ELECTRONIC NEWSLETTER ON
REASONING ABOUT ACTIONS AND CHANGE

Issue 98008 Editor: Erik Sandewall 25.1.1998

Today

We have two discussion items today:

1. Pat Hayes answers to the joint critique by Hector Geffner, David Poole, and Judea Pearl in the previous Newsletter issue
2. Erik Sandewall resumes the discussion with Wolfgang Bibel about his IJCAI-97 invited paper.

Debates

Ontologies for actions and change

From: Pat Hayes

I said:

> *... actions - or external events - do change ones beliefs about the past.... consider for example coming into a room in an empty house and finding a hot cup of coffee resting on a table. One immediately infers that somone else has been present there recently.*

This seems quite obvious and uncontroversial to me. One can be surprised by things that happen in one's world. Observations can, in fact, violate one's model of the world and show that it is faulty and needs to be updated; and these updates can involve changing one's beliefs about the past. However, many seem to disagree:

Hector Geffner:

> *It's important to distinguish* observations *from* actions.

David Poole:

> *No. No. No.* Observations *can change one's beliefs about the past. Doing an action doesn't change beliefs about the past. Observing the hot coffee made you update your beliefs about the past.*

Well, OK. I said 'actions or external events'.

I dont think this sharp distinction between sensing and acting is either necessary or even ultimately coherent, in fact. We are constantly monitoring our own actions in this kind of way, at various levels, so almost every action has observation involved in it. And similarly, it is hard to observe without somehow acting, and a good deal of our planning and acting is motivated by a perceived need to find out more about the world. So observation is intimately bound up with action. Our peripheral systems often blend motor action and lowlevel perception in tight feedback control loops, so that our bodies seem to 'move by themselves', but these lowlevel controls are the result of more cognitive decision-making (deciding to hit a tennis ball, say.)

But even if I agree, for the sake of argument; David's point about observation applies to beliefs about anything, past, present or future. You might predict what is going to happen when you act, but the only way to be sure it will happen is to do it and then take a look. So again, I see no reason why this distinction supports the original claim that we cannot draw new conclusions about the past from observations in the present.

Hector Geffner continues:

> *In dynamic systems the first are usually expressed as "initial conditions" e.g., $x(0) = 5$, loaded(0) = false, etc; while the latter are the inputs to the system. An* observation *at time i ("cup in the table") of course should have an effect on your beliefs at times $j \leq i$*

This seems so peculiar that I wonder if we are talking about the same thing. The whole point of making an observation, surely, is to gain information about the present state, not the initial conditions (which, once one has taken an action, are in the past.) Similarly, I wasnt saying that an observation could change one's past beliefs - nothing can change the past - but rather that it could change one's (present) beliefs *about* the past.

[Regarding Hector's later remarks on nonmonotonicity, I agree with Erik Sandewall's reply in 23.1(98007).]

Judea Pearl:

> **Correction**
>
> *The principle of causality is in fact stronger than Pat's interpretation of it. Not only the past, but also our beliefs about the past, do not change as a result of actions,* **unless** *the acting agent is part of our model. Moreover, if the agent* **is** *part of our model, then actions cease to be interesting and problematic as they are today (at least in some AI approaches to actions). To explain: The action of putting a cup of coffee on the table does not change the state of the coffee or the table before the action, and does not change our beliefs about the state of the coffee before the action.*

Well, I was talking about a later observation of the coffee put there by someone else, not the act of putting it there oneself. But in any case, this second statement again seems to me to be simply false. Suppose that one is walking along in NYC minding ones own business, when suddenly a mugger grabs one from behind and holds a knife to one's throat. Isnt this an example of an "exogenous" act (by an agent not in one's model) causing one to change one's beliefs about a whole lot of things, some of them involving the past, such as that you were alone in the street? (Examples like this are so blindingly obvious that I wonder if we are talking about the same thing??)

I fail to see why having beliefs about other agents (?I presume this is what is meant by an agent being part of a model) means that "actions cease to be interesting". Most of our daily lives are taken up with worrying about what other agents are doing, especially the ones we know a lot about. (Ever had kids to look after, or even a dog?)

