MANY AGENTS ARE BETTER THAN ONE

Technical Note 417

March 1987

By: Michael P. Georgeff
Program Director

Representation and Reasoning Program
Artificial Intelligence Center
Computer and Information Sciences Division
and
Center for the Study of Language and Information

APPROVED FOR PUBLIC RELEASE:
DISTRIBUTION UNLIMITED

An earlier version of this paper appears in *The Frame Problem in Artificial Intelligence: Proceedings of the 1987 Workshop*, published by Morgan Kaufmann Publishers, Los Altos, California, and is reproduced here with their kind permission.

This research has been made possible by a gift from the System Development Foundation, by the Office of Naval Research under Contract N00014-85-C-0251, and by the National Aeronautics and Space Administration, Ames Research Center, under Contract NAS2-12521.

The views and conclusions contained in this paper are those of the author and should not be interpreted as representative of the official policies, either expressed or implied, of the Office of Naval Research, NASA, or the United States Government.

SRI International

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

Report Documentation Page						Form Approved OMB No. 0704-0188
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.						
1. REPORT DATE **MAR 1987**		2. REPORT TYPE				3. DATES COVERED **00-03-1987 to 00-03-1987**
4. TITLE AND SUBTITLE **Many Agents Are Better Than One**						5a. CONTRACT NUMBER
						5b. GRANT NUMBER
						5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)						5d. PROJECT NUMBER
						5e. TASK NUMBER
						5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **SRI International,333 Ravenswood Avenue,Menlo Park,CA,94025**						8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)						10. SPONSOR/MONITOR'S ACRONYM(S)
						11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT						
15. SUBJECT TERMS						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**17**		

Abstract

This paper aims to show how much of the frame problem can be alleviated by using domain models that allow for the simultaneous occurrence of actions and events. First, a generalized situation calculus is constructed for describing and reasoning about events in multiagent settings. Notions of *independence* and *causality* are then introduced and it is shown how they can be used to determine the persistence of facts over time. Finally, it is shown how these notions, together with traditional predicate circumscription, make it possible to retain a simple model of action while avoiding most of the difficulties associated with the frame problem.

1 Introduction

Although the so-called frame problem has been regarded as presenting a major difficulty for reasoning about actions and plans, there is still considerable disagreement over what it actually is. Some researchers, for example, see the problem as largely a matter of combinatorics [14,19]; others view it as a problem of reasoning with incomplete information [15]; and yet others believe it relates to the difficulty of enabling systems to notice salient properties of the world [7].

I shall take the problem to be simply that of specifying adequate and reasonably natural "laws of motion," [14] – i.e., of constructing formulations in which it is possible to readily specify and reason about the properties of events and situations. This gives rise to at least five related subproblems, which I discuss below.

The first of these is what I shall call the *combinatorial problem.* While it does not appear too difficult to give axioms that describe the *changes* wrought by some given action, it seems unreasonable to have to write down axioms describing all the properties *unaffected* by the action. Axioms of the latter kind are usually called *frame axioms* [14] and, in general, they need to be given (or, at least, be deducible) for all property-action pairs. (Note that I use "unaffected" rather than "unchanging." This is an important distinction if we want to allow for concurrent events, as in most real-world domains.) The real problem here is to avoid *explicitly* writing down (or having to reason with) all the frame axioms for every property-action pair. Most solutions to this problem attempt to formalize some closed-world assumption regarding the specification of these dependencies.

Two further problems arise as a result of the fact that specifying the effects of actions is usually subject to qualification. The first sort of qualification has to do with the conditions under which the action effects certain *changes* in the world. This is what I call the *precondition qualification* problem (also variously called the intra-frame problem [20] and, simply, the qualification problem [13]). The second sort of qualification concerns the *extent of influence* of the action (or what remains *unaffected* by the action). I call it the *frame qualification* problem (also called the inter-frame problem [20] and the persistence problem [15]). Let me give examples of these two kinds of qualification.

For example, consider that we are trying to determine what happens if Mary fires a loaded gun [at point-blank range] at Fred [6]. Given such a scenario, we should be able to derive, without having to state a host of qualifications, that Fred dies as a result of the shooting. However, if we then discover (or are given an extra axiom to that effect) that the gun was loaded with a blank round, the conclusion (that Fred dies) should be defeasible – i.e., we should be able to accommodate the notion of Fred's possibly being alive after the firing. This is the precondition qualification problem. Most solutions to this problem aim to formalize the rule: "These are the only *preconditions* that matter as far as the performance of the action is concerned, *unless* it can be shown otherwise."

