main technical contributions are in formulating a wide class of monotonic causal theories that satisfy the criterion, and showing that each such theory can be reformulated succinctly in circumscription.

1 Introduction

The histories of the frame problem [18], and of the particular Yale Shooting Problem (YSP) which has become its best known illustration [6], have followed a disturbing pattern. The frame problem itself, although introduced in the context of formalizing common sense, was never formally defined, and was only illustrated through suggestive examples. This is an initial disturbing factor.

A second disturbing factor is that, despite the lack of a formal definition, arguments were made that a particular collection of formal tools, namely nonmonotonic logics, would 'solve' the problem. Again, no formal analysis was provided, and the claim was based only on sketchy examples.

A third disturbing factor is that the response in the community was not that the above claim is ill defined, but that it's false. In particular, the Yale Shooting Problem was proposed as an illustration of the falseness.

Given these shaky foundations, it is not surprising that subsequent research on the topic became increasingly splintered and controversial. From the outset there were arguments that the YSP is not a problem at all [16]. Simultaneously there were several proposed solutions to it [26, 10, 8, 11, 7]. Then there were counter-arguments that each of these solutions was 'wrong,' in that it didn't solve other problems such as the 'qualification problem' or the 'ramification problem,' or that that it supported 'prediction' but not 'explanation.' New examples were then devised, with names such as the Stolen Car Problem [8] and the Stanford Murder Mystery [1]. The responses again varied, including dismissals of some of the complaints, as well as new solutions to the YSP that allegedly avoided some of these problems ([19, 15, 2], etc.). Each solution has attracted some measure of criticism.

The lack of precise criteria against which to evaluate theories of action does not mean that the research has been worthless; quite the contrary. It is widely recognized that the frame problem is real and that its identification was insightful, even if it has not yet been formally defined. Similarly, the YSP led to major improvements in the understanding of

nonmonotonic logics and their applications, as well as to better understanding of formal temporal reasoning.

Nevertheless, in order to better understand what we have achieved so far, it is important to arrive at precise criteria for the adequacy of theories of action. In this paper we take a step in that direction. Specifically, we identify a formal yet intuitive adequacy criterion, prove a certain class of monotonic theories of action adequate relative to this criterion, and then show an equivalent nonmonotonic counterpart for a significant subclass of the theories. To our knowledge this is the first instance of provably-correct nonmonotonic temporal reasoning with respect to a general criterion.

2 The approach

Recall the following intuitive explanation of the frame problem. Suppose we are trying to formalize the effects of actions. Usually, an action causes only a small number of changes. For example, when we paint a block, only the color of the block changes. Most of the other facts, such as the location of the block, the smell of the paint, *et cetera*, do not change. The frame problem is the problem of representing concisely these numerous facts that are unaffected by an action.

Our approach to making the problem precise is conceptually very simple, and perhaps best illustrated by an analogy with simple databases. Consider a database of flight connections between pairs of cities. One way to structure the database is by a set of assertions of the form $Flight(x, y)$ and $\neg Flight(x, y)$, where for each pair of cities A, B exactly one of $Flight(A, B)$ and $\neg Flight(A, B)$ appears in the database. The semantics of this database are those of classical logic. This is an *epistemologically complete* representation since for any pair of cities it tells one whether the two are connected. However, while epistemologically adequate, the representation is pragmatically inadequate: it requires representation of all pairs of cities, whereas the connectivity graph is usually quite sparse.

The solution is, of course, to omit all the $\neg Flight(x, y)$ assertions, and infer $\neg Flight(A, B)$

'by default' in the absence of $Flight(A, B)$. This is a simple application of the *closed-world assumption (CWA)* [22], and the equivalent monotonic formulation can be regenerated from the abbreviated representation through *database completion* [3]. This concise representation is also epistemologically complete since it too entails either $Flight(A, B)$ or $\neg Flight(A, B)$ for all pairs of cities A and B, albeit nonmonotonically. Furthermore, the nonmonotonic version is sound and complete relative to the monotonic version in the sense that they entail the same facts.

Thus there are two criteria for evaluating the epistemological adequacy of a theory. Both monotonic and nonmonotonic theories can be tested for their epistemological completeness; this is an absolute criterion. In addition, nonmonotonic theories can be tested for equivalence to a given, monotonic, often better understood, and typically much larger, theory; this is a relative criterion.

In principle our treatment of theories of actions will be identical; we will require them to be complete, and furthermore evaluate a nonmonotonic theory relative to an equivalent and larger monotonic one. The complications will arise from a more complex definition of epistemological completeness, resulting difficulty in determining whether a given theory is indeed epistemologically complete, and a nonmonotonic mechanism that is more complex than CWA.

