

# Nonmonotonic Reasoning in the Framework of Situation Calculus

Andrew B. Baker  
Department of Computer Science  
Stanford University  
Stanford, CA 94305

**Abbreviated title: Nonmonotonic Reasoning in Situation Calculus**

## Abstract

Most of the solutions proposed to the Yale shooting problem have either introduced new nonmonotonic reasoning methods (generally involving temporal priorities) or completely reformulated the domain axioms to represent causality explicitly. This paper presents a new solution based on the idea that since the abnormality predicate takes a situational argument, it is important for the meanings of the situations to be held constant across the various models being compared. This is accomplished by a simple change in circumscription policy: when  $Ab$  is circumscribed, *Result* (rather than *Holds*) is allowed to vary. In addition, we need an axiom ensuring that every consistent situation is included in the domain of discourse. Ordinary circumscription will then produce the intuitively correct answer. Beyond its conceptual simplicity, the solution proposed here has additional advantages over the previous approaches. Unlike the approach that uses temporal priorities, it can support reasoning backward in time as well as forward. And unlike the causal approach, it can handle ramifications in a natural manner.

# 1 Introduction

The formalization of reasoning about change has proven to be a surprisingly difficult problem. Standard logics are inadequate for this task because of difficulties such as the frame problem [14]; nonmonotonic reasoning seems to be necessary. Unfortunately, as demonstrated by the Yale shooting problem of Hanks and McDermott [5], the straightforward use of standard nonmonotonic logics (such as circumscription) for reasoning about action leads to counter-intuitive results.

There have been a large number of solutions proposed to the shooting problem [2, 6, 7, 9, 10, 15, 17, 20, and others], but none of them are completely satisfactory. Some of these solutions cannot handle examples that require reasoning backward in time. And others require that the domain axioms be written in a rather restrictive format. This paper presents a new approach to the shooting problem that avoids these difficulties.

In the next section, we describe the shooting problem. Section 3 surveys some of the previous solutions and their limitations. Section 4 presents our solution, albeit in a slightly simplified form; this solution is refined in Section 5. In Section 6, we consider two additional scenarios in order to compare the various approaches to nonmonotonic temporal reasoning. Concluding remarks are contained in Section 7.

This paper is a revised version of [1].

## 2 The shooting problem

The Yale shooting problem arises regardless of which temporal formalism is used; we will use the situation calculus [14]. A *situation* is the state of the world at a particular time. Given an *action*  $a$  and a situation  $s$ ,  $Result(a, s)$  denotes the new situation after action  $a$  is performed in situation  $s$ . A truth-valued *fluent* (the only kind of fluent that will concern us) is a property that may or may not hold in a given situation. If  $f$  is a fluent and  $s$  is a situation,

then  $Holds(f, s)$  means that the fluent  $f$  is true in situation  $s$ .

With these conventions, one might use standard first-order logic to formalize the effects of various actions, but there are some well-known problems with this monotonic approach. The *frame problem* [14], to which this paper will be limited, is that we would need to write down a great many axioms specifying those properties that are *unchanged* by each action. And yet, intuitively, all of these frame axioms seem redundant; we would like to specify just the positive effects of an action, and then somehow say that nothing else changes. Part of the motivation behind the development of nonmonotonic reasoning was to formalize this notion, and thus to solve the frame problem; we will use McCarthy’s *circumscription* [12, 13]. If  $\mathcal{A}$  is a formula,  $P$  is a predicate, and  $Z$  is a tuple of predicates and functions, then the circumscription of  $P$  in  $\mathcal{A}$  with  $Z$  varied is written as  $Circum(\mathcal{A}, P, Z)$ . This abbreviates a formula in second-order logic that selects those models of  $\mathcal{A}$  in which the extension of the predicate  $P$  is minimal (in the set inclusion sense). Besides  $P$ , only those predicates and functions in  $Z$  are allowed to vary during this minimization process. For a more extensive discussion of circumscription, the reader is referred to [8].

Consider the standard default frame axiom:<sup>1</sup>

$$\neg Ab(f, a, s) \supset ( Holds(f, Result(a, s)) \equiv Holds(f, s)). \quad (1)$$

This says that the value of a fluent persists from one situation to the next unless something is abnormal. The original intention was to circumscribe  $Ab$  with  $Holds$  varied. (We will refer to this as the *standard* circumscription policy.) It was hoped that this minimization of abnormality would ensure that a fluent would persist unless a specific axiom forced this fluent to change.

Unfortunately, this approach does not work. In a sequence of events, often one can eliminate an expected abnormality at one time by introducing a totally gratuitous abnormality at another time. In this case, there will be multiple minimal models not all of which will correspond to our intuitions.

---

<sup>1</sup>Lower case letters represent variables. Unbound variables are universally quantified.

The standard example, of course, is Hanks and McDermott’s Yale shooting problem [5].<sup>2</sup> In this problem, which we are simplifying slightly from the Hanks and McDermott version, there are two fluents, *Alive* and *Loaded*, and two actions, *Wait* and *Shoot*. The story is that if the gun is shot while it is loaded, a person (named Fred) dies:

$$\text{Holds}(\text{Loaded}, s) \supset \neg \text{Holds}(\text{Alive}, \text{Result}(\text{Shoot}, s)). \quad (2)$$

There are no axioms about *Wait*, so one would hope that the general-purpose frame axiom would ensure that *Wait* does not change anything. In the original situation, Fred is alive, and the gun is loaded:

$$\text{Holds}(\text{Alive}, S0), \quad (3)$$

$$\text{Holds}(\text{Loaded}, S0). \quad (4)$$

If the actions *Wait* and then *Shoot* are performed in succession, what happens?