As an example of the frame qualification problem, consider the point at which Mary loads the gun prior to firing it at Fred. All things being equal, it should be possible to derive that Fred's state of being is unaffected by loading the gun. However, if we discover that Fred actually died while the gun was being loaded, it should be possible (without changing the theory, except for the additional axiom about the time of Fred's death) to accommodate this without inconsistency and, if desired, to derive other theorems about the resulting situation. Most solutions to this problem attempt to formalize the rule thus: "These are the only *effects* of the action (given the preconditions), *unless* it can be shown otherwise." Solutions to this last problem often provide a solution to the combinatorial problem; the two problems, however, are clearly distinct.

The fourth problem concerns the ability to write down certain axioms of invariance regarding world states. I shall call this the *consistency constraint* problem (also called the ramification problem [3]). Just as with deductive databases, it should not be necessary to write down explicitly every fact about a certain problem. For example, I should be able to formulate an axiom stating that everyone stops breathing when they die. Then I should be able to state that the effect of shooting a loaded gun at Fred results in his death without having to specify that it also results in cessation of his breathing. This latter effect of the shooting action should be inferable from the first axiom describing the consequences of dying. This seems straightforward, but the real problem arises from the fact that such axioms can complicate the solution of the previous problems.

The fifth problem, which I call the *overcommitment problem*, arises primarily in multiagent domains. In a dynamic world, or one populated with many agents, we want our solutions to allow for the independent activities of other agents. Most importantly, we do not want to have to specify explicitly all the external events that might conceivably occur. But if we leave the occurrence of such events unspecified, we do not want a solution to the previously mentioned problems to overcommit us to a world in which these events are thereby assumed *not* to have occurred.

In the next few sections, I shall lay the foundations for a solution to these problems. A more detailed description of the formalism can be found in reference [4].

2 Actions and Events

We consider that, at any given instant, the world is in a particular *world state.* Various properties may hold of each of these states. A sequence of world states is called a *world history.* A given world state has no duration; the only way the passage of time can be observed is through some change of state. The world changes its state by the occurrence of *events* or *actions.* An *event type* is a set of state sequences, representing all possible occurrences of the event *in all possible situations* [1,15]. An *event instance* is a particular occurrence of an event type in a particular state in a particular world history. Except where the distinction is important, we shall call event types simply events.

We shall restrict our attention herein to *atomic events.* An atomic event is one in which each state sequence comprising the event contains exactly two elements; it can thus be modeled as a transition relation on world states. The transition relation of a given event must comprise all possible state transitions, *including those in which other events occur simultaneously with the given event.* Consequently, the transition relation of an atomic event places restrictions on those world relations that are directly affected by the event, but leaves most others to vary freely (depending upon what else is happening in the world). This is in contrast to the classical approach, which views an event as changing some world relations but leaving most of them unaltered.

Of course, to specify events by listing all the possible transitions explicitly would, in any interesting case, be infeasible. We therefore need some formalism for describing events and world histories. The one we use here is a generalization of the situation calculus [14], although most of our remarks would apply equally to other logic-based formalisms.

We first introduce the notion of a *fluent* [14], which is a function defined on world states. We shall only be concerned with *propositional* fluents, whose value in a given state is a truth value (*true* or *false*). To denote the fact that the fluent f has value *true* in state s, we use the notation $holds(f, s)$. For example, if *alive* denotes the fluent representing Fred's being alive, $holds(alive, s)$ will be true if the fluent *alive* has value *true* in state s; that is, if Fred is alive in s.

As in the single-agent case, the well-formed formulas of this situation calculus may contain logical connectives and quantifiers; they can thus express general assertions about world histories. However, we do not use a "result" function to specify the state resulting from an event (or action). The reason is that, in our formalism, events are not functions on states but rather relations on states, and the occurrence of an event in a given state need not uniquely determine the resulting state. Therefore, for a given world history w containing state s, we let $succ(s, w)$ be the successor of s in w, and use a predicate $occurs(e, s, w)$ to mean that event e occurs in state s and history w. This formulation, in addition to allowing a wider class of events than in the standard situation calculus, also enables us to state arbitrary temporal constraints on world histories [17].

In that which follows, we shall use variables of four sorts: (1) state (situation) variables (usually given as s); (2) history variables, (w); (3) event variables (e); and (4) variables denoting propositional fluents (ϕ, ψ). We also simplify the notation in two ways. First, we shall assume throughout that all free variables are universally quantified. Second we assume that, in any formula involving the history variable w, both s and $succ(s, w)$ are elements of w, without stating this explicitly.

In reasoning about actions and events, one of the most important things we need to know is how they affect the world – that is, we must be able to specify the effects of actions and events when performed in given situations. To do this, we introduce a predicate $effects(\phi, e, \psi)$, which is intended to mean that, if event e occurs in a state (situation) in which the fluent ϕ holds, ψ will hold in the resulting state. (The predicate *effects* is essentially the same as the predicate *causes* as used by Lifschitz [12]). We can write this more formally as follows:

$$effects(\phi, e, \psi) \wedge holds(\phi, s) \wedge occurs(e, s, w) \supset holds(\psi, succ(s, w)) \quad (*) \qquad (1)$$

For example, an effect of shooting a gun at Fred might be described by the following axiom:

$$effects(loaded, shoot, \neg alive) \ .$$

This represents the fact that the shooting of a loaded gun will result in Fred's death.

