Observations on Observations in Action Theories (Position Paper)*

James M. Crawford
David W. Etherington
CIRL
1269 University of Oregon
Eugene, OR 97403-1269

Abstract

A number of authors have noticed that observations of specific facts about specific states of the world cause problems for existing theories of nonmonotonic reasoning about action. We argue that these problems are pervasive in action theories, and more subtle than previously noticed. Furthermore, we note that the proposed solutions are fundamentally procedural: they involve treating observations in a separate reasoning step. These insights engender a number of open questions about the precise nature of observations in reasoning about action, and how they should be treated.

Introduction

The development of a formal logic for reasoning about change has proven to be surprisingly difficult. We argue that one of the reasons for these difficulties is that knowledge about a changing world seems to come in two distinct modalities: knowledge about how the world "necessarily" behaves, and contingent knowledge derived from observations of the world.

The notion of distinguishing certain specific, contingent, facts about the world—called *observations*—from general knowledge, and treating them specially, has been gaining currency lately in the study of nonmonotonic reasoning about action. However, precisely what constitutes an observation, and why observations should be treated differently, is only partially understood.

There is now widespread agreement that $\neg alive(Shoot(Wait(Load(S_0))))$ is an observation. Similarly, the fact that things work if you treat observations specially which don't work if you don't is arguably sufficient motivation for special treatment. However, as is apparently always the case for theories of reasoning about action, things are not all that simple when more general cases are considered: the decision on whether to treat a particular bit of common-sense knowledge as an observation can be subtle, and have far-reaching consequences.

We consider some of the questions that arise when the role of observations is explored in more detail, provide some additional insight, and conclude that much more work is required before an adequate theory of observations is obtained.[1]

Motivation

Baker [1991] developed an intuitively-appealing formalism for reasoning about action. Essentially, the semantics of his approach amounts to building a state-machine model of the world, together with a mechanism to ensure that all consistent states are considered. He showed that this approach could be used to solve some historically-difficult problems, and it has been widely adopted.

In earlier work, however, we showed that Baker's approach yields pathological behaviours in certain situations [Crawford and Etherington, 1992]. In particular, we showed that explicit observations of unexpected future states produced undesirable results: when reasoning about devices with explicit failure modes, such observations prevented the obvious conclusions from being drawn, and no prioritization scheme could, a priori, remedy the situation. Either the expected results could not be obtained when not contradicted by observations, or the unexpected observations would be explained by "miracles" rather than by "normal" failures.

We argued that the appropriate formal abstraction separates the processes of constructing models of the world (in the sense of the device models of qualitative reasoning), of simulating them over all possible inputs, and of using observations to select the most normal model whose simulation matches

*This work has been supported by the Air Force Office of Scientific Research under contract 92-0693 and by ARPA/Rome Labs under contracts F30602-91-C-0036 and F30602-93-C-00031.

[1] Throughout the paper, we assume passing familiarity with the basic ideas of nonmonotonic approaches to reasoning about action based on the so-called "commonsense law of inertia".

the observations. This process clearly separates the building and simulation of device models from the process of selecting those models whose simulations agree with observed behaviour.

Later, Kartha [1994] showed that observations interact badly with nondeterministic actions in Baker's formalism, resulting in unexpected minimal models. He, too, showed that separating observations from the axiomatization of the world was sufficient to address the pathologies he observed.

Working in a different framework, Sandewall [1989] drew similar conclusions about the need to separate axioms about devices from observations of their behaviour in his trajectory-based semantic theories of action. This indicates that the phenomena is not limited to Baker's formalism.

All of this work has made a compelling prima facie case that the separation of observations about particular outcomes from the theory that describes general principles is an important representational discipline for reasoning about action. However, this distinction (as it is currently realized in practice) is nondeclarative, and little has been said about the principles underlying it: why does it work, and what distinguishes observations from theories?

In the sections that follow, we try to sketch answers to these questions, but argue that there is room for substantial further inquiry.

Why separate observations?

Systems for reasoning about action that rely on minimization of persistence violations to rule out "unnecessary" change (i.e., most nonmonotonic formalisms) can be viewed as running a simulator that makes the changes prescribed by the action descriptions, propagates those effects in ways entailed by the domain theory, and then carries over truth-values for the remaining fluents from previous states.