3 Logical preliminaries

We shall base our presentation on the situation calculus [18], although we believe that a similar treatment is possible in other frameworks such as temporal logics. For now we shall adopt the standard situation calculus, which precludes the representation of certain notions such as concurrent actions. In later sections, we shall consider a version of the situation calculus which allows concurrent actions. In this section we review the language for discussing the situation calculus. We do this briefly and almost apologetically since we realize that the situation calculus is very well known; however, we feel that in this paper it

is important to be precise about the language.

Our language $\mathcal{L}$ is a many-sorted first-order one with equality. Its three domain independent sorts are:

1. Situation sort: with situation constants $S_1, S_2, ...$, and situation variable $s, s_1, s_2, ...$ We will use $S, S', ...$ as meta-variables for ground situation terms.

2. Action sort: with action constants $A_1, A_2, ...$, and action variables $a, a_1, a_2, ...$ We will use $A, A', ...$ as meta-variables for ground action terms, which will be action constants in this paper.

3. Propositional fluent sort: with fluent constants $P_1, P_2, ...$, and fluent variables $p, p_1, p_2, ...$ We will use $P, P', ...$ as meta-variables for ground fluent terms, which will be fluent constants in this paper.

We have a binary function $Result$ which maps an action and a situation into another situation. Intuitively, $Result(a, s)$ is the resulting situation of performing a in situation s. We also have a binary predicate $H(p, s)$, which asserts that fluent p holds in situation s.

As with any other language, we may interpret $\mathcal{L}$ classically, assuming the standard notion of entailment, or nonmonotonically, using some form of nonmonotonic entailment.

4 Epistemologically complete theories of action

Suppose we want to use our language to state that action $Toggle$ changes the truth value of the fluent P_1. We may wish to use the following axiom:

$$\forall s. H(P_1, s) \equiv \neg H(P_1, Result(Toggle, s)). \tag{1}$$

Intuitively speaking, this axiom alone is not enough. For example, it tells us nothing about the effects of $Toggle$ on P_2; for that we would need to add the following so-called frame axiom:

$$\forall s. H(P_2, s) \equiv H(P_2, Result(Toggle, s)). \tag{2}$$

Clearly we need such a frame axiom for every fluent that is different from P_1. But, are those frame axioms enough? In other words, do these axioms together completely formalize our knowledge about $Toggle$? For this simple example it is easy to convince oneself that these axioms indeed do completely formalize the effects of the action $Toggle$. In general, however, the answer may not be obvious, and it is essential for us to have a precise definition of when a first-order theory is a complete formalization of an action.

In this paper we are only concerned with deterministic actions. Intuitively speaking, a theory $\mathcal{T}$ completely formalizes the effects of a deterministic action A if, given a complete description of the initial situation, it enables us to deduce a complete description of the resulting situation after A is performed. We now proceed to make this intuition precise. First, we notice that in actual applications, it is most convenient to talk about whether a description of a situation is complete with respect to a set of fluents in which we are interested. Thus we shall define conditions under which a theory is complete *about an action* and *with respect to a set of fluents*. This fixed set of fluents plays a role similar to Lifschitz's $Frame$ predicate [13]. In the following, let $\mathcal{P}$ be a fixed set of fluent constants.

Definition 4.1 *A set $\mathcal{S}$ of ground literals is a* state of the situation S *(with respect to $\mathcal{P}$) if there is a subset $\mathcal{P}'$ of $\mathcal{P}$ such that*

$$\mathcal{S} = \{H(P,S) \mid P \in \mathcal{P}'\} \cup \{\neg H(P,S) \mid P \in \mathcal{P} - \mathcal{P}'\}.$$

Therefore, if $\mathcal{S}$ is a state of S, then for any $P \in \mathcal{P}$, either $H(P,S) \in \mathcal{S}$ or $\neg H(P,S) \in \mathcal{S}$. Intuitively, states completely characterize situations with respect to the fluents in $\mathcal{P}$. Thus we can say that a first-order theory $\mathcal{T}$ is epistemologically complete about action A (with respect to $\mathcal{P}$) if it is consistent, and for any ground situation term S, any state $\mathcal{S}$ of S, and any fluent $P \in \mathcal{P}$, either $\mathcal{T} \cup \mathcal{S} \models H(P, Result(A,S))$ or $\mathcal{T} \cup \mathcal{S} \models \neg H(P, Result(A,S))$, where $\models$ is classical first-order entailment.