Let  $Y_0$  be the conjunction of axioms (1)-(4), and let  $\overline{Y_0}$  be the circumscription of Ab in  $Y_0$  with *Holds* varied:

$$\overline{Y_0} \equiv \text{Circum}(Y_0; \text{Ab}; \text{Holds}).$$

What does  $\overline{Y_0}$  have to say about the truth value of

$$\text{Holds}(\text{Alive}, \text{Result}(\text{Shoot}, \text{Result}(\text{Wait}, S0)))?$$

We might guess that the waiting has no effect, and thus the shooting kills Fred, but circumscription is not so cooperative. Another possibility according to circumscription is that the gun mysteriously becomes unloaded while waiting, and Fred survives. This second model contains an abnormality during the waiting that was not present in the first model, but there is no longer an abnormality during the shooting (since the gun is unloaded, Fred does

---

<sup>2</sup>A similar scenario, involving the qualification problem rather than the frame problem, was discovered independently by Lifschitz and reported by McCarthy [13, page 107].

not change from *Alive* to not *Alive*). So both models are minimal, and the formalization must be altered in some way to rule out the anomalous model.<sup>3</sup>

### 3 Previous approaches

There have been a large number of solutions proposed to the shooting problem. This section discusses the two most popular groups of solutions.

#### 3.1 Chronological minimization

One idea, proposed in various forms by Kautz [7], Lifschitz [10], and Shoham [20], is *chronological minimization* (the term is due to Shoham). This proposal claims that we should reason forward in time; that is, apply the default assumptions in temporal order. So in the shooting scenario, we should first conclude that the waiting action is not abnormal. Then, since the gun would remain loaded, we would conclude that Fred dies. Each of the above authors successfully constructs a nonmonotonic logic that captures this notion of chronological minimality. Kautz, for example, uses a modified version of circumscription in which abnormalities at earlier times are minimized at a higher priority than those at later times.

While this approach does in fact give the intuitively correct answer to the Yale shooting problem, it is nevertheless highly problematic. Its applicability seems to be limited to what Hanks and McDermott call *temporal projection* problems, or in other words, problems in which given the initial conditions, we are asked to predict what will be the case at a later time. One can also consider temporal *explanation* problems [5], i.e., problems requiring reasoning backward in time. For problems of this sort, chronological minimization generally does not work very well (see the example in Section 6.1). For this reason, chronological minimization is not a completely satisfactory solution.<sup>4</sup>

---

<sup>3</sup>The current formalization admits a third possibility: Fred might die during the waiting phase. Our solution will rule out this model also.

<sup>4</sup>Sandewall [19] has proposed *filter preferential entailment*, a modification of chrono-

## 3.2 Causal minimization

Another approach, developed by Haugh [6] and by Lifschitz [9], is that of *causal minimization*. This method represents causality explicitly by stating that a fluent changes its value if and only if a successful action causes it to do so. The intuition here is that there is a crucial difference between the abnormality of a gun becoming unloaded while waiting, and the abnormality of Fred dying when shot with a loaded gun: there is a *cause* for the second while the first is totally arbitrary. We will discuss the system of [9]. There, the effects of actions are represented with a predicate  $Causes(a, f, v)$  that indicates that if the action  $a$  is successful,<sup>5</sup> then the fluent  $f$  takes on the value  $v$ . With this formalism, one specifies all the known causal rules (for the current example,  $Causes(Shoot, Alive, False)$ ), and then circumscribes  $Causes$  with  $Holds$  varied. Since  $Causes$  does not take a situational argument, there obviously cannot be a conflict in minimizing it in different situations. Therefore, the shooting problem cannot arise.

The main drawback of this proposal is that it does not allow us to write our domain axioms in unrestricted situation calculus. Instead, we must use the  $Causes$  predicate. This is a severe restriction on our expressive power because there is simply no way to use the  $Causes$  predicate to express general context-dependent effects, domain constraints, or ramifications (see Section 6.2 for a discussion of this issue). In light of this difficulty, it would be useful to solve the Yale shooting problem in the original formalism.

## 4 The solution

Our approach to the Yale shooting problem consists of two innovations. First of all, when we circumscribe  $Ab$ , instead of letting the  $Holds$  predi-

---

logical minimization that avoids some of the difficulties of the ordinary version.

<sup>5</sup>An action is successful if all of its preconditions are satisfied. Lifschitz [9] formalizes preconditions with a special  $Precond$  predicate; for the current paper, the only precondition would be that  $Loaded$  must hold for  $Shoot$  to succeed.

cate vary, we will let the *Result* function vary. That is, we will not think of  $Result(Wait, S0)$  as being a fixed situation, with circumscription being used to determine which fluents hold in this situation. Instead, we will assume that for each combination of fluents, there is some possible situation in which these fluents hold. Circumscription will then be used to determine which of these situations might be the result of waiting in  $S0$ . Second, in order for this idea to succeed, we must already have every consistent situation in the domain of discourse; an axiom will be added to accomplish this.

To see why this approach makes sense, let us resist the temptation of appealing to causality or temporal priorities, and instead think about the problem in its own terms. Why is it that we prefer the model in which Fred dies to the one in which waiting unloads the gun? After all, if by making the world behave more abnormally in  $S0$ , we could make it behave less abnormally somewhere else, this would seem to be a fair trade. The problem is that if waiting unloads the gun, then the resultant situation is really a *different* situation than it would have been, so it is not the case that the world behaves more normally somewhere else. In our preferred model, the abnormality was that *Alive* changed to not *Alive* when *Shoot* was performed in a situation in which the gun was loaded. In the anomalous model, the world does not behave more normally in this situation; this situation just does not come about! But the standard circumscription policy, with *Holds* varied and *Result* fixed, completely misses this subtlety. It views the  $Result(Wait, S0)$ 's in the two models as the same situation, even though different fluents hold in them.

