

## REASONING ABOUT ACTION USING A POSSIBLE MODELS APPROACH

Marianne Winslett  
Computer Science Department  
University of Illinois  
Urbana, IL 61801

**Abstract.** Ginsberg and Smith [6, 7] propose a new method for reasoning about action, which they term a *possible worlds approach* (PWA). The PWA is an elegant, simple, and potentially very powerful domain-independent technique that has proven fruitful in other areas of AI [13, 5]. In the domain of reasoning about action, Ginsberg and Smith offer the PWA as a solution to the *frame problem* (What facts about the world remain true when an action is performed?) and its dual, the *ramification problem* [3] (What facts about the world must change when an action is performed?). In addition, Ginsberg and Smith offer the PWA as a solution to the *qualification problem* (When is it reasonable to assume that an action will succeed?), and claim for the PWA computational advantages over other approaches such as situation calculus.

Here and in [16] I show that the PWA fails to solve the frame, ramification, and qualification problems, even with additional simplifying restrictions not imposed by Ginsberg and Smith. The cause of the failure seems to be a lack of distinction in the PWA between the *state of the world* and the *description of the state of the world*. I introduce a new approach to reasoning about action, called the *possible models approach*, and show that the possible models approach works as well as the PWA on the examples of [6, 7] but does not suffer from its deficiencies.

### 1. Introduction

The possible worlds approach (PWA) is a powerful mechanism for incorporating new information into logical theories. The PWA has been studied in various guises by philosophers interested in belief revision and scientific theory formation ([8, 11, 12], and many others), by database theorists [1, 2, 14], and by AI researchers [13, 5]. The PWA philosophy of theory revision can be summed up as:

*To incorporate a set  $S$  of formulas into a theory  $T$ , take the maximal subset  $T'$  of  $T$  that is consistent with  $S$ , and add  $S$  to  $T'$ .<sup>1</sup>*

The elegance and simplicity of the PWA are offset by the fact, illustrated in Sections 4 through 6, that the PWA does not solve the frame, ramification, or qualification problems. The cause of the failure seems to be a lack of distinction between the *state of the world* and the *description*

*of the state of the world*. In particular, the frame principle says that as little as possible changes in the world when an action is performed. The PWA translates this into “as little as possible in the description of the world changes when an action is performed.” Unfortunately, a minimal change in the world does not necessarily correspond to a minimal change in the description of the world, and vice versa. This confusion gives the PWA a morbid sensitivity to the syntax of the description of the world, and leads to incorrect handling of incomplete information.

The philosophy of the possible models approach (PMA) is quite similar to that of the PWA; the essential difference is that under the PMA the *models* of  $T$ , rather than the formulas in  $T$ , are to be changed as little as possible in order to make  $S$  true. My goal in introducing the PMA is to produce a methodology that is as elegant, simple, and intuitively satisfying as the PWA, but which will produce correct results in the fashion of that plodding, awkward, unstructured workhorse, monotonic situation calculus [4].

In Section 2, I sketch a simple action scenario that will serve as an example throughout the remainder of the paper. Section 3 presents the possible models approach. In Sections 4 and 5, I show how the PWA fails to solve the frame and ramification problems, respectively, and show that the PMA does not suffer from the anomalies of the PWA. Section 6 discusses problems with the PWA treatment of multiple candidate result theories. Section 7 describes additional results concerning the PWA and PMA that are discussed in the full version of this paper [16].

### 2. An Example Action Scenario

In reasoning about action, it becomes clear that some formulas of  $T$ —for example, those stating inviolable properties of the physical world—should be designated as *protected*, in the sense that they should always be present in  $T'$ . For example, in trying to move a block to a position already occupied by another block, it is not reasonable to remove the PWA axiom stating that only one object can occupy any given position. Equivalently, no model produced by the PMA should have two objects occupying the same position. In this paper, I will assume that any formulas we try to add to  $T$  are consistent with the protected formulas of  $T$ , as otherwise the action is undefined.

