

Provably Correct Theories of Action (preliminary report)

Fangzhen Lin
Yoav Shoham

Computer Science Department
Stanford University
lin@cs.stanford.edu
shoham@cs.stanford.edu

Abstract

Research on nonmonotonic temporal reasoning in general, and the Yale Shooting Problem in particular, has suffered from the absence of a criterion against which to evaluate solutions. Indeed, researchers in the area disagree not only on the solutions but also on the problems. We propose a formal yet intuitive criterion by which to evaluate theories of actions, define a monotonic class of theories that satisfy this criterion, and then provide their provably-correct nonmonotonic counterpart.

1 Introduction

The histories of the frame problem [McCarthy and Hayes 1969], and of the particular Yale Shooting Problem (YSP) which has become its best known illustration [Hanks and McDermott 1986], 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. The attempt in [Shoham and McDermott 1988] to analyze issues in nonmonotonic temporal reasoning in a principled fashion was itself informal, and did not provide precise definitions. This is an initial disturbing factor.

A second disturbing factor is that, despite a 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 [Loui 1987]. Simultaneously there were several

proposed solutions to it [Shoham 1986, Lifschitz 1986, Kautz 1986, Lifschitz 1987, Haugh 1987]. Then there were counter arguments that each of these solutions was 'wrong,' in that they didn't solve other problems such as the 'qualification problem' or the 'ramification problem,' or that they supported 'prediction' but not 'explanation.' New examples were then devised, with names such as the Stolen Car Problem [Kautz 1986] and the Stanford Murder Mystery [Baker 1989]. 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 [Morgenstern and Stein 1988, Lifschitz and Rabinov 1989, Baker and Ginsberg 1989, 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 article 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.

The article is structured as follows. In the following section we illustrate our approach through an analogy with data bases and closed-world assumption. We then start a technical development; after short logical preliminaries in section 3, in section 4 we define the epistemological adequacy of a (monotonic or non-monotonic) theory of action. In section 5 we illustrate the notion through the YSP. In section 6 we define a wide class of monotonic theories of action, and show

those to be epistemologically adequate. In section 7 we show that a version of circumscription captures a subclass of the theories. In section 8 we point to our future work and make some concluding remarks.

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, etcetera, do not change. The frame problem is the problem of representing concisely these numerous facts that are unaffected by an action.

Our approach in making the problem precise is conceptually very simple, and perhaps best illustrated by an analogy with simple data bases. Consider a data base of flight connections between pairs of cities. One way to structure the data base is by 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 data base. The semantics of this data base 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 so-called *closed world assumption (CWA)* [Reiter 1978], and the equivalent monotonic formulation can be regenerated from the abbreviated representation through *data base completion* [Clark 1978]. This concise representation is epistemologically complete since it too entails $Flight(A, B)$ or $\neg Flight(A, B)$ for all pairs of cities A and B , albeit nonmonotonically. Furthermore, the nonmonotonic version is epistemologically correct in a stronger sense: it is sound and complete relative to the monotonic version, since 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, a 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 situation calculus [McCarthy and Hayes 1969]. although we believe that a similar treatment is possible in other frameworks such as temporal logics. The standard situation calculus, which we adopt here, precludes the representation of certain notions such as concurrent actions. In future publications we will address those. 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 article it is important to be precise about the language.

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

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

2. Action sort: with action constants A_1, A_2, \dots , and action variables a, a_1, a_2, \dots . We will use A, A', \dots as meta-variables for action constants.

3. Propositional fluent sort: with fluent constants P_1, P_2, \dots , and fluent variables p, p_1, p_2, \dots . We will use P, P', \dots as meta-variables for fluent constants.

We have a binary function *result* whose first argument is of action sort, second argument of situation sort, and whose value is of situation sort. Thus for any action term A , any situation term S , $result(A, S)$ is a situation term. Intuitively, $result(A, S)$ is the resulting situation when the action A is performed in the situation S . We also have a binary predicate *holds* whose first argument is of fluent sort, and second argument of situation sort. Intuitively, $holds(P, S)$ means that the fluent P is true in the 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:

$$\forall s. (holds(P_1, s) \equiv \neg holds(P_1, result(toggle, s))) \quad (1)$$

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:

$$\forall s. (holds(P_2, s) \equiv holds(P_2, result(toggle, s))) \quad (2)$$

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 the above axioms indeed do completely formalize 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 T is a complete formalization 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 that of the *Frame* predicate in [Lifschitz 1990]. In the following, let \mathcal{P} be a fixed set of fluent constants.