Observations about future states can interfere with this process. For example, in our "Jersey Drive-By" version[2] of the "Yale Shooting Problem" [Hanks and McDermott, 1986], shooting a person with a loaded gun causes them to cease to be alive, except in the abnormal situation in which you miss. If the minimization of "action abnormalities" (such as missing) is done with lower priority than minimization of persistence abnormalities, then shootings always fail (so that persistences are not interrupted). Clearly this is wrong—actions should not fail just to avoid persistence abnormalities. Conversely, if action abnormalities are minimized with higher priority, then the "normal" case works fine, but when the observation $alive(Shoot(Wait(Load(S_0))))$ is added to the theory, a *persistence* assumption is given up to explain the unexpected result: the only preferred model is one in which the bullets disappear during the *Wait*. In effect the world "steps out of the way"—the model constructed is such that the action sequence $Load, Wait, Shoot$ always produces *alive* (rather than *alive* being the result of a failure of *Shoot* in this particular case). The "carrying over" of *loaded* is prevented, not because of a direct or indirect effect of an action, but to allow for the observation of *alive* at a latter time.

In Kartha's examples [Kartha, 1994], too, conflating observations with axioms allows observations to impose structure on the models that can be considered by the simulator, resulting in pathological simulations. In his first example, a bus ticket is bought and then whichever bus (red or yellow, nondeterministically) shows up is boarded. In this example, knowing you end up on a red bus can make obviously-inappropriate models minimal, for pathological reasons. For example, a model in which a yellow bus shows up, and buying the ticket also "beams" you onto the red bus, would be minimal. (The obvious candidate for a minimal submodel, in which the red bus comes, has an incomparable set of abnormalities: in this model, boarding causes a change from not being on the red bus to being on it, while in the pathological model, boarding doesn't change the world, since you are *already* on board.) The observation that you *are* on the red bus is effectively interpreted to mean you *would* have been on the red bus even if the yellow bus had come, rather than simply ruling out the possibility that the yellow bus came!

Similarly, although much more subtly, Kartha's second example shows that treating observations as axioms can break Baker's automatic generation of so-called "existence of situations" axioms. In particular, not all the situations *allowed* by the axioms will be required by the existence of situations axioms. Thus, the information that an initial sequence of coin flips resulted in H–T–T–H–T will, if treated as a domain axiom, force all models to have structure corresponding to that chain. In some models, such structure can paradoxically use up all the situations constructed by the existence of situations default, without allowing for all consistent fluent-value combinations. Since the strength of Baker's approach lies in considering the complete situation space, this opens the door for as-yet-undetailed anomalies.

These pathological examples seem to hinge on whether observations are allowed to affect the *construction* of state transition graphs (or other parts of the underlying fabric of the models) or only the *selection* of particular models. On the former view, knowing facts about the future can change the way the world "works", either by precluding the exis-

[2]This example actually appears in a slightly different, but isomorphic, form in [Crawford and Etherington, 1992].

tence of certain states or by ruling out certain types of state transitions. Such a priori excision of states or transitions means that, in effect, observations about how the world *is* are interpreted as being about how it *could have been*, which seems inappropriate.

The latter view treats observations as deciding which of the *possible* behaviours the world *actually* exhibited. Knowing that no targets were hit chooses the model where the shooting missed, allowing no magical unloadings; knowing that one boarded the red bus simply means that the red bus came, allowing no miraculous teleportation when tickets are bought; knowing that a particular sequence of heads and tails occurred simply means that chance had its way, not that every model must have exhibited such outcomes.

Several important questions remain open. First, is the separation between domain theories and observations in some way a fundamental distinction in reasoning about action, or just an artifact of our decision to use some state-graph-based semantics? Second, can observations be syntactically recognized, and is there a difference in modality between the "necessary" facts in the domain theory and the "contingent" facts in the observations? Finally, can one write a less-operational (more declarative) semantics that does not require splitting the theory in two, minimizing persistence violations in one part, and then pasting it all back together again? At this point we cannot promise answers to these questions but we hope to shed some light on them by considering some new examples.

Domain Theories vs. Observations

The formalisms suggested by Sandewall, Crawford and Etherington, and Kartha all handle the distinction between domain theories and observations similarly. Essentially, they procedurally break down the problem of reasoning about action into two phases. They first construct all consistent situations and link these situations together with "result" arcs, generating a state graph. If there are sources of uncertainty about the effects of actions (e.g., because the actions have non-deterministic effects, or because the actions have certain effects by default, etc.), then multiple state graphs may be built, perhaps with some preference ordering among them. These state graphs are "platonic" in the sense that particular knowledge (e.g., "it rained on Tuesday") in no way informs their construction—only general domain knowledge is used. In the second phase, knowledge about particular states (i.e., observations) is used to prune the set of state graphs, and/or to select particular paths within state graphs.

The key difference between domain knowledge and observations in these formalisms is, thus, that general domain knowledge is used in building the state graphs, while observations provide selection functions among the results of this model-building phase. As Kartha [1994] points out, "[As] more observation axioms are added to a theory, the class of models of the theory should get smaller monotonically. That is, observation axioms should serve only to rule out some models, never to include new ones." Unless observations are treated separately, however, it appears to be impossible to prevent them from changing (or even expanding) the class of models.