Imagine Aunt Agatha's living room: two ventilation ducts on the floor, a bird cage, a newspaper, a television, and a magazine. The bird cage, newspaper, TV, and magazine must be either on the floor or on the ducts. Only one

<sup>1</sup> As it stands, this is not a complete description of the incorporation operation, because there may be more than one subset  $T'$  enjoying the maximality property, or there may be none. I will return to this point in Section 6; for now, we will only consider the case where there is a unique choice for  $T'$ .

object fits on a duct at a time, and if an object is on a duct, then the duct is blocked. If both ducts are blocked, then the room becomes stuffy. This living room is described by the following protected formulas, adapted from [7], which will be part of  $\mathcal{T}$  throughout this paper:

$$\begin{aligned}
\text{duct}(x) &\leftrightarrow [x=\text{duct1} \vee x=\text{duct2}] \\
\text{location}(x) &\leftrightarrow [\text{duct}(x) \vee x=\text{floor}] \\
[\text{on}(x, y) \wedge \text{on}(x, z)] &\rightarrow y=z & (1) \\
[\text{on}(x, y) \wedge \text{on}(z, y)] &\rightarrow [z=x \vee y=\text{floor}] & (2) \\
[\text{duct}(d) \wedge \exists x \text{on}(x, d)] &\leftrightarrow \text{blocked}(d) & (3) \\
[\text{blocked}(\text{duct1}) \wedge \text{blocked}(\text{duct2})] &\leftrightarrow \text{stuffy}(\text{room}) & (4) \\
\text{on}(x, y) &\rightarrow [\text{location}(y) \wedge \neg \text{location}(x)] & (5) \\
\exists y \text{on}(x, y) \vee \text{location}(x) & & (6)
\end{aligned}$$

These formulas are the mental model that Aunt Agatha has of her living room. Agatha herself is off washing dishes in the kitchen; her faithful robot servant, Tyro, will carry out any actions that she requests. In particular, Tyro is capable of moving living room objects from one spot to another. Note that formula (5) implies that no stacking of objects is permitted in Agatha's living room.

### 3. Definition of the Possible Models Approach

The PMA considers the possible states of the world to be the models of  $\mathcal{T}$ . To reason about the effect of performing an action with postconditions  $S$ , the PMA considers the effect of the action on each possible state of the world, that is, on each model of  $\mathcal{T}$ . The PMA changes the truth valuations of the atoms in each model as little as necessary in order to make both  $S$  and the protected formulas of  $\mathcal{T}$  true in that model. The possible states of the world after the action is performed are all those models thus produced.

This description is rather informal; for example, exactly what are models, and what constitutes a minimal change in a model? As do Ginsberg and Smith, we make a Herbrand universe assumption, so that models are simply subsets of the Herbrand base. The definitions and examples of this paper also carry over to the non-Herbrand case.

Let us say that models  $\mathcal{M}_1$  and  $\mathcal{M}_2$  differ on an atom  $\alpha$  if  $\alpha$  appears in exactly one of  $\mathcal{M}_1$  and  $\mathcal{M}_2$ . We can now formally define  $\text{Incorporate}(S, \mathcal{M})$ , the set of models produced by incorporating  $S$  into  $\mathcal{M}$ .

Let  $\mathcal{M}$  be a model of  $\mathcal{T}$  and let  $S$  be a set of formulas.  $\text{Incorporate}(S, \mathcal{M})$  is the set of all models  $\mathcal{M}'$  such that

- (1)  $S$  and the protected formulas of  $\mathcal{T}$  are true in  $\mathcal{M}'$ .
- (2) No other model satisfying (1) differs from  $\mathcal{M}$  on fewer atoms, where "fewer" is defined by set inclusion.

The possible states of the world resulting from applying an action with postconditions  $S$  are given by

$$\bigcup_{\mathcal{M} \in \text{Models}(\mathcal{T})} \text{Incorporate}(S, \mathcal{M}).$$

Note that the PMA semantics depends only on the models of  $\mathcal{T}$ , and not, beyond the division of formulas

into protected and unprotected statements, on the formulas used to describe those models; the PMA is syntax-independent.<sup>2</sup>

### 4. The Frame Problem

This section illustrates the difficulties that the PWA encounters with the frame problem. To sum up the conclusions of this section, difficulties arise because the frame problem cannot be solved by simply requiring that the changes made in  $\mathcal{T}$  as the result of an action be minimal.

**Example 1.** As an initial description of the state of the world, consider the set of unprotected formulas for  $\mathcal{T}$

$$\text{on}(\text{TV}, \text{duct1}) \quad \text{on}(\text{birdcage}, \text{duct2}) \quad \text{on}(\text{magazine}, \text{floor}).$$

Note that the whereabouts of the newspaper are not explicitly known. Ginsberg and Smith intend the PWA for use as a general mechanism for reasoning about action, in which incomplete information and non-atomic formulas are expected to occur. In this particular case, however, by the implicit Herbrand universe assumption that Ginsberg and Smith make, and by the protected formulas (2) and (6) given for "on", it follows that the newspaper must be on the floor<sup>3</sup>. In other words, the state of the world is completely determined by the information in  $\mathcal{T}$ .