From this perspective, the shooting problem arises from the failure to index abnormality correctly. Since the abnormality predicate *Ab* takes a situational argument, it is important for the meanings of the situations to be held constant across the various models being compared. We can accomplish this by varying *Result* instead of *Holds* during the circumscription.

First, some details need to be discussed. We will use a many-sorted language with object variables of the following sorts: for situations  $(s, s_1, s_2, \dots)$ , for actions  $(a, a_1, a_2, \dots)$ , and for fluents  $(f, f_1, f_2, \dots)$ . We have the situation

constant,  $S0$ , the action constants,  $Wait$  and  $Shoot$ , and the fluent constants,  $Alive$  and  $Loaded$ . Finally, we have the predicate constants  $Holds(f, s)$  and  $Ab(f, a, s)$  and the function constant  $Result(a, s):s$ . Axioms (1)–(4) should be interpreted relative to these declarations.

We need a uniqueness of names axiom for actions and also a domain-closure axiom for fluents:

$$Wait \neq Shoot, \tag{5}$$

$$f = Alive \vee f = Loaded. \tag{6}$$

Now, for the important part. Suppose that for every possible combination of fluents, there is some situation in which these fluents hold. We will discuss how to achieve this in general in the next section, but for the Yale shooting problem we add the following *existence of situations* axiom:

$$\begin{aligned} & \exists s(Holds(Alive, s) \wedge Holds(Loaded, s)) \\ \wedge & \exists s(Holds(Alive, s) \wedge \neg Holds(Loaded, s)) \\ \wedge & \exists s(\neg Holds(Alive, s) \wedge Holds(Loaded, s)) \\ \wedge & \exists s(\neg Holds(Alive, s) \wedge \neg Holds(Loaded, s)). \end{aligned} \tag{7}$$

This axiom ensures that for each of the four possible fluent combinations, there will be *at least* one corresponding situation; there may in fact be more than one such situation. We could add an axiom saying that there must be exactly one such situation (for each fluent combination); in this case, the models would be deterministic finite automata with four states. Alternatively, we could introduce the notion of time, and require that each fluent combination have a corresponding situation for each time point, and that actions always increment the time. We will not make either one of these additions since neither would affect any of the results to be presented. We will, however, use the second convention in illustrating the ideas in Figures 1 and 2.

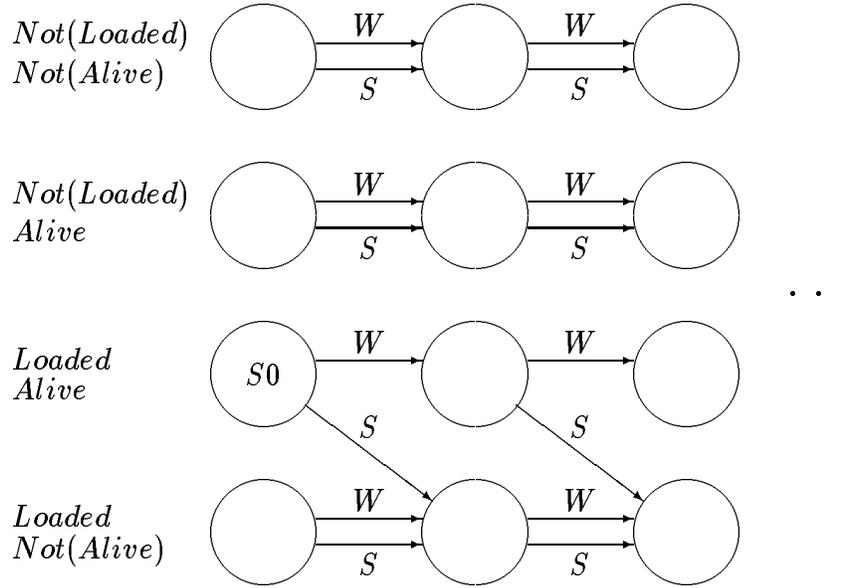


Figure 1: A minimal model of the Yale shooting problem

Let  $Y_1$  be the conjunction of axioms (1)–(7), and circumscribe  $Ab$  in  $Y_1$  with  $Result$  varied:

$$\overline{Y_1} \equiv \text{Circum}(Y_1; Ab; Result).$$

In other words, the models of  $\overline{Y_1}$  are precisely the minimal models (with respect to the above circumscription policy) of  $Y_1$ . Figures 1 and 2 are pictorial representations of two of the models of  $Y_1$ . Time flows horizontally, and each circle represents the set of situations at a given time in which the fluents to its left either hold or do not hold as indicated. The  $W$  arrows show the result of performing the *Wait* action, and the  $S$  arrows show the result of performing the *Shoot* action. Diagonal arrows represent actions that change at least one fluent, and hence are associated with at least one abnormality.

Since the  $S$  arrows from situations in which *Loaded* and *Alive* hold lead to situations in which *Alive* does not hold, these arrows must be diagonal.