Bacchus *et al* [1994] argue that the success of Baker's approach is due to the fact that it, in some sense, allows for the consideration of *counterfactual* states. (For example, a model in which a gun becomes unloaded during a *Wait* is not (in general) minimal because the fact that

> *If the gun had stayed loaded,* then
> firing it would have killed the target.

remains a property of the state transition graph that is constructed; the model where the gun unloads thus has two persistence abnormalities (the unloading and the counterfactual killing), compared to only one in the preferred model where loaded persists.) Their probabilistic framework for reasoning about action explicitly builds on this notion of counterfactuals. On the other hand, conflating observations and domain theories can expressly preclude some such counterfactual states, and this conflation can be argued to explain the infelicities of Baker's (and others') frameworks.

Recognizing Observations Syntactically

Consider the following domain theory:[3]

$$\forall s.\ aimed(Aim(s)) \tag{1}$$

$$\forall s.\ loaded(Load(s)) \tag{2}$$

$$\forall s.\ aimed(s) \wedge loaded(s) \wedge \neg broken \rightarrow \neg alive(Shoot(s)) \tag{3}$$

This is much like the Yale Shooting Problem, enhanced with the addition of the variable *broken*, which is introduced to make *Shoot* nondeterministic and force the creation of two state graphs: one in which *Shoot* causes $\neg alive$ and one in which *Shoot* is ineffective.[4]

Now assume that, after constructing the set of models for this theory, we add the observation:

$$alive(Shoot(Aim(Load(S_0)))) \ . \tag{4}$$

[3]The first version of this example was suggested to us by Matt Ginsberg.

[4]Obviously, we are assuming that we are using a formalism that holds *broken* fixed in the minimization of persistence abnormalities.

This results in the pruning away of those state graphs in which *broken* is false. Thus if we observe *alive* after *Load*, *Aim*, *Shoot*, we conclude that the gun is broken.

If, instead, we were to add (4) to the domain theory, we get a state graph in which aiming unloads an unbroken gun (otherwise axiom (3) would force $\neg alive$). Thus rather than eliminating state graphs, (4) forces modification of the graphs in which *broken* is false. Examples like this, and the more complex examples referred to above, have led various authors to suggest that facts about specific situations (in this case, the situation $Shoot(Aim(Load(s_0))))$ should in general be treated as observations.

Somewhat more subtly, consider the effects of adding the axiom:

$$\forall s.\ alive(s) \tag{5}$$

to the domain theory. Then (5) can be resolved against axiom (3) to yield:

$$\forall s.\ \neg aimed(s) \vee \neg loaded(s) \vee broken \tag{6}$$

Thus we get directly that aiming unloads unbroken guns. At first this also seems bizarre, but what actually seems to be going on here is that (6) is a reasonable ramification of a peculiar constraint (5).

To see this, consider a similar axiomatization of a lever in which axiom (3) is: $down_1(s) \wedge down_2(s) \wedge \neg broken \rightarrow bent(s)$, with the obvious changes to the other axioms. In this case it is clearly reasonable for $\forall s.\ \neg bent(s)$ to imply that pushing down one side of the lever forces the other side to come up.

Now assume that we add (5) as an observation (perhaps this is meant to be the inductive generalization of a bunch of individual observations—the result of watching too many hours of *Superman*). In this case we get *no* models! This is because an initial state with *alive* false must exist, since the state graph always has a node for all situations consistent with the domain theory at all time points. Even if we modify (5) to $\forall s.\ alive(Shoot(Aim(Load(s))))$ the problem persists, since there will be some initial state with $\neg alive$, and $\neg alive$ will persist since there is nothing in the domain theory to make it change.

Similarly, and perhaps more realistically, if the domain is such that all possible states for some problem can eventually be observed, the resulting set of ground observations will have the same anomalous effects as a universally-quantified observation.

This and similar examples suggest that universally-quantified observations are generally a bad idea. However it still seems likely that one will want to be able to inductively generalize sets of observations.[5] If we make such generalizations and add them to the domain theory, then we clearly will get quite different results than we would have got from the set of observations that led to the inductive generalization.

At this point we cannot present a definitive solution to this problem. We conjecture that a mechanism for writing universally-quantified observations that are relativized to some subset of the set of all situations—perhaps the set of all situations reachable in the "real world" (as opposed to all situations existing in the "platonic world" of the state graph) provides a solution. However, we are not convinced that such a solution is either the most intuitive or the most elegant.

Conclusions and Unanswered Questions

In some sense, our work to date raises more questions than it answers. Available evidence suggests that it is useful to separate observations from the domain theory, and to separate the generation of state graphs from the pruning of the set of state graphs using observations. However, we do not know whether this division is fundamental or is an artifact of current nonmonotonic approaches to reasoning about action.