Suppose Agatha now asks Tyro to move the TV to the floor. Under the PWA, necessary preconditions and postconditions for operator application are not represented in  $\mathcal{T}$ ;<sup>4</sup> rather, that information is kept separately. For example, preconditions for  $\text{move}(x, y)$  are that  $y$  be the floor,  $y$  be clear, or that  $x$  already be on  $y$ :  $\text{on}(x, y) \vee \neg \text{on}(z, y) \vee y=\text{floor}$ . The set of postconditions is  $\{\text{on}(x, y)\}$ . As do Ginsberg and Smith in [7], we will assume that the "move" action is unqualified, in the sense that it is guaranteed to succeed if the preconditions for "move" logically follow from  $\mathcal{T}$ .

In order, then, to reason about the effect of moving the TV to the floor, it suffices to incorporate the move postconditions  $\{\text{on}(\text{TV}, \text{floor})\}$  into  $\mathcal{T}$ . The result is the set of unprotected formulas

$$\text{on}(\text{TV}, \text{floor}) \quad \text{on}(\text{birdcage}, \text{duct2}) \quad \text{on}(\text{magazine}, \text{floor}).$$

The frame principle seems to tell us that the newspaper should still be where it was, i.e., on the floor, but this does not follow from the new theory; the newspaper may have flitted to duct1 when the TV was removed, according to the new theory. To see that this is not objectionable, imagine that duct1 and duct2 ventilate the room by powerfully sucking air in through the windows. In this case, the vacuum caused by moving the TV to the floor might

<sup>2</sup> Syntax dependence would not be a flaw in the PWA if there were some means of "normalizing" the syntactic form of  $\mathcal{T}$  so that the intuitively desirable effects of an action would be obtained. Unfortunately, the simple examples of this paper and [16] suggest that no such set of normalization guidelines exists.

<sup>3</sup> Technically, the newspaper must be mentioned somewhere else in the theory for this analysis to hold.

<sup>4</sup> This may be viewed as an epistemological deficiency of the PWA; however, it will not concern us here.

well result in the newspaper flying to duct1. Rather, the problem is that *only* the newspaper can have changed position. Both the magazine and the paper were lying on the floor; why should only the newspaper be affected by moving the TV?

The problem is that the PWA assumes that the frame problem will be solved by making a *minimal change* in the formulas of  $\mathcal{T}$ . Minimality is therefore measured only by the effect of a change on the formulas present in  $\mathcal{T}$ , rather than by considering the effect of a change on the world itself. Unfortunately, considering only the formulas of  $\mathcal{T}$  confers second-class status upon those formulas that can be derived from  $\mathcal{T}$ , such as the location of the newspaper, and also makes the PWA too fond of, too reluctant to retract, the formulas present in  $\mathcal{T}$ .

One might think that the anomaly of Example 1 could be prevented by applying the PWA not to  $\mathcal{T}$ , but rather to the *set of all logical consequences* of  $\mathcal{T}$ . Unfortunately, as pointed out in [5], this approach generates intuitively wrong answers. The theorem presented in Section 6 shows that if a formula  $\alpha$  is inconsistent with  $\mathcal{T}$ , then in adding  $\alpha$  to the logical closure of  $\mathcal{T}$  under the PWA one must remove essentially *all* unprotected formulas of  $\mathcal{T}$ ! Note also that the use of reason maintenance techniques, suggested in [5], would not suffice to eliminate the anomaly of Example 1.

What does the PMA do with Example 1? The model of  $\mathcal{T}$  is<sup>5</sup>

on(TV, duct1)	on(birdcage, duct2)
on(magazine, floor)	on(newspaper, floor)
blocked(duct1)	blocked(duct2) stuffy(room).