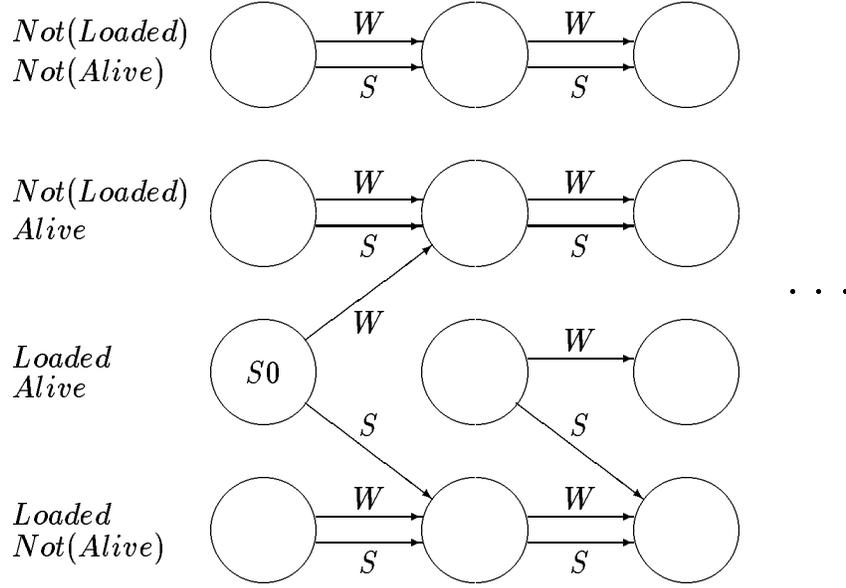


Figure 2: A nonminimal model of the Yale shooting problem

In the model of Figure 1, these are the *only* abnormalities, and therefore this is a minimal model of  $Y_1$  and hence a model of  $\overline{Y_1}$ . Figure 1 represents an intuitive model of the shooting problem. Figure 2, on the other hand, represents an unexpected model: one in which waiting in  $S0$  unloads the gun. Note, however, that this added abnormality does not reduce the abnormality anywhere else. This model is strictly inferior to the previous one, and so it is not a model of  $\overline{Y_1}$ . Therefore, our technique solves the Yale shooting problem.

To state the correctness formally, we will use the following criterion:

$$\begin{aligned}
 & [Holds(f, Result(a, s)) \equiv Holds(f, s)] \equiv \\
 & \neg[f = Alive \wedge a = Shoot \wedge Holds(Alive, s) \wedge Holds(Loaded, s)]. \quad (8)
 \end{aligned}$$

This states that the only fluent that changes its value is *Alive*, and this only happens when the *Shoot* action is performed in a situation in which the gun

is loaded and Fred is alive. Along with axiom (5), (8) immediately implies that *Wait* has no effect, and this, together with the initial conditions (3)–(4), ensures that Fred dies:

$$\neg \text{Holds}(\text{Alive}, \text{Result}(\text{Shoot}, \text{Result}(\text{Wait}, S0))).$$

**Proposition 1**  $\overline{Y}_1$  is consistent.

**Proof.** A model of  $\overline{Y}_1$  can be defined in a straightforward manner.  $\square$

**Proposition 2**  $\overline{Y}_1 \models (8)$

**Proof.** Let  $M$  be any model of  $\overline{Y}_1$ . Such a model consists of the various domains: the domain of situations  $|M|_s$ , the domain of actions  $|M|_a$ , and the domain of fluents  $|M|_f$ ; as well as interpretations for the constants:

$$\begin{aligned} M[\![S0]\!] &\in |M|_s, \\ M[\![Wait]\!] &\in |M|_a, \quad M[\![Shoot]\!] \in |M|_a, \\ M[\![Alive]\!] &\in |M|_f, \quad M[\![Loaded]\!] \in |M|_f, \end{aligned}$$

interpretations for the relations:

$$\begin{aligned} M[\![Holds]\!] &\subseteq |M|_f \times |M|_s, \\ M[\![Ab]\!] &\subseteq |M|_f \times |M|_a \times |M|_s, \end{aligned}$$

and an interpretation for the *Result* function:

$$M[\![Result]\!] \in (|M|_a \times |M|_s \rightarrow |M|_s).$$

We want to show that  $M$  must satisfy (8). Uniqueness of names for fluents follows from (7); this, together with the domain-closure axiom (6), ensures that without loss of generality, we can take:

$$|M|_f = \{\text{Alive}, \text{Loaded}\},$$

with *Alive* and *Loaded* interpreted as themselves. We can classify situations into four kinds with respect to which fluents hold in them, and if we let  $M[[Shoot]] = Shoot$ , we can classify actions into two kinds with respect to whether they are equal or not to *Shoot*. So there are eight cases that need to be considered.

Let  $S_1 \in |M|_s$  be any situation in which *Alive* and *Loaded* hold:

$$\begin{aligned} \langle Alive, S_1 \rangle &\in M[[Holds]], \\ \langle Loaded, S_1 \rangle &\in M[[Holds]], \end{aligned}$$

and let  $S_2$  be the result of performing *Shoot* in  $S_1$ :

$$M[[Result]](Shoot, S_1) = S_2.$$

It follows from (2) that *Alive* does not hold in  $S_2$ ; the only question is whether *Loaded* holds (intuitively, it should since we have not axiomatized the notion of running out of bullets). Suppose it does not. Then, from the frame axiom (1), we would have the abnormality:

$$\langle Loaded, Shoot, S_1 \rangle \in M[[Ab]]. \quad (9)$$

The existence of situations axiom (7) ensures that there is some situation  $S_3$  in which *Loaded* holds, but *Alive* does not. But then we would be able to further minimize abnormality by making  $S_3$  the result of the action, i.e., we could define another model  $M'$  of  $Y_1$  that is exactly like  $M$  except that:

$$\begin{aligned} M'[[Result]](Shoot, S_1) &= S_3, \\ M'[[Ab]] &= M[[Ab]] - \langle Loaded, Shoot, S_1 \rangle. \end{aligned}$$

This would eliminate abnormality (9) without introducing any new abnormalities (since there are only two fluents). Therefore  $M$  could not have been a minimal model of  $Y_1$ , contradicting the assumption that it was a model of  $\overline{Y_1}$ .