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

$$SS = \{holds(P, S) | P \in \mathcal{P}'\} \cup \{\neg holds(P, S) | P \in \mathcal{P} - \mathcal{P}'\}$$

Therefore, if SS is a state of S , then for any $P \in \mathcal{P}$, either $holds(P, S) \in SS$ or $\neg holds(P, S) \in SS$. Intuitively, states completely characterize situations with respect to the fluents in \mathcal{P} . Thus we can define that a first-order theory T is epistemologically complete about the action A (with respect to \mathcal{P}) if it is consistent, and for any ground situation term S , any state SS of S , and any fluent $P \in \mathcal{P}$, either $T \cup SS \models holds(P, result(A, S))$ or $T \cup SS \models \neg holds(P, result(A, S))$, 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_{\square} , we can define when a theory is epistemologically complete about an action according to the entailment \models_{\square} :

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

We notice here that for any sets T , SS , and formula φ , $T \cup SS \models \varphi$ if there is a finite subset SS' of SS such

that $T \models \bigwedge SS' \supset \varphi$. Thus if we replace \models_{\square} in Definition 4.2 by classical entailment \models , we get the same definition we have earlier for monotonic first-order theories.

Referring back to the data base 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 (A, B) , whether $Flight(A, B)$ is true.

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

In the later sections of the article, we first identify a broad class of epistemologically complete monotonic theories of action. These monotonic theories may contain a large number of explicit frame axioms. Then for a class of the monotonic theories, we develop equivalent nonmonotonic ones that do not appeal to explicit frame axioms.

However, we first re-examine the Yale Shooting Problem [Hanks and McDermott 1986] 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 *Load* is performed, the gun is loaded, and if the gun is loaded, then after *Shoot* is performed, Fred is dead. Thus we have the following two 'causal rules':

$$\forall s. holds(Loaded, result(Load, s)) \quad (3)$$

$$\forall s (holds(Loaded, s) \supset holds(Dead, result(Shoot, s))) \quad (4)$$

This theory is of course insufficient to fully capture 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 the action *Shoot*, we have that for each $P \in \mathcal{P}$,

$$\forall s (\neg holds(Loaded, s) \supset (holds(P, s) \equiv holds(P, result(Shoot, s)))) \quad (5)$$

$$\forall s (holds(Loaded, s) \equiv holds(Loaded, result(Shoot, s))) \quad (6)$$

For the action *Load*, we have

$$\forall s (holds(Dead, s) \equiv holds(Dead, result(Load, s))) \quad (7)$$

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

$$\forall s(\text{holds}(P, s) \equiv \text{holds}(P, \text{result}(\text{Wait}, s))) \quad (8)$$

Let $T_1 = \{(3), (4), (5), (6), (7), (8)\}$. It is possible to show that the monotonic theory 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

$$T_1 \models \forall s.\text{holds}(\text{Dead}, \text{result}(\text{Shoot}, \text{Wait}, \text{Load}, s))$$

where $\text{result}(\text{Shoot}, \text{Wait}, \text{Load}, s)$ is

$$\text{result}(\text{Shoot}, \text{result}(\text{Wait}, \text{result}(\text{Load}, s)))$$

That is, *Dead* holds after the *Load*, *Wait*, and *Shoot* actions are performed in sequence in any situation. As an example of ‘temporal explanation’, we have

$$T_1 \models \forall s[\text{holds}(\text{Dead}, \text{result}(\text{Shoot}, s)) \wedge \neg \text{holds}(\text{Dead}, s) \supset \text{holds}(\text{Loaded}, s)]$$

That is, in any situation, if *Dead* is not true initially but becomes true after the *Shoot* action, then *Loaded* must be true initially.

5.2 Nonmonotonic completion

Although the above monotonic theory is epistemologically complete, it does not solve the frame problem since it contains the explicit frame axioms. We now provide an equivalent nonmonotonic theory that avoids them.