Of course, as mentioned previously, statements such as this are often subject to qualification – under certain abnormal situations, the shooting of a loaded gun will not result in Fred's death. One way to handle this problem is to introduce an abnormality predicate, *ab* say, to allow for such situations. For example, we could replace axiom (1) by the axiom

$$\neg ab(e,s) \supset (\mathit{effects}(\phi,e,\psi) \wedge \mathit{holds}(\phi,s) \wedge \mathit{occurs}(e,s,w) \supset \mathit{holds}(\psi,\mathit{succ}(s,w))) \qquad (2)$$

Alternatively, we could regard such statements as *defeasible* inference rules [16,18] or, more simply, as inference rules that may be occasionally *unsound* [9]. However, instead of committing to any one of these views, we shall simply indicate with an asterisk ($*$) those axioms that are subject to qualification.

It is very important to note that those properties that are *unaffected* by an event need not persist during every occurrence of the event – whether or not they do will depend on what other events are occurring at the same time. For example, if shooting events, per se, do not influence the weather, it could well be false that $\mathit{effects}(\mathit{raining},\mathit{shoot},\mathit{raining})$. In contrast, this statement would be considered true in most single-agent formalisms.

3 Independence

We have been regarding atomic events as imposing certain constraints on the way the world changes while leaving other aspects of the situation free to vary as the environment chooses. That is, each event's transition relation describes all the potential changes of world state that could take place during the occurrence of the event. Which transition actually occurs in a given situation depends, in part, on the events that take place in the environment. However, unless we can reason about what happens when some subset of all possible events occurs – for example, when the only relevant events occurring are those initiated by the agent of interest – we could predict very little about the future and any useful planning would be impossible.

To handle this problem, we first introduce the concept of *independence*. We define a predicate $\mathit{indep}(\phi,e,\psi)$, which we take to mean that the fluent ψ is independent of (i.e., not directly affected by) event e in situations in which ϕ holds. Unlike classical models of actions and events, this does not mean that, if we are in a situation in which both ϕ and ψ hold, ψ will also hold in the resulting state. Rather, if ψ is independent of e in some situation, the transition relation associated with e will include transitions to states in which ψ does not hold, as well as ones in which ψ holds, while not constraining the values of any other fluents in the resulting state.

For example, the axiom

$$\mathit{indep}(\neg \mathit{loaded},\mathit{shoot},\mathit{alive})$$

states that, for all situations in which the gun is unloaded, Fred's state of health will be unaffected by the shooting of the gun (though it may be affected by other, possibly concurrent, events).

In our ontology, a world state can change only through the occurrence of events. Furthermore, in keeping with our intuitive notion of independence, events that are independent of some property cannot influence that property. We therefore have

$$holds(\phi, s) \wedge holds(\psi, s) \wedge \neg holds(\psi, succ(s, w)) \supset \exists e \, . \, (occurs(e, s, w) \wedge \neg indep(\phi, e, \psi)) \tag{3}$$

From this we can directly deduce the following *law of persistence*:

$$holds(\phi, s) \wedge holds(\psi, s) \wedge \forall e \, . \, (occurs(e, s, w) \supset indep(\phi, e, \psi)) \supset holds(\psi, succ(s, w)) \tag{4}$$

This rule states that, if we are in a state s in which both ϕ and ψ hold, and if all events that occur in state s and history w are independent of ψ under conditions ϕ, then ψ will also hold in the next (resulting) state. For example, we could use this rule to infer that, if *shoot* were the only event to occur in some state s, and the gun was unloaded in s, Fred would be alive after the shooting if he were alive prior to it.

Unlike many other approaches to persistence [2,6,8,15,19,20], the foregoing law does not involve any nonmonotonic operators or depend on any consistency arguments. Nor is it some "tendency" of nature or some fortuitous property of the world in which we live. Rather, it is a *direct consequence* of our notions of event and independence. What makes planning useful for survival is the fact that we can structure the world in a way that keeps most properties and events independent of one another, thus allowing us to reason about the future without complete knowledge of all the events that could possibly be occurring.

We now have means of specifying how events affect or determine the values of certain fluents in the domain (using *effects*) and a means of specifying which fluents are unaffected by events (using *indep*). Clearly, these obey certain constraints (for example, a fluent that is changed by the occurrence of an event cannot be independent of that event). Such constraints can be used to ease the specification of independence, but I shall not pursue these details here.