The arguments for the other seven cases are analogous. In each of these other cases, the action does not change the value of *any* of the fluents.  $\square$

The domain-closure for fluents axiom (6) plays a crucial role in the above argument. Suppose that there were a third fluent in addition to *Alive* and *Loaded*, and that this fluent held in situations  $S_1$  and  $S_2$ , but not in  $S_3$ . Then the above argument would fail. When we varied *Result* to go to  $S_3$  instead of  $S_2$ , we would be introducing an additional abnormality on this new fluent. The problem is that the existence of situations axiom (7) only ensures the existence of situations corresponding to the four combinations of *Alive* and *Loaded* — not to the eight combinations of *Alive*, *Loaded*, and the mystery fluent. Axiom (6) was not included in the preliminary version of this paper [1]; Lifschitz [11] pointed out that it was necessary.

## 5 Existence of situations

In the last section, we violated the spirit of the nonmonotonic enterprise by adding an existence of situations axiom (7) that explicitly enumerated all possible situations. For more complicated problems, this axiom would be a bit unwieldy. In this section, we show how to write this axiom in a more general way. We present one formalization using first-order logic, and then another using second-order logic.

### 5.1 First-order formalization

In order to talk about combinations of fluents, we will add a new sort to the language: generalized fluents (which are a superset of fluents), with the variables  $g, g_1, g_2, \dots$ ; and the functions  $And(g, g):g$  and  $Not(g):g$  to build up these generalized fluents. We will also alter the declaration  $Holds(f, s)$  to  $Holds(g, s)$  so that the first argument can be any generalized fluent. The following axioms define  $And$  and  $Not$ :

$$Holds(And(g_1, g_2), s) \equiv Holds(g_1, s) \wedge Holds(g_2, s), \quad (10)$$

$$\text{Holds}(\text{Not}(g), s) \equiv \neg \text{Holds}(g, s). \quad (11)$$

Generalized fluents will be used to reformalize the existence of situations axiom. Note that the frame axiom (1) is *not* modified to apply to all generalized fluents; it applies only to “regular” fluents (*Alive* and *Loaded*).

For our existence of situations axiom, we would like to introduce a function  $\text{Sit}(g):s$  that maps a generalized fluent into some situation in which the fluent holds:

$$\text{Holds}(g, \text{Sit}(g)).$$

This would ensure that there is some situation such that  $\text{And}(\text{Alive}, \text{Loaded})$  holds, and one such that  $\text{And}(\text{Alive}, \text{Not}(\text{Loaded}))$  holds and so on. But this would also mean that even inconsistent fluents like  $\text{And}(\text{Alive}, \text{Not}(\text{Alive}))$  would be true in some situation; this contradicts (10) and (11). More generally, any domain constraint will render certain fluent combinations inconsistent (in Section 6.2, we discuss an example with the fluent *Walking* and with the constraint that one has to be alive to be walking). So instead, the existence of situations axiom will be written as a default using a new abnormality predicate  $\text{Absit}(g)$ :

$$\neg \text{Absit}(g) \supset \text{Holds}(g, \text{Sit}(g)). \quad (12)$$

We will circumscribe  $\text{Absit}$ .

Axiom (12) can be viewed as a closed-world assumption for domain constraints. We know from the other axioms that there cannot be a situation in which both *Alive* and  $\text{Not}(\text{Alive})$  hold. Suppose, however, that there were also no situation in which both *Loaded* and  $\text{Not}(\text{Alive})$  held. Since we are thinking of the situations as possible states of the world, this would be like saying that there is an extra domain constraint that we do not know about; axiom (12) excludes this possibility.

In order for the circumscription of  $\text{Absit}$  to have its intended effect, some uniqueness of names axioms will be necessary. We will use an abbreviation

from [9]:  $\text{UNA}[f_1, \dots, f_n]$ , where  $f_1, \dots, f_n$  are (possibly 0-ary) functions, stands for the axioms:

$$f_i(x_1, \dots, x_k) \neq f_j(y_1, \dots, y_l)$$

for  $i < j$  where  $f_i$  has arity  $k$  and  $f_j$  has arity  $l$ , and:

$$f_i(x_1, \dots, x_k) = f_i(y_1, \dots, y_k) \supset (x_1 = y_1 \wedge \dots \wedge x_k = y_k)$$

for  $f_i$  of arity  $k > 0$ . These axioms ensure that  $f_1, \dots, f_n$  are injections with disjoint ranges. We state that uniqueness of names applies to generalized fluents, and to our special *Sit* function:

$$\text{UNA}[\textit{Alive}, \textit{Loaded}, \textit{And}, \textit{Not}], \quad (13)$$

$$\text{UNA}[\textit{Sit}]. \quad (14)$$

Let  $Y_2$  be the conjunction of axioms (1)–(6) and (10)–(14). In addition to circumscribing *Ab* as before, we now circumscribe *Absit* with *Ab*, *Holds*, *Result*, and *S0* allowed to vary:

$$\overline{Y_2} \equiv \text{Circum}(Y_2; \textit{Absit}; \textit{Ab}, \textit{Holds}, \textit{Result}, \textit{S0}) \wedge \text{Circum}(Y_2; \textit{Ab}; \textit{Result}).$$