The PMA agrees with the PWA that when the TV is moved, the other objects can stay where they are, or the newspaper or the magazine can fly to duct1. In addition, under the PMA the *bird cage* can move from duct2 to duct1, and the resulting void at duct2 can be left open or filled by the newspaper or the magazine. (To see this, recall that the protected formulas of  $\mathcal{T}$  must be true in every result model. The protected formulas (3) and (4), governing “stuffy” and “blocked”, are key players in computing the result models.) The six result models are:

on(TV, floor)	on(TV, floor)	on(TV, floor)
on(birdcage, duct2)	on(birdcage, duct2)	on(birdcage, duct2)
on(magazine, floor)	on(magazine, floor)	on(magazine, duct1)
on(newspaper, floor)	on(newspaper, duct1)	on(newspaper, floor)
blocked(duct2)	blocked(duct1)	blocked(duct1)
	blocked(duct2)	blocked(duct2)
	stuffy(room)	stuffy(room)
on(TV, floor)	on(TV, floor)	on(TV, floor)
on(birdcage, duct1)	on(birdcage, duct1)	on(birdcage, duct1)
on(magazine, duct2)	on(magazine, floor)	on(magazine, floor)
on(newspaper, floor)	on(newspaper, duct2)	on(newspaper, floor)
blocked(duct1)	blocked(duct1)	blocked(duct1)
blocked(duct2)	blocked(duct2)	
stuffy(room)	stuffy(room)	

Are these extra models intuitively acceptable? As Ginsberg and Smith point out, the physics of the ducts

<sup>5</sup> For brevity, I do not list the “location” and “duct” atoms in the example models of this paper.

is unspecified by the protected formulas of  $\mathcal{T}$ ; nothing in  $\mathcal{T}$  indicates that changes of location should be minimized in preference to changes in stuffiness, hence one cannot eliminate the unwanted models using vanilla PMA. The PWA does not have any semantic means of eliminating these models either; they were only eliminated under the PWA because the location of the bird cage was explicitly stated in  $\mathcal{T}$ , as opposed to being derivable. In other words, the physics knowledge needed to keep the bird cage from moving was encoded syntactically into the PWA theory, rather than being stated declaratively.

One can, however, specify preferences for minimizing certain PMA predicates in a manner analogous to prioritization in circumscription. For example, suppose the physics of the living room is such that changes of location are minimized in preference to changes in blockage and stuffiness. If this is done, then the sole minimally-changed model in which on(TV, floor) and all protected formulas are true is the intuitively desirable model:

on(TV, floor)	on(birdcage, duct2)
on(magazine, floor)	on(newspaper, floor) blocked(duct2).

Reference [16] includes a formal definition of prioritization under the PMA.

Prioritization can be applied to the PWA by preferentially removing certain formulas. However, it is not obvious how to establish a correct a priori ordering on *formulas* rather than predicates, without severely restricting the formulas that can appear in the unprotected section of  $\mathcal{T}$ . Further, prioritization cannot prevent the anomaly of Example 1, because the troublesome fact  $\neg$ on(newspaper, duct1) was never present in  $\mathcal{T}$ .

The remaining examples of this paper do not make use of the magazine; for that reason, let us assume that Tyro has removed the magazine from the living room, and it ceases to exist from the viewpoint of  $\mathcal{T}$ .

The frame problems of the PWA are exacerbated in the presence of incomplete information. Even if the locations of all objects are known initially, anomalies will occur if Agatha makes abstract requests. For example, a useful robot *should* be able to deal with requests like “Take the top off the toothpaste,” although performing this request will make the location of the toothpaste top uncertain. As Example 1 illustrates, this type of incomplete information can lead to anomalies.

## 5. The Ramification Problem

The PWA fails to solve the frame problem, so it cannot solve the ramification problem. However, Example 1 was benign in the sense that the PWA failed to draw certain desirable conclusions about the state of the world after an action was performed, but did not draw any false conclusions about the intuitively correct state of the world: the PWA was weak but did not lie. Example 2 shows that the PWA can lie in the presence of incomplete information.

**Example 2.** Aunt Agatha, still working in the kitchen, remembers that the TV overheats and turns off unless it

gets extra ventilation. She asks Tyro to put the TV on one of the ducts. If her initial set of unprotected formulas is

$$\text{on}(\text{TV}, \text{floor}) \quad \text{on}(\text{birdcage}, \text{floor}) \quad \text{on}(\text{newspaper}, \text{floor}),$$

and her request is modeled as the postcondition  $\text{on}(\text{TV}, \text{duct1}) \vee \text{on}(\text{TV}, \text{duct2})$ , then the result theory is

$$\begin{array}{l} \text{on}(\text{TV}, \text{duct1}) \vee \text{on}(\text{TV}, \text{duct2}) \\ \text{on}(\text{birdcage}, \text{floor}) \quad \text{on}(\text{newspaper}, \text{floor}). \end{array}$$