4 Causality

One problem that we have not properly addressed is the apparent complexity of the axioms that describe the effects of actions. For example, while it might seem reasonable to state that the location of block B is independent of the movement of block A, this is simply untrue, as everyone knows, in most interesting worlds. Whether or not the location of B is independent of the movement of A will depend on a host of conditions, such as whether B is in front of A, on top of A, atop A but tied to a door, and so on.

One way to handle this problem is to introduce another abnormality predicate to handle the special cases in which independence between a fluent and an event may be compromised. While this may work in relatively simple domains, it is not clear that such an approach would be expressive enough for reasoning about complex, real-world problems.

An alternative scheme is to introduce a notion of inter-event *causality* [4] [1]. I shall use the expression $causes(\phi, e_1, e_2)$ to denote the fact that, under conditions ϕ, event e_1 causes event e_2. In the general case, it is desirable to allow two kinds of inter-event causation, one in which an event causes the simultaneous occurrence of another event, and the other in which an event causes the occurrence of a consecutive event. For simplicity, however, I shall assume herein that all causal relations are between simultaneous events. Under this restriction, the axiom expressing the effects of causation is

$$causes(\phi, e_1, e_2) \wedge holds(\phi, s) \wedge occurs(e_1, s, w) \supset occurs(e_2, s, w) \quad (*) \tag{5}$$

As with the axiom describing the effects of events, this law is defeasible.

For example, we might have a causal law to express the fact that, whenever a block A is moved, any block on top of A and not somehow restrained (e.g., by a string tied to a door) will also move. We could write this as

$$causes(on(B, A) \wedge \neg restrained(B), moveA, moveB)$$

If this axiom holds, the movement of A will *cause* the simultaneous movement of B whenever B is on top of A and is not restrained.

We also require that all events that occur are either agent-initiated or caused by other events. We shall write $act(a, e, s, w)$ to denote the fact that agent a initiates the event e in state s and history w and will use $caused(e, s, w)$ to indicate that the occurrence of e in s and w is caused by some other event. The predicate *caused* can be defined as follows:

$$caused(e_2, s, w) \equiv \exists e_1, \phi \,.\, (causes(\phi, e_1, e_2) \wedge holds(\phi, s) \wedge occurs(e_1, s, w) \wedge occurs(e_2, s, w)) \tag{6}$$

We can now express the requirement that all events must have some reason for occurring:

$$occurs(e, s, w) \supset \exists a \,.\, act(a, e, s, w) \vee caused(e, s, w) \tag{7}$$

In one sense, agent-initiated events can be viewed as arising from the free-will of the agents that populate the domain, whereas caused events are determined by the other happenings that take place. Of course, it may be that events identified as arising from free-will turn out to result from some previously unknown causal relationship.

We use the notion of causality in a purely technical sense and, while it has many similarities to commonsense usage, we are not proposing it as a fully-fledged theory of causality. Essentially, we view causation as a relation between atomic events that is conditional on the state of the world. We also relate causation to the temporal ordering of events, and assume that an event cannot cause another event that precedes it. However, as stated above, we do allow an event to cause another that occurs simultaneously. This differs from most other formal models of causality [11,15,20], although Allen [1] also allows simultaneous causation.

[1] It is important to note that I use causality in a purely technical sense and do not intend that it reflect our intuitive notions of causation. As used herein, causality corresponds most closely to what Goldman [5] calls *generation*.

Before concluding this section, it is important to understand how the notions of independence and causality relate to one another. The first issue is a technical one: if an event e_1 causes an event e_2 to occur simultaneously, we allow that the fluents affected by e_2 be independent of e_1, even if e_1 *always* causes e_2. Thus, viewed intuitively, our notion of independence between a fluent and an event corresponds to the fluent not being *directly* affected by the event, even though it may be indirectly affected. Of course, in this case, the fluents affected by e_2 (and, similarly, e_1) will not be independent of the *composite* event consisting of e_1 and e_2 occurring simultaneously.

The second issue concerns whether to view axioms about independence as being subject to qualification when applied to real-world domains. For example, assume that we specify that Fred's being alive is independent of the loading of the gun:

$$indep(true, load, alive) \ .$$

But say that, on loading the gun, we notice that Fred dies. There are two ways we could explain this observation. The first would be to say that it is not really true that Fred's state of health is independent of loading the gun – under certain abnormal conditions, Fred's health *is* affected by gun loading. The second view is that, while the loading of the gun, per se, does not affect Fred's health, some other event occurs simultaneously with the loading, and this other event is the cause of Fred's dying.

In commonsense reasoning, both forms of explanation seem to be used. However, in trying to formalize planning and reasoning about action, we can afford to be more frugal: we don't need to support both kinds of explanation. Thus, the view I choose to adopt is that independence is *not* subject to qualification, and that unusual happenings result, not from unexpected dependencies, but from the occurrence of unexpected or unlikely events. This allows us to provide very simple axioms regarding independence, at the expense of introducing more complex causal laws.