**Proposition 3**  $\overline{Y_2}$  is consistent.

**Proof.** A model of  $\overline{Y_2}$  can be defined in a straightforward manner.  $\square$

**Proposition 4**  $\overline{Y_2} \models (8)$

**Proof.** Let  $M$  be any model of  $\overline{Y_2}$ . Consider the following four generalized fluents:

$$\begin{aligned} &\textit{And}(\textit{Alive}, \textit{Loaded}), & \textit{And}(\textit{Alive}, \textit{Not}(\textit{Loaded})), \\ &\textit{And}(\textit{Not}(\textit{Alive}), \textit{Loaded}), & \textit{And}(\textit{Not}(\textit{Alive}), \textit{Not}(\textit{Loaded})). \end{aligned} \quad (15)$$

Axiom (13) ensures that these are all unequal, and axiom (14) ensures that the situations formed by applying *Sit* to them are also unequal. Suppose

that one of these four generalized fluents were “abnormal” (with respect to *Absit*); that is, suppose for some  $g$  in (15), we had:

$$M[[g]] \in M[[Absit]].$$

This cannot happen, however, because it would violate the first circumscription: we would be able to define a model  $M'$  that is preferred to  $M$  by modifying the interpretation of *Absit* such that:

$$M'[[g]] \notin M'[[Absit]],$$

for all  $g$  in (15), and then varying *Ab*,  *Holds*,  *Result*, and *S0* in order to maintain consistency with the various axioms in  $Y_2$ . (These predicates and functions have to be allowed to vary during the circumscription of *Absit* for the following reason. When a generalized fluent  $g$  is not *Absit*, axioms (10)–(12) will tell us which fluents must hold in *Sit*( $g$ ). Then, the law of change (2) might place some constraints on the *Result* function for actions performed in *Sit*( $g$ ), and hence the frame axiom (1) might place some constraints on the *Ab* predicate. Finally, the initial conditions (3)–(4) might require that  $M'[[S0]] \neq M[[S0]]$ .)

Therefore by (12), it follows that  $|M|_s$  contains situations for each of the four fluent combinations. The proposition then follows from the same argument used in the proof of Proposition 2.  $\square$

## 5.2 Second-order formalization

The above approach for ensuring the existence of situations only works correctly for a finite number of fluents. If we had some fluent *Interesting*( $x$ ), for instance, where  $x$  could range over integers, (12) would ensure the existence of a situation in which

$$And(Interesting(0), Interesting(1))$$

held, but with only finite conjunctions, it would not ensure the existence of a situation in which all integers were interesting. We can fix this by writing an existence of situations axiom in second-order logic:

$$\neg Absit2(h) \supset (Holds(f, Sit2(h)) \equiv h(f)) \quad (16)$$

where  $h$  is a predicate variable, with this predicate taking a fluent argument ( $Absit2$  is a predicate on fluent predicates, and  $Sit2$  is a function from fluent predicates to situations). The axiom states that for every non-abnormal set of fluents, there is some situation in which exactly these fluents hold.

Let  $Y_3$  be the conjunction of (1)-(5),(16), and the following uniqueness of names axioms:

$$UNA[Alive, Loaded], \quad (17)$$

$$UNA[Sit2]. \quad (18)$$

Note that the domain-closure axiom (6) is no longer necessary; if there are additional fluents besides *Alive* and *Load*, then (16) will apply to them as well. Note also that we no longer have any need for generalized fluents. We have:

$$\overline{Y_3} \equiv Circum(Y_3; Absit2; Ab, Holds, Result, S0) \wedge Circum(Y_3; Ab; Result).$$

Since  $Absit2$  is a second-order predicate, its circumscription will be a formula in *third-order* logic; strictly speaking, this generalizes the standard definition of circumscription from [8], but this generalization is entirely straightforward.

**Proposition 5**  $\overline{Y_3}$  is consistent.

**Proof.** A model of  $\overline{Y_3}$  can be defined in a straightforward manner.  $\square$

**Proposition 6**  $\overline{Y_3} \models (8)$

**Proof.** Axiom (18) tells us that for each predicate  $h$  on fluents (i.e., for each  $h \subseteq |M|_f$ ),  $Sit2(h)$  will be a distinct situation. Since we have no domain

constraints, it must be the case that  $Absit2 = \emptyset$ , (or else, by the argument used in the proof of Proposition 4, the first circumscription in the definition of  $\overline{Y_3}$  would not be satisfied.) Therefore, for every possible combination of fluents, there is a corresponding situation. The proposition then follows by the argument that was used in the proof of Proposition 2.  $\square$

## 6 Examples

In order to compare the different approaches to the Yale shooting problem, this section will discuss two additional temporal reasoning problems. Section 6.1 contains an example that requires reasoning backward in time, while Section 6.2 discusses the issue of ramifications.

### 6.1 The murder mystery

Consider the following temporal explanation problem, which we will call the murder mystery. In this variation on the Hanks–McDermott problem, Fred is alive in the initial situation, and after the actions *Shoot* and then *Wait* are performed in succession (the opposite of the Yale shooting order), he is dead:

$$\begin{aligned} & Holds(Alive, S0), \\ S2 = & Result(Wait, Result(Shoot, S0)), \\ & \neg Holds(Alive, S2). \end{aligned}$$

The system is faced with the task of determining when Fred died, and whether or not the gun was originally loaded. If we used the obvious monotonic frame axioms, we would be able to conclude that the gun was originally loaded, and that Fred died during the shooting. Unfortunately, the standard circumscription policy is unable to reach this same conclusion. It has no preference for when Fred died, and even if it were told that Fred died during

the shooting, it still would not conclude that the gun was originally loaded. Surely, assuming that the gun was loaded is the only way to explain the

$$Ab(Alive, Shoot, S0)$$

that would be entailed by Fred being shot to death, but circumscription is in the business of minimizing abnormalities — not explaining them.