Next she remembers that she can't see the TV from the couch if the TV is on duct2, and she asks Tyro to put the TV on duct1. Incorporating  $\text{on}(\text{TV}, \text{duct1})$  into  $\mathcal{T}$  produces the new set of unprotected formulas

$$\text{on}(\text{TV}, \text{duct1}) \quad \text{on}(\text{TV}, \text{duct1}) \vee \text{on}(\text{TV}, \text{duct2}) \quad \text{on}(\text{birdcage}, \text{floor}) \quad \text{on}(\text{newspaper}, \text{floor}).$$

(Note that the formula  $\text{on}(\text{TV}, \text{duct1}) \vee \text{on}(\text{TV}, \text{duct2})$  is still part of the theory.) Then Agatha remembers that the heat from duct1 melts the little plastic feet on the TV, and she asks Tyro to take the TV off duct1 (using a "remove" action, discussed in [16]):

$$\begin{array}{l} \text{on}(\text{TV}, \text{duct1}) \vee \text{on}(\text{TV}, \text{duct2}) \\ \neg \text{on}(\text{TV}, \text{duct1}) \quad \text{on}(\text{birdcage}, \text{floor}) \quad \text{on}(\text{newspaper}, \text{floor}), \end{array}$$

which logically implies that the TV is on duct2, when intuitively the TV could be anywhere but on duct1! The PWA has led Agatha to a false conclusion, by being too reluctant to retract a formula in the face of new information.

What does the PMA do with Example 2? Because Agatha's theory does not include physical principles, under the plain PMA the objects in her living room could be nearly anywhere after her series of requests. If instead changes in the location of objects are minimized in preference to changes in other predicates, then the two final PMA models are

$$\begin{array}{ll} \text{on}(\text{TV}, \text{duct2}) & \text{on}(\text{TV}, \text{floor}) \\ \text{on}(\text{birdcage}, \text{floor}) & \text{on}(\text{birdcage}, \text{floor}) \\ \text{on}(\text{newspaper}, \text{floor}) & \text{on}(\text{newspaper}, \text{floor}). \\ \text{blocked}(\text{duct2}) & \end{array}$$

## 6. Multiple Extensions

Example 3 illustrates a PWA anomaly that arises when more than one possible world can result from an action.

**Example 3.** Suppose that the unprotected formulas of  $\mathcal{T}$  are

$$\begin{array}{ll} \text{on}(\text{TV}, \text{floor}) & \text{on}(\text{newspaper}, \text{duct2}) \\ \text{on}(\text{birdcage}, \text{floor}) & \neg \text{stuff}(\text{room}). \end{array}$$

Ginsberg and Smith show that moving the TV to duct1 leads to two candidate PWA result theories: one in which the newspaper flies off duct2 and the room remains unstuffy, and one in which the newspaper stays put and the room becomes stuffy. As mentioned earlier, these two possibilities are both reasonable.

Suppose, however, that  $\neg \text{stuff}(\text{room})$  is *not* present in  $\mathcal{T}$  initially. It is, of course, still derivable from  $\mathcal{T}$ , yet moving the TV to duct1 now gives only one candidate result theory:

$$\text{on}(\text{TV}, \text{duct1}) \quad \text{on}(\text{newspaper}, \text{duct2}) \quad \text{on}(\text{birdcage}, \text{floor}).$$

Once again, the decision to represent a fact explicitly rather than to have it merely derivable has had a major impact on the meaning of an action.

What does the PMA do with Example 3? Whether  $\neg \text{stuff}(\text{room})$  is included in  $\mathcal{T}$  or not, the two intuitively desired result models are produced:

$$\begin{array}{ll} \text{on}(\text{TV}, \text{duct1}) & \text{on}(\text{TV}, \text{duct1}) \\ \text{blocked}(\text{duct1}) & \text{blocked}(\text{duct1}) \\ \text{on}(\text{newspaper}, \text{duct2}) & \text{on}(\text{newspaper}, \text{floor}) \\ \text{on}(\text{birdcage}, \text{floor}) & \text{on}(\text{birdcage}, \text{floor}). \\ \text{blocked}(\text{duct2}) & \\ \text{stuff}(\text{room}) & \end{array}$$