It was implied in [McCarthy 1986] that the frame axioms can be replaced by the single

$$\forall p a s (\neg ab(p, a, s) \supset (\text{holds}(p, s) \equiv \text{holds}(p, \text{result}(a, s)))) \quad (9)$$

and minimizing the abnormality predicate *ab* with *holds* allowed to vary. The main technical result of [Hanks and McDermott 1986] 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 [Shoham 1986], although other proposals, such as [Lifschitz 1986] and [Kautz 1986], would work as well.

For the full definition of chronological minimization the reader is referred to the above publication; 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

$$S < \text{result}(\text{Shoot}, S) < \dots$$

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

$$\text{Loaded} \neq \text{Dead} \neq \text{Shoot} \neq \text{Load} \neq \text{Wait} \quad (10)$$

We now simply take T_2 to be the conjunction of (3), (4), (9), and (10), and chronologically minimizing *ab* in T_2 . It is now possible to show that $T_1 \cup \{(10)\}$ and T_2 are equivalent: for any φ in the language of T_1 , $T_1 \cup \{(10)\} \models \varphi$ iff $T_2 \models_{\square} \varphi$, where \models is classical entailment and \models_{\square} is the nonmonotonic entailment. In particular, we have that the nonmonotonic theory T_2 is epistemologically complete.

We remark here that we have avoided formally claiming that 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 A class of complete causal theories

In this section we identify a class of monotonic causal theories that are epistemologically complete.

In reasoning about action, our knowledge can be generally divided into two kinds. One is about the environment where the actions are taken place, and is commonly referred to as domain constraints. The other is about actions themselves, and is usually called causal rules. Causal rules only tell us the direct effects of the actions. In different environments, an action may have different side effects. Side effects are determined by the direct effects and the domain constraints. A fact that is neither a direct effect nor a side effect of an action is assumed to be unchanged by the action. This motivates the following definition. Again in the following, we fix a set of fluents \mathcal{P} .

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

$$\forall s.C(s) \quad (11)$$

For each $1 \leq i \leq n$, the causal rules:

$$\forall s(R_i(s) \supset \text{holds}(P_i, \text{result}(A, s))) \quad (12)$$

For any subset M of $N = \{1, \dots, n\}$, and any $P \in \mathcal{P}$, if

$$\not\models \forall s(\bigwedge_{i \in M} \text{holds}(P_i, s) \wedge C(s) \supset \text{holds}(P, s))$$

and

$$\not\models \forall s(\bigwedge_{i \in M} \text{holds}(P_i, s) \wedge C(s) \supset \neg \text{holds}(P, s))$$

then the following frame axiom:

$$\forall s.(\bigwedge_{i \in N-M} \neg R_i(s) \supset (\text{holds}(P, s) \equiv \text{holds}(P, \text{result}(A, s)))) \quad (13)$$

We notice here that the above definition is for single action. The causal theory for a set of actions is the union of the causal theories for the actions in the set.

We now formulate a sufficient condition under which the monotonic causal theories are epistemologically complete.

Let T_0 consist of the domain constraint (11), and for each $1 \leq i \leq n$, the causal rule (12).

For any state SS of S , let

$$\begin{aligned} \text{result}(A, SS) = & \{ \text{holds}(P, S') \mid P \in \mathcal{P}, \\ & SS \cup T_0 \models \text{holds}(P, S') \} \\ \cup & \{ \neg \text{holds}(P, S') \mid P \in \mathcal{P}, \\ & SS \cup T_0 \models \neg \text{holds}(P, S') \} \\ \cup & \{ \text{holds}(P, S') \mid \text{holds}(P, S) \in SS, \\ & SS \cup T_0 \not\models \neg \text{holds}(P, S') \} \\ \cup & \{ \neg \text{holds}(P, S') \mid \neg \text{holds}(P, S) \in SS, \\ & SS \cup T_0 \not\models \text{holds}(P, S') \} \end{aligned}$$

where $S' = \text{result}(A, S)$. We see that $\text{result}(A, SS)$ is a state of S' if $SS \cup T_0$ is consistent.

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

Condition 1. $C(s), R_1(s), \dots, R_n(s)$ do not contain any situation term other than s .

Condition 2. For any state SS of S , either $SS \models \varphi(S)$ or $SS \models \neg \varphi(S)$, where $\varphi(S) \in \{C(S), R_1(S), \dots, R_n(S)\}$.