Chronological minimization only makes the situation worse. It tries to delay abnormalities as long as possible, so it avoids any abnormality during the shooting phase by postponing Fred’s death to the waiting phase. It therefore concludes that the gun must have been *unloaded*!

Causal minimization yields the intuitive answer since it is only by assuming that the gun was originally loaded that it can explain the death without introducing an additional causal rule.

Our method also gives the right answer, although there is a fine point that we have glossed over so far. In addition to *Result*, the situation constants  $S0$  and  $S2$  also must be allowed to vary during the circumscription of  $Ab$ . Since *Holds* is fixed during this circumscription, it is only by varying  $S0$  and  $S2$  that circumscription can be of any use in determining which fluents hold in these situations. In general with our approach, all situation-valued constants and functions should be allowed to vary; this did not matter for the original shooting problem since in that problem, the situation  $S0$  was fully specified. With this elaboration, the correctness criterion (8) is satisfied as before.

## 6.2 Ramifications

Often, it is impractical to list explicitly all the consequences of an action. Rather, some of these consequences will be *ramifications*; that is, they will be implied by domain constraints [4]. One of the main advantages of our method over causal minimization is that ours can handle ramifications, while causal minimization cannot.

A simple example of this limitation of causal minimization can be obtained from Hanks and McDermott’s original version of the Yale shooting

problem in which the gun was unloaded in the initial situation, and a *Load* action was performed before the waiting (the following example is based on one from Ginsberg [3]). Using the causal notation from [9], we assert that *Load* causes the gun to be loaded, and that *Shoot* causes Fred to be not alive (provided the appropriate preconditions are satisfied):

$$\begin{aligned} & \text{Causes}(\text{Load}, \text{Loaded}, \text{True}), \\ & \text{Causes}(\text{Shoot}, \text{Alive}, \text{False}). \end{aligned}$$

Suppose that we now add the fluent *Walking* and a domain constraint stating that in order to be walking, one must be alive:

$$\text{Holds}(\text{Walking}, s) \supset \text{Holds}(\text{Alive}, s). \quad (19)$$

Suppose also that in addition to the usual initial conditions (3) and (4), we also know that Fred is originally walking:

$$\text{Holds}(\text{Walking}, S_0).$$

It would be nice if by minimizing *Causes* we could conclude that not only does shooting make Fred not alive; it also makes him not walking:

$$\text{Causes}(\text{Shoot}, \text{Walking}, \text{False}). \quad (20)$$

Unfortunately, there is another causally minimal model in which *Load* kills Fred. This model contains:

$$\text{Causes}(\text{Load}, \text{Alive}, \text{False}), \quad (21)$$

$$\text{Causes}(\text{Load}, \text{Walking}, \text{False}). \quad (22)$$

In this model, when *Load* loads the gun, it also kills Fred. Thus, there need be no situation in which the gun is loaded while Fred is walking. To see why this model is causally minimal, observe that if there were a situation in which the gun were loaded and Fred were walking, then due to the domain

constraint (19), shooting Fred in this situation would force him to cease walking; and since causal minimization allows no effects without causes, (20) would have to be added (which is what we would like). In the unintended causally minimal model, however, by adding the new causal laws (21) and (22), such a situation need not occur, and thus (20) does not have to be added. Note the similarity of this difficulty to the original shooting problem: by adding spurious changes, certain states of the world never come about and thus the abnormalities that would have been associated with these states are ignored.

There have been some attempts at resolving this difficulty while remaining within the causal minimization framework, but so far none seem adequate. Lifschitz [9], for example, requires that fluents be divided into two groups, primitive and nonprimitive, with the causal laws and the frame axiom limited in application to the primitive fluents, and with the values of the nonprimitive fluents determined by their definitions in terms of the primitive fluents. This would work if we were axiomatizing *Alive* and *Dead* as we could make *Alive* a primitive fluent and *Dead* a nonprimitive fluent with the definition *Not(Alive)*. It would not apply, however, to the current example. Neither *Walking* nor *Alive* can be defined in terms of the other, so the frame axiom must apply to both. Thus, with the causal minimization approach, we would have to explicitly assert (20). It has been argued that it will generally be intractable to precompute all the ramifications in this manner [4].<sup>6</sup>

The approach advocated in this paper handles ramifications correctly. Domain constraints determine which situations can exist in the model; in the above example, there are only six types of situations that can exist: *Loaded* can be true or false, *Alive* can be true or false, and if *Alive* is true, then *Walking* can be true or false (otherwise, it must be false). The causal laws, like (2), further constrain the resultant situation of an action. So if

---

<sup>6</sup>Even if this problem were solved, causal minimization has still another problem: since the *Causes* predicate does not take a situational argument, it cannot be used to formalize context-dependent effects.

the gun is loaded, and Fred is alive, then (2) will demand that Fred must be not alive after being shot, and the domain constraint (19) will ensure that Fred is not walking in this resulting situation. Finally, the minimization of abnormality will keep the gun loaded. The strange behavior of the causal approach cannot occur here because we require all possible situations to exist in the model.