Ginsberg and Smith propose that the unprotected formulas resulting from actions with ambiguous results be those formulas that appear in every candidate result theory. In other words, one takes the intersection of all candidate result theories, giving in the case of Example 3 the set of unprotected formulas

$$\text{on}(\text{TV}, \text{duct1}) \quad \text{on}(\text{birdcage}, \text{floor}).$$

As shown in [14, 16], despite certain advantages, this "when in doubt throw it out" philosophy has the fatal flaw of performing extra formula deletions. That is, one knows progressively less and less about the state of the world, as intuitively true propositions become unprovable. Since the PWA behaves poorly in the presence of incomplete information, the method chosen for dealing with multiple candidate result theories should do no more deletions than absolutely necessary. Other approaches to the multiple extension problem are discussed in [1, 2, 14].

There is an occasion, however, when "when in doubt throw it out" might make good sense: if the result theory is defined as the intersection of the logical consequences  $\text{Cn } \mathcal{T}_i$  of all the candidate result theories  $\mathcal{T}_i$ . If there is a finite number of candidate result theories, then the models of the result theory  $\mathcal{T}'$  will be exactly  $\cup_i \text{Models}(\mathcal{T}_i)$ , so no information is lost. As suggested in Section 4, it would seem that good results could be obtained by using  $\text{Cn } \mathcal{T}$  instead of  $\mathcal{T}$  and taking the "when in doubt throw it out" approach when multiple candidate result theories arise. Unfortunately, the following theorem from [16], an extension of Theorem 3 of [1], shows that almost all information will be lost if the postcondition of an action is inconsistent with  $\mathcal{T}$ . In particular, all formulas of  $\mathcal{T}$  will be removed except those that are consequences of  $\alpha$  and the protected formulas of  $\mathcal{T}$ .

**Theorem.** *Let  $\alpha$  be a formula and let  $\mathcal{T}$  be a consistent theory with the set of protected formulas  $P$ , such that  $\alpha$  and  $P$  are consistent. Then under the PWA, the result of incorporating  $\alpha$  into  $\text{Cn } \mathcal{T}$  has the same models as*

- $\mathcal{T} \cup \{\alpha\}$ , if  $\alpha$  is consistent with  $\mathcal{T}$ ;
- $P \cup \{\alpha\}$ , if  $\alpha$  is inconsistent with  $\mathcal{T}$ .

The anomalies associated with multiple candidate result theories do not arise under the PMA, as the models resulting from a PMA action are exactly the union of all candidate models.

## 7. Additional Results

The full version of this paper [16] also shows that the PWA fails to solve the qualification problem, when we drop the

assumption that an action is guaranteed to succeed if its preconditions are satisfied. In particular, in [6] Ginsberg and Smith propose techniques for dealing with “minor” preconditions, those which we are willing to assume are satisfied in the absence of clear information to the contrary. An action is *qualified* if it follows from  $T$  that execution of an action must fail because of a minor precondition. In [16], I show via counterexamples that the algorithm given for testing qualification in [6] will produce incorrect results.

The PMA has a circumscriptive flavor: it is model-theoretic, uses set inclusion as a measure of minimality, and uses priorities. The full version of this paper [16] shows that the PMA has close ties to pointwise circumscription [9]. In particular, minimizing the changes made to a model under the PMA can be rephrased as minimizing the extent of an appropriately defined “changes” predicate, using a new generalization of pointwise circumscription called *setwise circumscription* [15]. The relationship of setwise circumscription to the PMA is spelled out in [15, 16].

The full version of this paper [16] also describes three potential weaknesses of the PMA:

*Language dependence and measures of minimality.* The PMA does not suffer from the syntax-dependence anomalies of the PWA. However, the PMA is still language-dependent, in the sense that the PMA is affected by the choice of language used to describe the world. This reflects the PMA assumption that the possible states of the world are the models of a theory, and therefore a minimal change in the world is a minimal change in a model. Reference [16] illustrates this point with an example, and argues that this drawback is not likely to prove important in practice.

*Standards of correctness.* To have faith in the PMA as a means of reasoning about action, one must show that the PMA is sound and complete with respect to some formal theory of the meaning of actions. As I am no philosopher, I have no theory of the meaning of actions, and it would seem that general proofs of correctness lie out of reach. Indeed, this paper is just as likely as [6, 7] to suffer from hidden epistemological inaccuracies. I do, however, have faith that a given situation can be laboriously but correctly encoded in monotonic situation calculus, and that the PMA can be tested for correctness by comparison with the monotonic encoding. This work is now under way.