Condition 3. $\forall s. C(s)$ is consistent.

Condition 4. For any SS of S , if $SS \models C(S)$, then $\text{result}(A, SS)$ is consistent, and $\text{result}(A, SS) \models C(\text{result}(A, S))$.

In most cases, Conditions 1 to 3 are easy to check. Condition 4 is the only one that is nontrivial.

Although Theorem 1 is about a single action, it is easily extendable to multiple actions. This is because that the five conditions guarantee that all of the actions must be independent. For example, Condition 1 excludes any action terms from appearing in $C(s)$, $R_1(s)$, \dots , and $R_n(s)$.

The class of the causal theories that satisfy the five conditions includes most of the blocks world examples found in the literature. If we ignore the predicates *frame* and *possible* in [Lifschitz 1990a], then our class includes the causal theories in the main theorem of [Lifschitz 1990a].

7 Capturing the causal theories in circumscription

We now proceed to see how to have a succinct representation for the frame axioms in the causal theories. We

shall use a simple version of circumscription. We expect that other formalisms and theories like chronological ignorance [Shoham 1986], pointwise circumscription [Lifschitz 1986], and Baker's method [Baker 1989] will also work.

Again, we fix a set of fluents \mathcal{P} . We assume that $p \in \mathcal{P}$ can be formalized by a first-order formula. Formally, we assume that $\text{Frame}(p)$ is a formula with a free variable p such that for any interpretation M , and any P^* in the fluent domain of M , $M \models \text{Frame}(P^*)$ iff there is a $P \in \mathcal{P}$ such that P is interpreted as P^* . For example, if $\mathcal{P} = \{P_1, P_2\}$, then $\text{Frame}(p)$ can be $p = P_1 \vee p = P_2$.

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

$$\text{Frame}(p) \wedge (\text{holds}(p, s) \equiv \neg \text{holds}(p, \text{result}(a, s)))$$

Our circumscriptive policy will be that for any given situation S and any give action A , we minimize $ab(p; S; A)$ as a unary formula of p with $\text{holds}(p, S)$ fixed but $\text{holds}(p, S')$ allowed to vary for every S' that is different from S . In order to do that, we extend our language to include a new predicate holds' that is similar to holds . Then for any formula W , we minimize ab (as a unary formula of p) in the following formula with holds' fixed and holds allowed to vary:

$$\forall p (\text{holds}(p, s) \equiv \text{holds}'(p, s)) \wedge W$$

Also, in order to use circumscription, we need to have some unique names axioms. We suppose there is an axiom U_1 that captures the unique names assumption for fluents in \mathcal{P} . We also suppose U_2 is the following axiom:

$$\begin{aligned} & \forall s. \text{earlier}(s, \text{result}(a, s)) \wedge \\ & \forall s_1 s_2 (\text{earlier}(s, s_1) \wedge \text{earlier}(s_1, s_2) \supset \text{earlier}(s, s_2)) \\ & \wedge \forall s_1 (\text{earlier}(s, s_1) \supset s \neq s_1) \end{aligned}$$

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

$$S \neq \text{result}(A, S) \neq \text{result}(A, \text{result}(A, S)) \neq \dots$$

Now we can have the following result

Theorem 2 *Under the assumptions and conditions in Theorem 1, for any formula φ in \mathcal{L} (our original language without holds'), $T \cup \{U_1, U_2\} \models \varphi$ iff $\forall s. \text{Circum}(W; ab(p; s; A); \text{holds}) \models \varphi$, where $\text{Circum}(W; ab; \text{holds})$ is the circumscription of ab in W with holds allowed to vary, W is the following formula*

$$\forall p (\text{holds}(p, s) \equiv \text{holds}'(p, s)) \wedge U_1 \wedge U_2 \wedge (\bigwedge T_0)$$

and T_0 consists of the domain constraint (11) and for each $1 \leq i \leq n$, the causal rule (12).