However, as we said earlier, the notion of epistemological completeness is not limited to monotonic first-order theories. In general, for any given monotonic or nonmonotonic

entailment $\models_{\sqsubset}$, we can define when a theory is epistemologically complete about an action according to the entailment $\models_{\sqsubset}$:

Definition 4.2 *A theory $\mathcal{T}$ is* epistemologically complete *about action A (with respect to $\mathcal{P}$, and according to $\models_{\sqsubset}$) if $\mathcal{T} \not\models_{\sqsubset} False$, and for any ground situation term S, any state $\mathcal{S}$ of S, and any fluent $P \in \mathcal{P}$, there is a finite subset $\mathcal{S}'$ of $\mathcal{S}$ such that either $\mathcal{T} \models_{\sqsubset} \bigwedge \mathcal{S}' \supset H(P, Result(A, S))$ or $\mathcal{T} \models_{\sqsubset} \bigwedge \mathcal{S}' \supset \neg H(P, Result(A, S))$.*

We note that for any sets $\mathcal{T}$, $\mathcal{S}$, and formula φ, $\mathcal{T} \cup \mathcal{S} \models \varphi$ if there is a finite subset $\mathcal{S}'$ of $\mathcal{S}$ such that $\mathcal{T} \models \bigwedge \mathcal{S}' \supset \varphi$. Thus if we replace $\models_{\sqsubset}$ in Definition 4.2 by classical entailment $\models$, we get the same definition we had earlier for monotonic first-order theories.

Referring back to the database example from section 2, we note that requiring a theory to be epistemologically complete is analogous to requiring a representation for $Flight$ to tell us, for any pair (A, B) of cities, whether $Flight(A, B)$ is true.

In the following, we shall say that $\mathcal{T}$ is a monotonic (nonmonotonic) theory if the entailment relation associated with $\mathcal{T}$ is classical (nonmonotonic). Thus if $\mathcal{T}$ is an epistemologically complete theory according to classical entailment $\models$, then we say that $\mathcal{T}$ is an epistemologically complete monotonic theory.

In the following, we first re-examine the Yale Shooting Problem [6] as an extended example of the foregoing definitions. The YSP turns out to be a very special case of the class of theories we consider later.

5 The Yale shooting problem revisited

In the YSP we consider three actions: $Shoot$, $Load$, and $Wait$. After performing $Load$, the gun is loaded. If the gun is loaded, then after $Shoot$ is performed, Fred is dead. Thus we have the following two 'causal rules':

$$\forall s. H(Loaded, Result(Load, s)), \tag{3}$$

$$\forall s. H(Loaded, s) \supset H(Dead, Result(Shoot, s)). \tag{4}$$

This theory is, of course, insufficient to capture fully the effects of the three actions.

5.1 Monotonic completion

One way to achieve epistemological completeness is to supply frame axioms. Let $\mathcal{P} = \{Dead, Loaded\}$. For action $Shoot$, we have that for each $P \in \mathcal{P}$,

$$\forall s. \neg H(Loaded, s) \supset [H(P, s) \equiv H(P, Result(Shoot, s))], \tag{5}$$

$$\forall s. H(Loaded, s) \equiv H(Loaded, Result(Shoot, s)). \tag{6}$$

For action $Load$, we have

$$\forall s. H(Dead, s) \equiv H(Dead, Result(Load, s)). \tag{7}$$

For $Wait$, we have that for any $P \in \mathcal{P}$:

$$\forall s. H(P, s) \equiv H(P, Result(Wait, s)). \tag{8}$$

Let $\mathcal{T}_1 = \{(3), (4), (5), (6), (7), (8)\}$. It can be shown that the monotonic theory $\mathcal{T}_1$ is epistemologically complete about the actions $Wait$, $Shoot$, and $Load$ with respect to $\mathcal{P}$. Using first-order logic only, we can answer queries about the theory. As a 'temporal projection' example, we have

$$\mathcal{T}_1 \models \forall s. H(Dead, Result(Shoot, Result(Wait, Result(Load, s)))),$$

that is, Fred is dead after $Load$, $Wait$, and $Shoot$. As an example of 'temporal explanation', we have

$$\mathcal{T}_1 \models \forall s. \neg H(Dead, s) \wedge H(Dead, Result(Shoot, s)) \supset H(Loaded, s),$$

that is, in any situation, if Fred is initially alive, but dead after the gun is fired, then the gun must be initially loaded.

5.2 Nonmonotonic completion

Although the above monotonic theory is epistemologically complete, it does so at the expense of explicit frame axioms which are generated manually. We now provide an equivalent nonmonotonic theory that avoids these frame axioms.