*Algorithms.* The PWA may not always do the right thing, but at least there is a simple procedure<sup>6</sup> for reasoning about PWA actions. Algorithms are only known for special cases of the PMA, and it is too early to say whether the PMA will prove amenable to algorithmization in common applications. If good algorithms are not forthcoming for the PMA, it can still serve, to the extent that it is proven correct, as a standard of correctness for more easily computed methods of reasoning about action. Then the

<sup>6</sup> An important point to remember here is that the procedure for the PWA [7] is not an algorithm in the technical sense of the term: the problem that the procedure addresses is *not semi-decidable*. In other words, there cannot be an algorithm for the PWA that always gives correct answers and never goes into an infinite loop.

tradeoffs and inaccuracies introduced by the more efficient approaches can at least be identified and understood.

## 8. Conclusions

The problems with the possible worlds approach (PWA) stem from its differential treatment of explicitly stated and derivable information. As shown here (and in more detail in the full version of this paper [16]), anomalies quickly creep in if the PWA is forced to operate with incomplete information.

The possible models approach (PMA) has as elegant a definition as does the PWA. The PMA behaves well in the presence of incomplete information, and is oblivious to the distinction between derived and explicitly represented information. However, there is an algorithm that approximates the PWA, and algorithms are only known for special cases of the PMA.

Finally, the PMA is a special case of a new type of circumscription. The relationship between the PMA and circumscription, and the difficulties encountered in attempting to use pointwise circumscription for reasoning about action, are explored in [15, 16].

## Acknowledgments

I would like to thank Matt Ginsberg, H. Hirsh, V. Lifschitz, M. Mitchell, R. Reiter, R. Thomason, and M. Vardi for their helpful comments. Matt’s counterexamples were particularly valuable.

## References

- [1] R. Fagin, J. D. Ullman, and M. Y. Vardi, “On the Semantics of Updates in Databases,” *Proc. ACM PODS*, April 1983.
- [2] R. Fagin, G. M. Kuper, J. D. Ullman, and M. Y. Vardi, “Updating Logical Databases,” *Advances in Computing Research* 3, 1986.
- [3] J. J. Finger, *Exploiting Constraints in Design Synthesis*, PhD thesis, Dept. of Computer Science, Stanford University, 1987.
- [4] M. Georgeff, “Many Agents Are Better Than One,” *Proc. 1987 Wkshp on the Frame Problem in AI*, Lawrence, 1987.
- [5] M. Ginsberg, “Counterfactuals”, *AIJ*, 30:1, 1986.
- [6] M. Ginsberg and D. E. Smith, “Possible Worlds and the Qualification Problem”, *AIJ*, to appear.
- [7] M. Ginsberg and D. E. Smith, “Reasoning About Action I: A Possible Worlds Approach”, in *Readings in Nonmonotonic Reasoning*, M. Ginsberg, ed., Morgan Kaufmann, Los Altos, 1987.
- [8] P. Gärdenfors, “Conditionals and Changes of Belief”, from *The Logic and Philosophy of Scientific Change*, in *Acta Philosophica Fennica*, 30:2–4, 1978.
- [9] V. Lifschitz, “Pointwise Circumscription”, in *Readings in Nonmonotonic Reasoning*, M. L. Ginsberg, ed., Morgan Kaufmann, Los Altos, 1987.
- [10] V. Lifschitz, “On the Declarative Semantics of Logic Programs with Negation”, in *Readings in Nonmonotonic Reasoning*, M. L. Ginsberg, ed., Morgan Kaufmann, Los Altos, 1987.
- [11] G. Oddie, “Verisimilitude”, from *The Logic and Philosophy of Scientific Change*, in *Acta Philosophica Fennica*, 30:2–4, 1978.
- [12] J. Pollack, *Subjunctive Reasoning*, Reidel, Dordrecht, 1976.
- [13] R. Reiter, “A Theory of Diagnosis from First Principles,” *AIJ*, April 1987.
- [14] M. Winslett, “A Framework for Comparison of Update Semantics”, *Proc. ACM PODS*, Austin, March 1988.
- [15] M. Winslett, “Setwise Circumscription”, in preparation.
- [16] M. Winslett, “Reasoning About Action Using a Possible Models Approach”, TR UIUC-DCS-R-88-1428, Department of Computer Science, University of Illinois at Urbana, May 1988.