5 The Yale Shooting Problem

We have now laid all the foundations we need for our theory of actions and plans. Most importantly, the theory is a standard first-order theory – it does not contain any nonmonotonic operators nor have we invoked (yet) any minimization criteria. To see how everything works, we shall consider in detail the problem posed by Hanks and McDermott [6], called the Yale shooting problem.

The Yale shooting problem is as follows. At some point of time, Fred is alive. Mary then loads a gun, waits a moment, and finally fires the gun at Fred. We take the fluents of the domain to include *loaded* (representing the fact that the gun is loaded) and *alive* (Fred is alive), and the events to include *load* (the gun gets loaded), *wait* (Mary waits a moment), and *shoot* (Mary shoots the gun at Fred). We can then describe the shooting scenario using the following axioms:

Effects of Actions:

1. $effects(true, load, loaded)$
2. $effects(loaded, shoot, \neg alive)$

Independence:

3. $indep(true, load, alive)$
4. $indep(\neg loaded, shoot, alive)$
5. $indep(true, wait, \phi)$

Law of effects:

6. $effects(\phi, e, \psi) \land holds(\phi, s) \land occurs(e, s, w) \supset holds(\psi, succ(s, w))$ $(*)$

Law of persistence:

7. $holds(\phi, s) \land holds(\psi, s) \land \forall e . (occurs(e, s, w) \supset indep(\phi, e, \psi)) \supset holds(\psi, succ(s, w))$

Law of causality:

8. $causes(\phi, e_1, e_2) \land holds(\phi, s) \land occurs(e_1, s, w) \supset \exists e_2 . occurs(e_2, s, w)$ $(*)$

Reasons for events:

9. $occurs(e, s, w) \supset \exists a . act(a, e, s, w) \lor caused(e, s, w)$

This problem:

10. $holds(alive, s_0)$
11. $occurs(load, s_0, w_0)$
12. $occurs(wait, s_1, w_0)$
13. $occurs(shoot, s_2, w_0)$
14. $equals(w_0, seq(s_0, s_1, s_2, s_3))$

Axioms (1) and (2) simply state that loading a gun always results in it being loaded and shooting a loaded gun always results in Fred's death. Axioms (3) and (4) state that Fred's being alive is not influenced by loading a gun nor by shooting an unloaded gun, and axiom (5) states that *no* fluent is influenced by waiting. Axioms (6) and (7) are the basic laws of effects and persistence, axiom (8) defines causality, and axiom (9) says that there must be a reason for every event. Axioms (10) through (14) describe the sequence of events that occur in this scenario.

As such, there is almost nothing of interest we can prove about this domain. To progress further, we need to make some *assumptions* about the problem. For example, we could assume that the only relevant events that occur are those that are explicitly specified. (Of course, if we were to observe that that the world behaved differently than predicted, we would have to withdraw some of these assumptions.) Alternatively, we could have much more sophisticated metatheories to describe the process of making (and retracting) assumptions in the problem domain of interest.

However, my aim in this paper is to show how traditional circumscription can be used as a means of capturing an important class of assumptions, while avoiding the pitfalls identified by Hanks and McDermott.

6 Circumscribing Causality

At this point, I want to defer the problem of specifying independence; for now, I shall assume that independence can be specified as fully as we need it for the domain of interest.

The independence axioms as given above meet this requirement. (Those who believe I have thus side-stepped the frame problem should perhaps read the section on the specification of independence before proceeding further.)

Let us begin by examining what happens if we choose to minimize the inter-event causality predicate, *causes*. In doing this, we allow *holds*, *occurs*, and *caused* to vary. Furthermore, let us assume that Mary is the only agent that can initiate events – all others must somehow be *caused* by the activities of Mary. We can specify this quite simply:

$act(a, e, s, w) \supset (a = Mary)$

Finally, let us assume that the only events performed by Mary are the ones specified: *load*, *wait*, and *shoot*. We thus have the following axioms:

$act(Mary, e, s_0, w_0) \supset (e = load)$
$act(Mary, e, s_1, w_1) \supset (e = wait)$
$act(Mary, e, s_2, w_2) \supset (e = shoot)$

It is then not difficult to show that the resulting circumscribed theory yields only intended models and that, in each of these, Fred is dead in state s_3. Moreover, the extension of the minimized predicate *causes* is empty.

Furthermore, the same results hold if, in addition, we choose to handle the precondition-qualification problem by introducing abnormality predicates into those axioms subject to qualification (indicated by $*$) and minimize their extension. In this case, the extension of these abnormality predicates would be empty and we again obtain the desired results.

But note that, given the above axioms, we cannot conclude anything about whether or not the gun is loaded after the shooting (in contrast to the solution described by Hanks and McDermott). We thus avoid, in this case, overcommitting ourselves to a maximally static world. This means that we do not get a unique model, but the ones we do get are indeed intended ones.