> *BUT, Pat will argue: seeing the coffee on the table allows us to infer that "someone else has been present there recently." True, but only if we are concerned about the actor's whereabout and if the limitations or motivations of the action-producing agents are* **in** *the model (e.g., that someone must be in a house to do it, and will probably do it if he/she is thirsty, etc.)*

Well of course. If one knew nothing of people and coffee, one wouldnt come to the conclusion I mentioned. It follows from one's knowledge of such things. That is simply a nonsequitur. Notice however that the inference is that an agent exists, whose existence one had not previously considered. I'm not sure what your doctrine allows, but this doesn't seem quite the same as having that agent in one's model before seeing the cup. Not to mention the mugger.

> *Once this action are perceived as produced by a modeled agent, it is no different than any other events, say the agent tripping or being hit by a car, because then it is licensed to trigger the usual inferences of predictions and abduction that observations trigger.*

Well, all this talk of 'licence' seems like the reiteration of a doctrine of some kind (one I am not familiar with, being an atheist); but leaving that aside, if I can make sense of this at all, then Judea seems to simply be agreeing with me. The point is that (within one's "licence") it is possible to come to new conclusions about the past on the basis of new information about the present. The direction of inference can be opposed to time's arrow. Is that not what we were talking about?

> *Hector said all this by using the term "exogenous". Indeed, the problematic aspects of actions surface when we try to treat actions as exogenous, namely produced by external agents* **not** *modeled in our system. Only by considering an action exogenous can we talk about the world "changing" (anyone ever saw a world truly changing?.)*

Of course I have seen a world changing, in the relevant sense. It happens even if the actions are entirely within my own conceptual scope, eg as when I flick the light switch, confident that the light will in fact come on, and it does indeed come on. That's an observation of a changing world. (Sometimes the light doesnt come on, and I am forced to update my beliefs, often about the past history of the light bulb or the electrical system.) ....

> *One problematic aspect of dealing with external actions is that ordinary abduction, from the immediate consequences of those actions, must be suppressed. For example, we do not want to abduce that it rained when we decide to pour water on the driveway and get it wet.*

I confess to failing completely to follow this point. Why is *my* putting water on my driveway considered an *external* action? External to what?

Pat Hayes

P.S. back to David Poole:

> *P.S. I am wondering why such confusion between observing and acting may have arisen in the first place.*
>
> *I conjecture is that it has to do with the preconditions of actions. For example, the pickup(x) action has the precondition that there is*

> *nothing on x. Then by carrying out the action, can you infer that there was nothing on x?*

No, of course not. the implication runs **from** precondition **to** result, not in reverse. (You might also have a reverse implication if the precondition was necessary as well as sufficient; but then this would be a valid inference to make. Consider for example that the only way to stay alive is to not hit the oncoming truck; you make a wild swerve, survive, and say with a sigh of relief, Thank God I *didnt* hit the truck.)

> *But this doesn't make sense. What happens if I had attempted to pickup x when there was something on it?*

You would have failed, and maybe (if the axiom had been an iff) concluded that there must have been something on the block after all, or at any rate that *something* had prevented it being picked up. It would have been an abnormal state, in McCarthy's middle-period sitcalc using ab-minimisation.

> *What if I didn't know whether there was something on x when I tried to pick it up? It seems that the only sensible interpretation of the precondition is that if there was nothing on x and I carried out pickup(x), then the expected thing would happen. If something was on x and I carried out pickup(x) then who knows what may happen. The role of the precondition is that it is only sensible to attempt to carry out the action when the preconditions hold.*

No no. This is all a *description* of an action. What actually happens when you do the actual action may be much more complicated than your description (your beliefs about the action) are able to predict. Maybe something was there that you didnt know about; maybe your idea of lifting is defective in some way. We can never guarantee that our beliefs are accurate, and still less that they are complete. But whatever actually happens, if you are able to deduce from your beliefs that X should happen at time t, and then when you actually do it, (you observe that) X doesnt happen at time t, you want to be able to recover from this situation without dissolving into inconsistency. Hence the (original) need for nonmonotonic logics.