Evidence also suggests that observations are generally statements about particular states, while axioms in the domain theory are generally universally quantified. However, we do not know whether this is a hard-and-fast rule, and the distinction does not seem semantically well-motivated. Even if it is a general principle, it does not necessarily constitute a complete syntactic characterization of observations, nor does such a separation make all the problems we have noted disappear (c.f. the problems with finite state spaces and "complete" sets of observations mentioned above).

On another view, observations describe the way the world was, but not the way it might have been. They thus have the flavour of a "possibility" modality. This parallel is only suggestive, at present, and what (if any) relationship exists between domain axioms/observations and necessity/possibility is largely unexplored. However, such an treatment at least holds out the possibility that a purely-declarative framework might result, rather than the procedurally-flavoured representation mechanisms (e.g., circumscribe the domain theory, add the observations, and circumscribe again) that seem to arise from attempts to patch existing nonmonotonic frameworks to make them deal with observations adequately.

[5] If for no other reason, learning programs will eventually have to interact with mechanisms for reasoning about action.

Finally, there are likely to be connections between observations in theories of action and observations/evidence in probabilistic approaches, but, here again, the exact relationship between the two is not clear. Goldszmidt and Darwiche [1994a] have developed a causal theory of action in the *Kappa* calculus, which they claim can handle observations correctly [Goldszmidt and Darwiche, 1994b], but we have yet to explore how their approach handles the problematic examples we (and others) have identified.

Bacchus *et al* [1994; 1992] do not provide sufficient details to work through examples such as the Jersey Drive-By in detail [Bacchus, 1995], but their comments (including the result embedding Kartha's language, $\mathcal{A}$, in their framework), suggest that, if nothing else, some care will be required to ensure that axiomatizations like theirs do not fall victim to inadequate protection of the full range of counterfactual states, at least in some cases, in ways like those that pose problems for existing nonmonotonic approaches. Since their representational mechanisms are richer than Kartha's, however, it may be straightforward to ensure this.

While we have not yet been able to answer this question definitively, however, their framework *is* apparently sensitive to the role of observations, which are handled by conditioning. Furthermore, because they are free to assign non-integral probabilities to facts, it may be that the pathological worlds that create havoc for purely logical approaches will have "vanishingly-small" probabilities, and essentially no effect [Bacchus, 1995] in practice.

We expect that such probabilistically-motivated approaches will provide useful insights into questions of exactly how observations should be defined and handled. We expect this symposium to be a good forum for discussing these connections. At this point we can only conclude that this is an interesting and promising line of work, but much of the basic work remains to be done.

Acknowledgements: We would like to thank (but not blame) the members of CIRL, particularly Matt Ginsberg and Tania Bedrax-Weiss, for useful discussions of this material. We would also like to thank Fahiem Bacchus, Daphne Koller, and Adam Grove for last-minute clarification of some details of [Bacchus *et al.*, 1994] in this context. Remaining deficiencies are, of course, the responsibility of the authors.

References

F. Bacchus, A. Grove, J. Halpern, and D. Koller. From statistics to belief. In *Proceedings Tenth National Conference on Artificial Intelligence (AAAI-92)*, pages 602–608, 1992.

F. Bacchus, A. Grove, J. Halpern, and D. Koller. Forming beliefs about a changing world. In *Proceedings Twelfth National Conference on Artificial Intelligence (AAAI-94)*, pages 222–229, 1994.

F. Bacchus. Personal communication, 1995.

A. Baker. Nonmonotonic reasoning in the framework of the situation calculus. *Artificial Intelligence*, 49:5–23, 1991.

J.M. Crawford and D.W. Etherington. Formalizing reasoning about change: A qualitative reasoning approach. In *Proceedings Tenth National Conference on Artificial Intelligence (AAAI-92)*, pages 577–583, 1992.

M. Goldszmidt and A. Darwiche. Action networks: a framework for reasoning about actions and change under uncertainty. In *Working Notes: AAAI-94 Spring Symposium Series in Decision Theoretic Planning.* American Association for Artificial Intelligence, 1994.

M. Goldszmidt and A. Darwiche. Personal communication, 1994.

S. Hanks and D. McDermott. Default reasoning, nonmonotonic logics, and the frame problem. In *Proceedings Fifth National Conference on Artificial Intelligence (AAAI-86)*, pages 328–333, Philadelphia, PA, 1986.

G.N. Kartha. Two counterexamples related to Baker's approach to the frame problem. *Artificial Intelligence*, 69:379–391, 1994.

E. Sandewall. Filter preferential entailment for the logic of action in almost continuous worlds. In *Proceedings Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89)*, pages 894–899, 1989.