It was implied in [17] that the frame axioms can be replaced by minimizing the following abnormality predicate ab with H allowed to vary:

$$\forall pas.\neg ab(p,a,s) \supset [H(p,s) \equiv H(p, Result(a,s))]. \tag{9}$$

The main technical result of [6] is that this does not work. We mentioned in the introduction the slew of proposed solutions, all criticized on some grounds or others. Surprisingly, most of them are actually correct relative to the above monotonic theory. We pick as an example chronological minimization [26], although other proposals, such as those in [10] and [8], would work as well.[1]

For the full definition of chronological minimization the reader is referred to [26]; we only remind the reader that in this framework the preferred models are those in which the minimized predicate is true as *late* as possible, rather than as *infrequently* as possible. In our framework, the obvious (partial) temporal ordering on situations is

$$S < Result(Shoot, S) < Result(Wait, Result(Shoot, S)) < ...$$

and so on. Like circumscription, we also need unique names assumptions for chronological minimization:

$$Loaded \neq Dead \neq Shoot \neq Load \neq Wait. \tag{10}$$

We now simply take $\mathcal{T}_2$ to be the conjunction of (3), (4), (9), and (10), and chronologically minimize ab in $\mathcal{T}_2$. It is now possible to show that $\mathcal{T}_1 \cup \{(10)\}$ and $\mathcal{T}_2$ are equivalent:

[1] Unfortunately, although chronological minimization works for the YSP, it does not work in general for the class of causal theories that we consider in the next section. For instance, given the theory $H(P,s) \supset \neg H(P, Result(A,s))$, the chronological minimization will incorrectly conclude that $\forall s.\neg H(P,s)$. We have not yet investigated whether other theories such as Lifschitz's pointwise circumscription will work in the general case.

for any φ in the language of $\mathcal{T}_1$, $\mathcal{T}_1 \cup \{(10)\} \models \varphi$ iff $\mathcal{T}_2 \models_{\sqsubset} \varphi$, where $\models$ is classical entailment and $\models_{\sqsubset}$ is the nonmonotonic entailment. In particular, we have that the nonmonotonic theory $\mathcal{T}_2$ is epistemologically complete.

We observe that previous arguments against chronological minimization and other solutions, for example by appeal to temporal explanation, involved incorrect use of the theories. For example, instead of $\mathcal{T}_2$, such arguments would use

$$\mathcal{T}_2 \cup \{H(Dead, Result(Shoot, S_0))\}.$$

Other arguments referred not to the YSP but to extended or modified examples; most of those are covered by the general treatment in the following sections.

We remark here that we have avoided formally claiming that $\mathcal{T}_2$ solves the frame problem for YSP. The reason is that we do not yet have a formal criterion to decide when a representation is *concise* enough to qualify as a solution to the frame problem. Until we have one, the frame problem will continue to contain an informal factor. However, this does not affect our claim about provable correctness of theories of action.

6 Causal theories

In this section we define a class of causal theories that includes the YSP as a special case.

In reasoning about action, our knowledge can be generally divided into two kinds. First we have knowledge about the environment where the actions are taken place, commonly referred to as *domain constraints*. Second we have knowledge about effects of actions themselves, usually called *causal rules*. This leads us to the following definition.

Definition 6.1 *Let $C(s)$, $R_i(s)$, $i = 1, ..., n$, $n \geq 0$, be formulas with a free variable s. The* causal theory *of the action A with the domain constraint C, and the direct effects $P_1, ..., P_n$ under the preconditions $R_1, ..., R_n$, respectively, is the following set of sentences: The domain constraint:*

$$\forall s.C(s), \tag{11}$$

and for each $1 \leq i \leq n$*, the causal rules:*

$$\forall s. R_i(s) \supset H(P_i, Result(A, s)). \tag{12}$$

Notice that the above definition is for a single action. The causal theory of a set of actions is the union of the causal theory of each individual action.

Of course, causal theories are generally incomplete. As with our solution to the YSP, we can provide both monotonic and nonmonotonic completions to them.

6.1 A monotonic completion of causal theories

There are two kinds of changes an action can cause. One is direct effects, which are determined by the causal rules. The other is side effects, which are determined by the direct effects and the domain constraints. Those that are neither direct effects nor side effects are assumed to be unchanged by the action.

Essentially, making causal theories epistemologically complete amounts to supplying frame axioms that capture facts which are not changed by actions. When the frame axioms are explicitly listed, the completions are called *monotonic*. Again in the following, we fix a set of fluents $\mathcal{P}$.