> *An alternative explanation if the confusion may be that the action pickup(x) is an observation of the effect of my motor control on a particular state of the world. If I carry out a particular motor control when there is nothing on x, then a pickup(x) action arises. When I do the same motor control when there is something on x then some other action arises. Then I do not choose the action, but I only choose the motor control (consider the case when I don't know whether there is something on x or not, and I try to pick it up). Is this what people mean by an action? Surely then it is imperative to distinguish that motor control (for a better description) that I get to choose, from the observation of the effect of that motor control.*

In the sitcalc (any variety), actions are changes in the world, not motor commands. One plans by thinking about the changes, not by thinking about the muscles one is going to use. Putting such a system into a robot requires one to somehow connect these actions with motor controls, no doubt, but they shouldnt be identified. (Murray and Ray, do y'all agree??)

Discussion with Wolfgang Bibel about his IJCAI lecture

From: Erik Sandewall

Dear Wolfgang,

With respect to your IJCAI article, I have a number of questions several of which relate to the research methodology or paradigm being used. I observe that when Marc Friedman ended one of his questions with "why prefer one solution to the other", your answer was

> *To the best of my knowledge Michael Thielscher (...) was the first who gave a solution to the ramification problem which overcomes deficiencies in any previous solution (...) no previous solution would model reality in a correct way. A better solution in this sense must be preferred to a deficient one. (...)*

In this way, you appeal to the traditional method of validating approaches to common-sense reasoning by ways of counterexamples: a method is accepted until disproved by an example. Unfortunately, this research methodology can not provide any reliable conclusions. As has been recognized in core computer science since a long time, the lack of knowledge of a counterexample does not prove that a proposed solution (or a program) is correct.

On this background, I have the following questions or observations:

1. What is your perspective on current research in reasoning about actions and change where we provide *assessments* or *validation results* for new theories, instead of merely proposing new logics based on a combination of intuition and toy examples? The new approach is being used by several of us, including Shoham (who started this trend), Lifschitz, and myself. In spite of being a survey paper, your article does not mention this development at all.

2. In the section on qualification, you describe a solution using transition logic: "TL opens a new way to deal with this problem". The solution is exemplified using McCarthy's old "potato in the tailpipe" problem, but not motivated in any other way. In particular, there is no statement of when the method is correct or is not correct, nor any proof or reference to a proof of its correctness under some precisely stated assumptions.

Unfortunately, however, it is well known that simple solutions to the qualification problem will easily fail if the naive potato and tailpipe example is modified ever so slightly. For example, suppose there are two cars, A and B, it is known that a potato is put into the tailpipe of one of them, and one asks whether car B will start properly. In such a case, the absence of positive knowledge that the tailpipe of car B has been plugged, does not allow one to draw the default conclusion. Or, suppose one does not know whether the plugging of the pipe preceded or succeeded the attempt to start the car; the same difficulty arises again.

Given that previously proposed solutions have failed so easily, one would like to have some evidence that the method proposed in your article is not going to encounter those problems, or other ones which may come up. In other words, here is a case where the shortcomings of the older, example-driven methodology are evident.

My question is what can be said in general about the method you propose for qualification: when is it known to work, and when not?

3. In section 4, you write "LCM has been the first method which actually solved these aspects of the frame problem and did so in the optimally possible way". Since there is no obvious definition for optimality in this respect, I wonder which quality measure do you use, and what is the proof that the LCM method is optimal with respect to it?

4. Another example of an incompletely substantiated claim occurs in your section on ramification. You write "Lifschitz' categorization of fluents does *not* work in this example. We need to *categorize the actions* into primary and secondary ones (rather than the fluents) as done in the solution presented in this section". However, the second sentence does not follow from the first one. The fact that the particular variety of fluent categorization that was proposed by Lifschitz doesn't work for the example, does not prove that **all** fluent categorization methods fail for it, or for some reasonable class of examples. So in what sense do we "need to" categorize the actions?