Thus, the problem posed by Hanks and McDermott presents no difficulties for us – we do not get any unintended models. But have we, in solving the problem, thrown away the other ingredients that are essential for a commonsense theory of actions and plans? To help answer this question, let us consider some variations of the problem.

First, let us assume that we discover that Fred is not dead in state s_3. That is, what happens if we add the axiom $holds(alive, s_3)$ to the circumscribed theory given above? In that case, we obtain models corresponding to each of the following scenarios:

1. Because of an abnormality in state s_0, the loading of the gun fails to result in it actually getting loaded,
2. Because of an abnormality in state s_2, the shooting of the loaded gun fails to result in Fred's death,
3. Mary's waiting in state s_1 *causes* the occurrence of another event which results in the gun being unloaded in state s_2.

All these are reasonable explanations for the fact that Fred remains alive after the shooting. Thus, the approach seems to provide exactly the intended models even as we vary the facts of the shooting.

Had we also chosen to minimize the actions of Mary (rather than explicitly delimit them), we would have had yet another alternative corresponding to the case in which Mary herself unloaded the gun prior to the shooting. In some situations, one might want to allow this (e.g., when testing the psychic skills of Uri Geller). However, in most cases, this is not what one wants, particularly when planning one's own course of action.

It is important to be aware of the consequences of adopting one or more of these various explanations. For example, in the case that an abnormality occurs in state s_0 (case (1) above), we shall not be able to conclude anything about how the attempted loading of the gun affects world properties that are not independent of the event. If, in normal situations, the loading of the gun results in a reduction in the number of bullets in the ammunition pack, we could not, in abnormal situations, predict whether or not this would still be the case. That is, it would be unknown (unless we had other sources of evidence) whether the number of bullets had decreased or not. However, given appropriate axioms of independence, we shall still be able to conclude that the properties independent of *load* (such as the fluent *alive*) are not affected by the attempted loading. This, it seems to me, is just what we want. If some abnormality occurs, we may well not know *how* things went wrong, and thus it would be unwise to conclude anything about the properties possibly affected by the event in question. However, it would be reasonable to assume that properties completely unrelated to the event are unaffected by the event, whether it is performed in a normal situation or not. Indeed, the definition of independence requires that this be the case.

In the situation in which some extraneous event is assumed to have occurred (case (3) above), we are in quite a different position. In particular, unless we have other evidence, we shall not know *anything* about the consequences of that event. For example, we shall not be able to determine whether the event simply resulted in the unloading of the gun, or whether other things happened as well. Again, I maintain that this is just what we want of the theory – if an event is completely unanticipated, we shall not know its likely side-effects. Indeed, in many situations such unanticipated events have quite significant side-effects (such as one might experience in opening a package containing a bomb). This is quite different from the results that would be obtained using most other approaches to the frame problem [2,12,13,15,19,20] – in essence, these all minimize the *effects* of actions.

One can also restrict the event types that can possibly occur to specific classes. For example, consider that there exist two other events, *unload* and *melt*, both of which result in the unloading of the gun:

$effects(true, unload, \neg loaded)$
$effects(true, melt, \neg loaded)$
$effects(true, melt, hot)$

Furthermore, let these two event types, together with *load*, *wait*, and *shoot*, be the only events that can occur:

$$occurs(e, s, w) \supset (e = unload) \vee (e = melt) \vee (e = load) \ldots$$

In the case that Fred is still alive in state s_3, we now have the same set of possible explanations as before but the unexpected event caused by Mary's waiting (case (3) above) is now forced to be either the event *unload* or the event *melt*. Again note that this approach does not require that *changes* be minimal; in particular, one of the possible [minimal] models has the gun changing temperature as well as becoming unloaded.

Similar considerations apply when we have certain domain constraints that must hold of all world states. For example, let's assume that the following statement is true:

$$holds(alive \vee inheaven \vee miserable, s)$$

That is, in all situations, Fred is either alive, in heaven, or miserable.

Assuming that Fred is alive and neither in heaven nor miserable in state s_2, we can use the law of persistence to deduce that shooting Fred must have caused the simultaneous occurrence of another event that results in Fred going to heaven or becoming miserable, or both. As before, circumscribing *causes* will leave us unable to predict the other consequences of such an event occurrence. However, if we have specified that only certain specific events cause misery, or travel to heavenly places, we can use the properties of these events to determine all the possible ramifications of the shooting of Fred.

Thus, it is not necessary to explicity specify causal laws that ensure domain constraints are maintained: the existence of appropriate causal relations can be inferred using the law of persistence. However, if we do adopt this approach, we need some way of determining (either monotonically or nonmonotonically) which of all the potentially appropriate events is the intended one.