Definition 6.2 *Let* $\mathcal{T}$ *be a causal theory of* A *with the domain constraint* $C(s)$*, and the direct effects* $P_1, ..., P_n$ *under the preconditions* $R_1, ..., R_n$*, respectively. The* monotonic completion *of* $\mathcal{T}$*, written* $Comp(\mathcal{T})$*, is the union of* $\mathcal{T}$ *with the following frame axioms. For any subset* I *of* $N = \{1, ..., n\}$*, and any* $P \in \mathcal{P}$*, if*

$$\not\models \forall s. \bigwedge_{i \in I} H(P_i, s) \wedge C(s) \supset H(P, s)$$

and

$$\not\models \forall s. \bigwedge_{i \in I} H(P_i, s) \wedge C(s) \supset \neg H(P, s),$$

then the following is a frame axiom:

$$\forall s. \bigwedge_{i \in N-I} \neg R_i(s) \supset [H(P, s) \equiv H(P, Result(A, s))]. \tag{13}$$

Example 6.1 The causal theory $\mathcal{T}$ of the action $Wait$ with the domain constraint $C(s)$, and no direct effect is $\forall s.C(s)$. For each $P \in \mathcal{P}$, if

$$\not\models \forall s.C(s) \supset H(P,s)$$

and

$$\not\models \forall s(C(s) \supset \neg H(P,s)),$$

then the following sentence is a frame axiom:

$$\forall s.H(P,s) \equiv H(P, Result(Wait,s)).$$

Thus $Comp(\mathcal{T})$ is equivalent to the following two axioms:

$$\forall s.C(s),$$
$$\forall s.H(P,s) \equiv H(P, Result(Wait,s)).$$

where P is any fluent in $\mathcal{P}$. ■

Our formulation of $Comp(\mathcal{T})$ can be considered a generalization of that in [20] where domain constraints have to be hand-coded into the effect axioms. Our class of monotonic theories is also related to that in [25] and [24]. The exact relationships between the two are delicate. It seems that Schubert and Reiter made certain closure assumptions about actions, while we made similar ones about fluents.

We now formulate some sufficient conditions for $Comp(\mathcal{T})$ to be epistemologically complete.

Let $\mathcal{T}$ be a causal theory of A. For any state $\mathcal{S}$ of S, let

$$\begin{aligned} RES(A,\mathcal{S}) &= \{H(P,S') \mid P \in \mathcal{P}, \mathcal{S} \cup \mathcal{T} \models H(P,S')\} \\ &\cup \{\neg H(P,S') \mid P \in \mathcal{P}, \mathcal{S} \cup \mathcal{T} \models \neg H(P,S')\} \\ &\cup \{H(P,S') \mid H(P,S) \in \mathcal{S}, \mathcal{S} \cup \mathcal{T} \not\models \neg H(P,S')\} \\ &\cup \{\neg H(P,S') \mid \neg H(P,S) \in \mathcal{S}, \mathcal{S} \cup \mathcal{T} \not\models H(P,S')\}, \end{aligned}$$

where $S' = Result(A,S)$. We see that $RES(A,\mathcal{S})$ is a state of S' if $\mathcal{S} \cup \mathcal{T}$ is consistent.

Theorem 1 *Let $\mathcal{T}$ be a causal theory of A with the domain constraint C, and the direct effects $P_1, ..., P_n$ under the preconditions $R_1, ..., R_n$, $n \geq 0$, respectively. $Comp(\mathcal{T})$ is an epistemologically complete monotonic theory about A with respect to $\mathcal{P}$ if the following conditions are satisfied:*

Condition 1. *$C(s), R_1(s), ..., R_n(s)$ do not contain any situation terms other than s.*
Condition 2. *For any state $\mathcal{S}$ of S, either $\mathcal{S} \models \varphi(S)$ or $\mathcal{S} \models \neg\varphi(S)$, where $\varphi(S) \in \{C(S), R_1(S), ..., R_n(S)\}$, and $\varphi(S)$ is the result of substituting S for s in $\varphi(s)$.*
Condition 3. *$\forall s.C(s)$ is consistent.*
Condition 4. *For any state $\mathcal{S}$ of S, if $\mathcal{S} \models C(S)$, then $RES(A, \mathcal{S})$ is consistent, and $RES(A, \mathcal{S}) \models C(Result(A, S))$.*

The proof of this theorem is given in Appendix A. The class of the causal theories given by the theorem includes many of the blocks world examples found in the literature. If we ignore the predicates *frame* and *possible* in [14], then our class also includes the causal theories in the main theorem of [14]. Normally, the first three conditions of the theorem will be straightforward to verify. Condition 4 is the only nontrivial one. Let us see an example.

Example 6.2 The Extended Yale Shooting Problem [1]: We add one more fluent *Alive* into the YSP. Let $\mathcal{P} = \{Dead, Alive, Loaded\}$. The domain constraint $\forall s C(s)$ is:

$$\forall s.H(Alive, s) \equiv \neg H(Dead, s).$$

The causal rules are (3) and (4). Let us check the four conditions for the example. We do so for action *Shoot*. The cases for *Wait* and *Load* are trivial.