5. Also with respect to your formalization of Thielscher's example with three switches and a relay, it is remarkable that the electric circuit in question can easily be understood in terms of dependencies and persistence, but the proposed formalization requires the axioms to represent the propagation of *changes*: "if this fluent changes in such-and-such a way, then that fluent also changes in such a way". This seems clumsy and counterintuitive. Do you claim that it is the best possible representation in the present state of the art?

6. As you correctly observe, a rational agent must be capable of reasoning about the timing of actions and about changes within the duration of an action. In section 6.4 of the paper you describe in outline how to introduce timing of actions into TL. However, the most obvious way of arranging this is by approaches that use explicit, metric time, as in Shoham's work in the 1980's, the Features and fluents approach, and the modern event calculus. The PMON(RCs) logic presented by Doherty and Gustavsson at KR-96 is an example of such a logic. To summarize,

- PMON(RCs) uses explicit time, which allows one to characterize the duration of actions and the intermediate changes within its duration;
- PMON(RCs) also allows one to characterize the timing of concurrent actions with overlapping duration periods;
- PMON(RCs) allows delayed effects of actions;
- The circumscription policy used in PMON(RCs) has been shown to be reducible to the first-order case, which insures that standard theorem-proving techniques and their optimizations can be used to compute entailment.

In other words, it does in concrete detail what you only describe in gross outline. My questions are:

1. What are the advantages of your approach over the one I just referred to? (This is a valid question in view of your statement "A better solution in this sense must be preferred to a deficient one").
2. In PMON(RCs) and other approaches that use explicit metric time, it is straightforward to make statements about durations, comparing

durations of actions, and so on, basically because each interpretation represent an entire history of the world. How can this be done in a logic like TL where the => operator takes one from one state to the next, and updating the current time in the process?

7. You write "section 6 shows how the various aspects involved in reasoning about actions and causality can be taken into account within TL". However, nothing in section 6 or elsewhere in the article presents any concrete results about how spontaneous change in the world can be represented - it is as though the world were entirely static when no actions are taken. The resulting concept of causality is quite meagre.

Here again, there *exists* in fact a body of results in this area, ranging from the Situated Action Theory of Morgenstern and Stein and my own early work on integrating differential equations into logic (presented at KR-89 and IJCAI-89) to Murray Shanahan's earlier and recent work which uses a very similar approach, and which has also been used for implementing a simulated robot. When you write "another issue concerns the integration of differential equations and their computation within a logic such as TL", the uninformed reader will not easily guess that the integration occurred nine years ago, although not for TL.

8. In section 4, on the topic of how to obtain a specific plan as a solution to a planning problem, you write "[Bibel 1986a] introduced state literals, $S(x)$, which keep track of the states passed through while executing a plan. (...) By unification the variable denoting the goal situation will then along with a successful proof always provide a term that expresses the linearized sequence of actions." Yes, but how is this different from the use of the Answer predicate which was proposed and used by Cordell Green in the late 1960's?

9. In subsection 6.3 you refer to an "easy solution" for allowing action laws for special cases to take precedence over a more general law for the same action. However, in the example you quote, the specialized case only contains an *additional* effect besides the effect in the general case. Does the same solution also apply if some of the effects of the general case *do not* arise in the special case?

10. If the answer to the previous question is positive, does the planning method that you propose using unification against the state literal (answer literal) still work when such specificity is allowed? The question arises since unification against answer literals results in planning backwards from goals, whereas the solution for specificity is defined in terms of forward simulation of the plan.

11. A final observation: The title and the introduction of your article focusses on planning, as shown when you write: "In this paper I review the state of the art in deductive planning...", but the major part of the contents (after introducing the TL logic as such) deals with problems in reasoning about actions and change: the frame problem, ramification, and so on. May I suggest that some additional coverage of modern results in the latter area would be appropriate, in particular since this is claimed to be a survey article.

Sincerely

Erik Sandewall

This newsletter is issued whenever there is new news, and is sent by auto-

matic E-mail and without charge to a list of subscribers, in plaintext format. The present postscript issue is a bona fide rendering of the same information in alternative typography. Instruction for contributors and other additional information is found at the following URL:

`http://www.ida.liu.se/ext/etai/actions/njl/`