7 Other Agents

In all the above scenarios, we have assumed the existence of a single agent, Mary, and consequently, any events that occurred were either initiated by her or caused (directly or indirectly) by her. In this section, I wish to briefly consider domains populated with other agents.

In multiagent domains, one simple approach is to minimize the activities of various agents in the domain. For example, assume that an agent called Anne is present during the shooting of Fred and that, as long she is not disturbed, she is content to stack blocks on a table. In addition, we shall assume that shooting a gun makes a loud noise (usually), and that loud noises disturb Anne (usually). These facts could be represented by using the following axioms, together with various block-stacking axioms with which I assume we are all familiar.

$causes(true, shoot, makenoise)$
$causes(true, makenoise, disturb)$
$occurs(disturb, s, w) \equiv \neg occurs(stack, s, w)$

As such, there is very little we can conclude about state s_3. For example, Anne may have unloaded the gun or performed a variety of other acts during the course of the shooting. However, by circumscribing $act(Anne, e, s, w)$ (as well as the other predicates as discussed above), we can prove that Fred dies in state s_3, while Anne stacks blocks up until the moment (s_2) that she is disturbed by the shooting of the gun.

It is not necessary to minimize the activities of all other agents. For example, if Joe is sitting on a park bench in Palo Alto, and the shooting of Fred occurs at Yale, there is no reason to minimize the activities of Joe. Moreover, in other situations, we may wish to minimize other kinds of events based on domain-specific knowledge.

In complex domains, it appears useful to structure the domain as a set of interacting processes [4,10]. With each process, we associate a set of internal and external events, and require that there be no *direct* causal relationship between elements from each of these classes of events. Thus, the only way the internal events of a given process can influence external events (or vice versa) is through indirect causation by an event that belongs to neither category. These intermediary events are often called *ports*. Processes thus impose causal boundaries and independence properties on a problem domain, and can thereby substantially reduce combinatorial complexity.

For domains structured in this way, the event occurrences we choose to minimize will depend on the problem that we are attempting to solve. For example, in considering the behavior of some given process, it might be sensible to minimize the occurrence of interface events. In this way, it is possible to determine how the process in question operates in as much isolation as possible from the rest of the domain, without overcommitting ourselves to a static external environment. Thus, assuming that changes in the weather were external to the process of Fred's being shot, we would not be able to predict whether or not it continued raining throughout the episode. Such techniques allow us to avoid making needless – and usually invalid – assumptions about the activities of the other agents and processes.

8 Specifying Independence

Of course, we are left with the problem of *specifying* independence. There are essentially two problems here: (1) the apparent combinatorial difficulties in expressing all the required independence axioms; and (2) the complexity to be expected of many, if not most, of these axioms in real-world applications.

The first of these problems is probably overstated in much of the literature on the frame problem. Thus, just as in the axiomatization of any large or infinite domain, the number of axioms required to specify independence can be substantially reduced by the use of general axioms that allow specific instances of independence to be deduced as needed. For example, it may be that events outside a particular region (or process) R cannot affect properties inside that region:

$$internal_f(\phi, R) \wedge external_e(e, R) \supset indep(true, e, \phi)$$

In this way, a single axiom can specify independence for an entire class of fluent/event pairs. In large real-world domains, such axioms almost invariably lead to a substantial reduction in the combinatorics of the problem. In small blocks worlds, on the other hand, they may not – but writing down all the independence axioms in such cases is not much of a problem either. Elsewhere, I discuss this approach in more detail [4].

The second problem is handled by keeping independence relationships simple and introducing causal laws that describe how actions and events bring about (cause) others. Of course, the causal laws can themselves be complex (just as is the physics of the real world), but the representation and specification of actions and events are thereby kept simple. As mentioned above, appropriate structuring of the problem domain can be used to define causal boundaries and thus limit the influence of events on one another. This can considerably simplify the task of providing an adequate set of causal laws [4,11].

For those who are still unhappy with the explicit specification of independence, we can simplify the specification by again appealing to circumscription. All we need do is maximize the extension of the predicate *indep* (or, equivalently, minimize its complement).

In concluding this section, it is important to note that the root of the problem raised by Hanks and McDermott [6] is *not* that the minimization was performed over a predicate that involved situation terms. For example, in the formalism presented here, minimizing the agent-initiated events that occur in any given state does not lead to any difficulties or unintended models. The real point is that we need some *justification* for the minimizations we make; it is not often sensible to try minimizing changes in *all* world properties when told no more than that some particular event has occurred.

9 Conclusions

Our approach to the frame problem has been to construct a theory of actions and events that allows for concurrent activity, and to introduce notions of independence and causality to simplify the specification of these entities.