1. Conditions 1, 2, and 3: Trivial.

2. Condition 4: Let $\mathcal{S}$ be a state of S. Suppose $\mathcal{S} \models C(S)$, i.e.

$$H(Dead, S) \in \mathcal{S} \quad \text{iff} \quad \neg H(Alive, S) \in \mathcal{S}.$$

We can show that

$$H(Dead, Result(Shoot, S)) \in RES(Shoot, \mathcal{S})$$

iff

$$H(Dead, S) \in \mathcal{S} \text{ or } H(Loaded, S) \in \mathcal{S},$$

and similarly,

$$H(Alive, Result(Shoot, S)) \in RES(Shoot, \mathcal{S})$$

iff

$$H(Alive, S) \in \mathcal{S} \text{ and } \neg H(Loaded, S) \in \mathcal{S}.$$

Thus we have

$$H(Dead, Result(Shoot, S)) \in RES(Shoot, \mathcal{S})$$

iff

$$H(Alive, Result(Shoot, S)) \notin RES(Shoot, \mathcal{S})$$

iff

$$\neg H(Alive, Result(Shoot, S)) \in RES(Shoot, \mathcal{S}).$$

Therefore $RES(Shoot, \mathcal{S}) \models C(Result(Shoot, S))$.

■

6.2 A nonmonotonic completion

We now proceed to provide a nonmonotonic completion of causal theories. We shall use a version of circumscription.

Given a causal theory $\mathcal{T}$ of the action A, for any initial state

$$\{H(p, s) \mid p \in \mathcal{P} \text{ holds in } s\},$$

we shall minimize the difference between the initial state and the resulting state

$$\{H(p, Result(A, s)) \mid p \in \mathcal{P} \text{ holds in } Result(A, s)\}.$$

Semantically, we have the following definition:

Definition 6.3 *Let W be a sentence. A model M of W is a* preferred *one (w.r.t. the action A, and the set of fluents $\mathcal{P}$) if there does not exist another model M' of W such that:*

1. *M and M' share the same domains;*

2. *M and M' interpret everything except H the same;*

3. *There exists a variable assignment σ such that for any $P \in \mathcal{P}$,*

$$M, \sigma \models H(P, s) \quad \Leftrightarrow \quad M', \sigma \models H(P, s),$$
$$M, \sigma \models H(P, s) \equiv H(P, Result(A, s)) \quad \Rightarrow \quad M', \sigma \models H(P, s) \equiv H(P, Result(A, s))$$

and there is at least one $P \in \mathcal{P}$ such that

$$M', \sigma \models H(P, s) \equiv H(P, Result(A, s)) \text{ but } M, \sigma \not\models H(P, s) \equiv H(P, Result(A, s)).$$

We shall write $M' \sqsubset_{A,\mathcal{P}} M$ if M and M' satisfy the conditions in the above definition. For any sentence W and φ, we write $W \models_{\sqsubset}^{A,\mathcal{P}} \varphi$, or $W \models_{\sqsubset} \varphi$ when there are no possibility of confusion, if φ is true in every preferred model of W with respect to A and $\mathcal{P}$.

Under the assumption that the assertion $p \in \mathcal{P}$ can be formalized by a first-order formula, the semantic entailment $\models_{\sqsubset}$ can be captured by circumscription. Formally, suppose that $Frame(p)$ is a formula with a free variable p such that for any interpretation M, and any variable assignment σ of fluents, $M, \sigma \models Frame(p)$ iff there is a $P \in \mathcal{P}$ such that P is interpreted as $\sigma(p)$. For example, if $\mathcal{P} = \{P_1, P_2\}$, then $Frame(p)$ can be $p = P_1 \vee p = P_2$.

Let $ab(p)$ be the abbreviation of the following formula

$$Frame(p) \wedge [H(p, s) \equiv \neg H(p, Result(A, s))].$$

Our circumscriptive policy will be that for any situation S, we minimize $ab(p)(s/S)$ as a unary formula of p with $H(p, S)$ fixed but $H(p, S')$ allowed to vary for every S' that is different from S, where $ab(p)(s/S)$ is the result of replacing s in $ab(p)$ by S.