Thus if we define \models_{\square} such that $\{W\} \models_{\square} \varphi$ iff $\forall s. \text{Circum}(W; ab(p; s; A); \text{holds}) \models \varphi$, then we have the following corollary

Corollary 2.1 *Under the assumptions in Theorem 2, if W is classically consistent, then $\{W\}$ is an epistemological complete nonmonotonic theory about A according to \models_{\square} .*

8 Future work and concluding remarks

We have argued that a useful way to tackle the frame problem is to consider a monotonic theory with explicit frame axioms first, and then to show that a succinct and provably equivalent representation using, for example, nonmonotonic logics, captures the frame axioms concisely. The idea is not startling. A similar idea is used in verifying negation-as-failure [Clark 1978]. It seems that several researchers have entertained the idea of verifying a nonmonotonic theory of action against a monotonic one (for example, it is listed as future work in [Winslett 1988]), even if no one to date has actually followed this course. Central to our project is the definition of an epistemological completeness condition for deterministic actions, whose intuitive purpose is to determine whether sufficient axioms are included in a theory. Our main technical contribution is in formulating a wide class of epistemologically complete monotonic causal theories, and showing that for each of causal theories, there is a succinct representation using a version of circumscription.

We notice here that our reformulation of the causal theories in circumscription does not address the qualification problem. This can be easily done by using the method in [Lifschitz 1987].

There are many directions for future work. At the present time, the most important one is to extend our work to allow concurrent actions. Just as fluents can partially describe a situation, we may have action descriptions that partially describe the set of actions taken in any situation. As a result we will be able to infer by default not only what is true in situations, but also what actions have been taken; in the current framework that is not even expressible.

Acknowledgement

We would like to thank Vladimir Lifschitz for stimulating discussions related to the subject of this paper.

References

- [1] Baker, A. B. (1989), A simple solution to the Yale shooting problem, in *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, 1989.
- [2] Baker, A. B. and M. L. Ginsberg (1989), Temporal projection and explanation, in *Proceedings of IJCAI-89*, 1989.
- [3] Clark, K. (1978), Negation as failure, in *Logics and Databases*, eds. by Gallaire and Minker, Plenum Press, 293-322, 1978.
- [4] Hanks, S. and D. McDermott (1986), Nonmonotonic logics, and temporal projection, *Artificial Intelligence*, 33 (1987), 379-412.
- [5] Haugh, B. (1987) Simple causal minimizations for temporal persistence and projection, in *Conference Proceedings of AAAI-87*.
- [6] Kautz, H. (1986), The logic of persistence, in *Proceedings of AAAI-1986*.
- [7] Lifschitz, V. (1986), Pointwise circumscription, in *AAAI - 1986*.
- [8] Lifschitz, V. (1987), Formal theories of action, *IJCAI-87*.
- [9] Lifschitz, V. (1990), Frames in the Space of Situations, *Artificial Intelligence*, 46 (1990) 365-376.
- [10] Lifschitz, V. (1990a), Toward a Meta-theory of Action, Draft, 1990.
- [11] Lifschitz, V. and A. Rabinov (1989), Miracles in formal theories of action, *Artificial Intelligence*, 38 (1989) 225-237.
- [12] Loui, R. (1987), Response to Hanks and McDermott: temporal evolution of beliefs and beliefs about temporal evolution, *Cognitive Science* 11 (1987).
- [13] McCarthy, J. (1986), Applications of circumscription to formalizing commonsense knowledge, *Artificial Intelligence* 28 (1986), 89-118.
- [14] McCarthy, J. and P. Hayes (1969), Some philosophical problems from the standpoint of artificial intelligence. In *Machine Intelligence 4*, Meltzer, B and Michie, D. (eds), Edinburgh University Press.
- [15] Morgenstern, L. and L. A. Stein (1988), Why things go wrong: A formal theory of causal reasoning, in *Proceedings of AAAI-1988*.
- [16] Reiter, R. (1978), On closed world data bases, in H. Gallaire and J. Minker (eds.), *Logics and Data Bases*, Plenum Press, New York, 1978.
- [17] Shoham, Y. (1986), Chronological ignorance: time, nonmonotonicity and necessity. In *Proceedings of AAAI-86*.
- [18] Shoham, Y. and D. McDermott (1988), Problems in formal temporal reasoning, *Artificial Intelligence*, 36 (1988) 49-62.
- [19] Winslett, M. (1988), Reasoning about actions using a possible models approach, in *Conference Proceedings of AAAI-88*.