Although the new approach can handle ramifications without introducing absurdities, it does not “solve the ramification problem.” Consider a robot that is standing at location  $L_1$ , where this robot is holding an ice-cream cone. Suppose the robot now moves to location  $L_2$ . Does the ice-cream cone’s absolute position persist, or does the fact that it is being held by the robot persist? (equivalent scenarios have been discussed by many authors, e.g., [4, 16, 18]). Based on our understanding of the domain, we would probably guess that after the robot moves, it will continue to hold the ice-cream cone. Yet, if we simply write down the domain constraint:

$$\text{Holds}(\text{Holding}(x, y), s) \wedge \text{Holds}(\text{At}(x, l), s) \supset \text{Holds}(\text{At}(y, l), s),$$

then both persistence possibilities will correspond to legitimate minimal models. Clearly this is all that can be expected unless the system is supplied with additional information. If domain constraints are to be of much use for reasoning about action, we will need convenient methods for encoding this additional information that is used in resolving ambiguities. Ginsberg and Smith [4] discuss a few such methods, but more work is required.

## 7 Conclusion

This paper has presented a new approach to nonmonotonic temporal reasoning. The approach correctly formalizes problems requiring reasoning both forward and backward in time, and it allows for some of an action’s effects to be specified indirectly using domain constraints.

It should be noted that, in a certain sense, our solution works for the same reason that causal minimization does. Causal minimization is not tempted to unload the gun because it is minimizing the extent of the *Causes* predicate rather than actual changes in the world; unloading the gun would prevent the *Causes(Shoot, Alive, False)* fact from *being used*, but it would not eliminate the fact itself. Similarly, our solution minimizes even those abnormality facts associated with situations that do not really happen. But since we stick with the standard axioms (rather than introducing a special *Causes* predicate), our approach appears to be the more robust of the two: for us, even those abnormalities that arise as ramifications can not be eliminated by the assumption of unmotivated changes.

## Acknowledgements

I would like to thank Vladimir Lifschitz, Karen Myers, Arkady Rabinov, Yoav Shoham, David Smith, and especially Matthew Ginsberg for helpful comments, discussions, and encouragement. The author is supported by an AFOSR graduate fellowship.

## References

- [1] A.B. Baker, A simple solution to the Yale shooting problem, in: R.J. Brachman, H.J. Levesque, and R. Reiter, eds., *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, Toronto, Canada (Morgan Kaufmann, San Mateo, CA, 1989) 11–20.
- [2] M. Gelfond, Autoepistemic logic and formalization of common-sense reasoning, in: M. Reinfrank, ed., *Non-Monotonic Reasoning: Second International Workshop (Lecture Notes in Computer Science 346)* (Springer-Verlag, Berlin, 1989) 176–186.

- [3] M.L. Ginsberg, personal communication (1988).
- [4] M.L. Ginsberg and D.E. Smith, Reasoning about action I: A possible worlds approach, *Artificial Intelligence* **35** (1988) 165–195.
- [5] S. Hanks and D. McDermott, Nonmonotonic logics and temporal projection, *Artificial Intelligence* **33** (1987) 379–412.
- [6] B.A. Haugh, Simple causal minimizations for temporal persistence and projection, in: *Proceedings AAAI-87*, Seattle, WA (1987) 218–223.
- [7] H.A. Kautz, The logic of persistence, in: *Proceedings AAAI-86*, Philadelphia, PA (1986), 401–405.
- [8] V. Lifschitz, Computing circumscription, in: *Proceedings IJCAI-85*, Los Angeles, CA (1985) 121–127; also in: M.L. Ginsberg, ed., *Readings in Nonmonotonic Reasoning* (Morgan Kaufmann, Los Altos, CA, 1987) 167–173.
- [9] V. Lifschitz, Formal theories of action, in: F.M. Brown, ed., *The Frame Problem In Artificial Intelligence: Proceedings of the 1987 Workshop*, Lawrence, KS (Morgan Kaufmann, Los Altos, CA, 1987) 35–58; also in: M.L. Ginsberg, ed., *Readings in Nonmonotonic Reasoning* (Morgan Kaufmann, Los Altos, CA, 1987) 410–432.
- [10] V. Lifschitz, Pointwise circumscription, in: M. L. Ginsberg, ed., *Readings in Nonmonotonic Reasoning* (Morgan Kaufmann Publishers, Los Altos, CA, 1987) 179–193.
- [11] V. Lifschitz, personal communication (1989).
- [12] J. McCarthy, Circumscription — a form of non-monotonic reasoning, *Artificial Intelligence* **13** (1980) 27–39.
- [13] J. McCarthy, Applications of circumscription to formalizing common-sense knowledge, *Artificial Intelligence* **28** (1986) 89–116.

- [14] J. McCarthy and P.J. Hayes, Some philosophical problems from the standpoint of artificial intelligence, in: B. Meltzer and D. Michie, eds., *Machine Intelligence 4* (American Elsevier, New York, 1969) 463–502.
- [15] P.H. Morris, The anomalous extension problem in default reasoning, *Artificial Intelligence* **35** (1988) 383–399.
- [16] K.L. Myers and D.E. Smith, The persistence of derived information, in: *Proceedings AAAI-88*, St. Paul, MN (1988) 496–500.
- [17] J. Pearl, On logic and probability, *Computational Intelligence* **4** (1988) 99–103.
- [18] M. Reinfrank, Multiple extensions, where is the problem?, in: F.M Brown, ed., *The Frame Problem in Artificial Intelligence: Proceedings of the 1987 Workshop*, Lawrence, KS (Morgan Kaufmann, Los Altos, CA, 1987) 291–295.
- [19] E. Sandewall, Filter preferential entailment for the logic of action in almost continuous worlds, in: *Proceedings IJCAI-89*, Detroit, MI (1989) 894–899.
- [20] Y. Shoham, Chronological ignorance: Experiments in nonmonotonic temporal reasoning, *Artificial Intelligence* **36** (1988) 279–331.