In order to do that, we extend the language $\mathcal{L}$ to include a new predicate H' that is similar to H. Then for any formula W, we minimize $ab(p)$ in the following formula with H' fixed and H allowed to vary:

$$W \wedge \forall p.H(p,s) \equiv H'(p,s). \tag{14}$$

The following proposition follows straightforwardly from the standard results about circumscription (cf. [9]):

Proposition 6.1 *For any sentence W, M is a preferred model of W iff M' is a model of $\forall s Circum(W'; ab; H)$, where M' is obtained from M by interpreting H' the same as H, W' is (14), and $Circum(W'; ab; H)$ is the circumscription of $ab(p)$ in W' with H allowed to vary.*

It follows from this proposition that for any sentence φ in $\mathcal{L}$,

$$W \models_{\sqsubset} \varphi \Leftrightarrow \forall s Circum(W'; ab; H) \models \varphi.$$

In the rest of this section, however, we shall choose to use the semantic entailment $\models_{\sqsubset}$.

We now proceed to show that this minimization policy yields the same frame axioms generated by the monotonic completion in last section. In order to use circumscription, we need to have some unique names axioms. We suppose that U_1 captures the unique names assumptions for the fluents in $\mathcal{P}$, and U_2 is the following axiom:

$$\begin{aligned}
&\forall as.Earlier(s, Result(a,s)) \wedge \\
&\forall ss_1s_2.Earlier(s,s_1) \wedge Earlier(s_1,s_2) \supset Earlier(s,s_2) \wedge \\
&\forall ss_1.Earlier(s,s_1) \supset s \neq s_1,
\end{aligned}$$

where *Earlier* is a new binary predicate. The purpose of U_2 is to capture the following unique names axiom:

$$S \neq Result(A,S) \neq Result(A, Result(A,S)) \neq ...$$

Now we can state the following result:

Theorem 2 *Under the assumptions and conditions in Theorem 1, for any formula φ in $\mathcal{L}$ (our original language without Earlier),*

$$Comp(\mathcal{T}) \cup \{U_1, U_2\} \models \varphi \Leftrightarrow W_{\mathcal{T}} \models_{\sqsubset} \varphi,$$

where $W_{\mathcal{T}}$ is the conjunction of U_1, U_2, and the axioms in $\mathcal{T}$.

The proof of this theorem is given in Appendix B. Now define $\models_{\sqsubset}$ as: For any formula W and φ, $\{W\} \models_{\sqsubset} \varphi$ iff $W \models_{\sqsubset} \varphi$. We then have the following corollary:

Corollary 2.1 *Under the assumptions in Theorem 2, $\{W_{\mathcal{T}}\}$ is an epistemologically complete nonmonotonic theory about A according to $\models_{\sqsubset}$.*

7 Limitations and possible extensions

What we have developed so far is a theory of a *single, primitive, and deterministic* action. What we hope for is that this theory of action will be the core from which theories about complex actions can be developed. In the remaining sections of this paper we show how this can be done for concurrent actions.

There is, however, an important limitation on our definition of epistemological completeness. We have assumed that fluent constants are the only closed fluent terms, and more importantly, that every fluent that we are interested in can be named by one of such fluent constants. We can imagine situations where such assumptions are false. For these cases, we need the following more general semantic definition:

> A theory $\mathcal{T}$ of the action A is *epistemologically complete* if for any two models M_1 and M_2 of $\mathcal{T}$ that share the domains, we have that for any variable assignment of situations σ_1, if
>
> $$M_1, \sigma_1, \sigma_2 \models H(p, s) \Leftrightarrow M_2, \sigma_1, \sigma_2 \models H(p, s)$$
>
> holds for any variable assignment σ_2 of fluents, then
>
> $$M_1, \sigma_1, \sigma_2 \models H(p, Result(A, s)) \Leftrightarrow M_2, \sigma_1, \sigma_2 \models H(p, Result(A, s))$$

holds for any variable assignment σ_2 of fluents as well.

In other words, if $\mathcal{T}$ is epistemologically complete, then for any domain D_p of fluents, $\mathcal{T}$ defines a function from the power set of $\{H(P,S) \mid P \in D_p\}$ to the power set of $\{H(P, Result(A,S)) \mid P \in D_p\}$.

Notice that the above notion was defined with respect to classical entailment. It can be defined with respect to any nonmonotonic entailment as well by substituting, for example, preferred models for models.

However, we notice that if the assumption that every fluent in $\mathcal{P}$ is named by one of the closed fluent terms is not true, then the meta-logical process of monotonically completing a causal theory in section 6.1 will in general no longer work. But if $\mathcal{P}$ can be expressed by a formula, then the circumscription in section 6.2 will continue to work under similar conditions.

8 Representing concurrent actions

Recall that the version of the situation calculus we use so far is a many-sorted first-order logic with the following domain independent sorts: situation sort (s), propositional fluent sort (p), and action sort (a). There is a domain independent function $Result(a,s)$, which represents the resulting situation when a is performed in s, and a domain independent predicate $H(p,s)$, which asserts that p holds in s.