We have shown how this framework, together with traditional circumscription, appears to provide a satisfactory solution to five major problems in planning and reasoning about action; namely, the combinatorial, precondition-qualification, frame-qualification, consistency-constraint, and overcommittment problems. Indeed, the approach described here, together with means for determining event interference and for specifying processes [4], satisfies nine of the ten requirements for a theory of action as given by Shoham [21]. The only requirement not met is that concerning continuous change, which the proposed theory does not attempt to address. Moreover, the manner of describing actions and events is quite natural and the use of circumscription seems to correspond closely to much of commonsense reasoning.

However, while we have here used circumscription as a means of making reasonable assumptions about an incompletely specified domain, the underlying theory has a clearly defined semantics and is independent of how we choose to use it. In many cases, we may not have to make any assumptions about the problem domain whatsoever. For example, in planning a possible course of action it is often most sensible to consider only idealizations of the problem domain. Such domains can often be readily described using general axioms about independence and causality, especially if the domain is structured appropriately [4,11]. We may then be able to prove all the results we need without recourse to any form of assumption making – reasoning about plans and actions does not have to be nonmonotonic.

If we do have to make assumptions about the problem domain, or ease the specification of certain predicates, circumscription appears to be a powerful tool. However, even in this case, there are other options. For example, it may sometimes be preferable to use domain-specific rules for defining which assumptions are appropriate or, alternatively, to employ information-theoretic approaches based on quantitative probabilities of event occurrences.

Acknowledgments

I wish to thank Randy Goebel, David Israel, Amy Lansky, Vladimir Lifschitz, and Martha Pollack for some very enlightening discussions and for their critical reading of this paper.

References

[1] J. F. Allen. Towards a general theory of action and time. *Artificial Intelligence*, 23:123–154, 1984.

[2] T. Dean. Planning and temporal reasoning under uncertainty. In *Proceedings of the IEEE Workshop of Knowledge-Based Systems*, Denver, Colorado, 1984.

[3] J. J. Finger. *Exploiting Constraints in Design Synthesis.* PhD thesis, Stanford University, Stanford, California, 1986.

[4] M. P. Georgeff. Actions, processes, and causality. In *Reasoning about Actions and Plans: Proceedings of the 1986 Workshop*, Morgan Kaufmann, Los Altos, California, 1987.

[5] A. I. Goldman. *A Theory of Human Action.* Prentice-Hall, Englewood Cliffs, New Jersey, 1970.

[6] S. Hanks and D. McDermott. Default reasoning, nonmonotonic logics, and the frame problem. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, pages 328–333, Philadelphia, Pennsylvania, 1986.

[7] J. Haugeland. *Artificial Intelligence: The Very Idea.* MIT Press, Cambridge, Massachusetts, 1985.

[8] P. J. Hayes. The frame problem and related problems in artificial intelligence. In Elithorn A. and D. Jones, editors, *Artificial and Human Thinking*, pages 45–59, Jossey-Bass, 1973.

[9] D. Israel. What's wrong with non-monotonic logic. In *Proceedings of the Third National Conference on Artificial Intelligence*, Stanford, California, 1980.

[10] A. L. Lansky. Localized representation and planning methods for parallel domains. In *Proceedings of the National Conference on Artificial Intelligence*, Seattle, Washington, 1987.

[11] A. L. Lansky. A representation of parallel activity based on events, structure, and causality. In *Reasoning about Actions and Plans: Proceedings of the 1986 Workshop*, Morgan Kaufmann, Los Altos, California, 1987.

[12] V. Lifschitz. Formal theories of action. In *The Frame Problem in Artificial Intelligence: Proceedings of the 1987 Workshop*, Morgan Kaufmann, Los Altos, California, 1987.

[13] J. McCarthy. Applications of circumscription to formalizing common sense knowledge. In *Proceedings of the AAAI Non-Monotonic Reasoning Workshop*, pages 295–324, 1984.

[14] J. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, 4:463–502, 1969.

[15] D. McDermott. A temporal logic for reasoning about processes and plans. *Cognitive Science*, 6:101–155, 1982.

[16] D. Nute. *LDR: A logic for Defeasible Reasoning.* Technical Note 01-0013, Advanced Computational Methods Center, University of Georgia, Athens, Georgia, 1986.

[17] R. Pelavin. *A Formal Logic for Planning with a Partial Description of the Future.* PhD thesis, Department of Computer Science, University of Rochester, Rochester, New York, forthcoming.

[18] D. L. Poole, R. G. Goebel, and R. Aleliunas. *Theorist: a logical reasoning system for defaults and diagnosis.* Springer-Verlag, New York, New York, 1986.

[19] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[20] Y. Shoham. Chronological ignorance: time, nonmonotonicity, necessity and causal theories. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, pages 389–393, Philadelphia, Pennsylvania, 1986.

[21] Y. Shoham. Ten requirements for a theory of change. *New Generation Computing*, 3:467–477, 1985.