It is not surprising that this version of the situation calculus cannot represent concurrent actions. We see that actions here are assumed to be primitive. Furthermore, it is clear that there is an asymmetry between actions and situations in this picture. While situations are 'rich' objects, as manifested by the various fluents that are true and false in them, actions are 'poor,' primitive objects. In this paper, we propose to model concurrent actions in the situation calculus by correcting this asymmetry. We introduce the notions of *global actions*, *primitive actions*, and the binary predicate In, whose roles will be completely analogous to those of situations, fluents, and the predicate H, respectively. Intuitively, a global action is

a set of primitive actions, and In expresses the membership relation between global actions and primitive actions. When a global action is performed in a situation, all of the primitive actions in it are assumed to be performed simultaneously.

Formally, the extended situation calculus is a many-sorted first-order logic with four domain-independent sorts: situation sort (s), propositional fluent sort (p), global action sort (g), and primitive action sort (a). We again have a binary function $Result$. But now it maps a *global* action and a situation into another situation. Intuitively, $Result(g, s)$ is the resulting situation of performing the primitive actions in g simultaneously in the situation s. In addition to H, we have another binary predicate In. Intuitively, $In(a, g)$ means that the primitive action a is an element of g.

For any finite set of primitive actions, $A_1, ..., A_n$, we assume that $\{A_1, ..., A_n\}$ is the unique global action satisfying the following properties:

$$\forall a. In(a, \{A_1, ..., A_n\}) \equiv (a = A_1 \vee \ldots \vee a = A_n), \tag{15}$$

and

$$\forall g. [\forall a. In(a, g) \equiv (a = A_1 \vee \ldots \vee a = A_n)] \supset g = \{A_1, ..., A_n\}. \tag{16}$$

If A is a primitive action, then we shall write $\{A\}$ as A. Most often, whether A is a primitive action or the corresponding global action $\{A\}$ will be clear from the context. For example, A is a global action in $Result(A, s)$, but a primitive action in $In(A, g)$.

To be sure, there are other proposals in the literature for extending the situation calculus to allow expressions for concurrent actions. Most introduce new operators on actions (cf. [4, 25]). For example, in [4] a new operator "+" is introduced, with the intuitive meaning that $a + b$ is the action of executing a and b simultaneously. This approach is common also in the programming languages community. The relationships between our formalism and those with new operators for concurrent actions are delicate. It is probably the case that they all have the same power of expressiveness, but some concurrent actions may be more easily formalizable in one than the other.

Another way to extend the situation calculus is to think of $Result$ as a relation, rather

than as a function. For example, we can introduce $RES(a, s_1, s_2)$ with the intuitive meaning that s_2 is one of the situations resulted from executing a, along with possibly some other actions, in s_1. This is essentially the approach taken in [5, 21], and others. The drawback of this approach is that it does not explicitly list the additional actions that cause the transition from s_1 to s_2. Thus Georgeff [5] had difficulty in formalizing the effect of a single action when performed exclusively.

We have been only recently aware of a closely related work by Weber [27]. In essense, what we call primitive actions are called action types in [27], and what we call global actions are simply actions in [27]. However, there are some significant differences in how the formalisms are used. Whereas Weber appeals directly to causal rules of (global) actions, as we shall see, the effects of the global actions here are often inherited from their sub-actions, in particular their primitive actions. Weber, however, investigates certain issues that are not addressed here. In particular, in our terms, he discusses how nonmonotonic logics can be used to formalize the default assumption that a global action contains only those primitive actions that are explicitly stated.

9 The frame problems

We notice here that all of the previous definitions and results about actions in the standard situation calculus can be extended straightforwardly to that about global actions. In the following, we shall use "action" and "global action" interchangeably. For instance, the new version of the epistemological completeness is:

Definition 9.1 *A theory $\mathcal{T}$ is* epistemologically complete *about the (global) action G (with respect to $\mathcal{P}$, and according to $\models_{\sqsubset}$) if $\mathcal{T} \not\models_{\sqsubset} False$, and for any ground situation term S, any state $\mathcal{S}$ of S, and any fluent $P \in \mathcal{P}$, there is a finite subset $\mathcal{S}'$ of $\mathcal{S}$ such that either $\mathcal{T} \models_{\sqsubset} \bigwedge \mathcal{S}' \supset H(P, Result(G, S))$ or $\mathcal{T} \models_{\sqsubset} \bigwedge \mathcal{S}' \supset \neg H(P, Result(G, S))$.*

In section 4 we analyzed the frame problem in terms of the notion of the epistemological completeness. For concurrent actions, a key problem is what we shall call the