ELECTRONIC NEWSLETTER ON REASONING ABOUT ACTIONS AND CHANGE

Issue 98009 Editor: Erik Sandewall 26.1.1998

Today

The research methodology discussion that resumed a week ago has now shifted to ontological issues concerning actions: "what is the relation between actions and observations?" and "what is an action anyway?" Today's issue contains answers to Pat Hayes by Judea Pearl and Erik Sandewall.

Also, Michael Thielscher communicates a reference to his existing work that addresses the case of nonmonotonicity wrt observations.

Dear Reader, if you feel that some of your own earlier works has a bearing on the topic(s) of the present discussion, then why don't you send us a note for inclusion in the debate? One of the problems in contemporary research is that a lot of good work tends to be overlooked, maybe because so much is being published, and because it's difficult for anyone to have a full overview of everything that's contributed. You can help both your colleagues and yourself by sending us a note that clarifies where you stand on the present issues: what is an action in your system; how do you relate actions to observations; do you agree or disagree with those who participated in this discussion up until now?

Please feel free to advertise your own published works by showing how it is relevant for the topics discussed here. (If you get to be asked hard questions as a result of appearing on this stage, then just remember that all advertising is usually good advertising). - References containing a URL for where the paper can be picked up are of course particularly useful.

Debates

Ontologies for actions and change

From: Michael Thielscher

On 23.1., Erik wrote

> *... but I can't think of any work in mainstream actions and change that has addressed nonmonotonicity with respect to observations in a serious way.*

Well, I at least intended to be serious in my KR'96 paper on "Qualification and Causality" [c-kr-96-51], where I address the Qualification Problem. The latter is inherently concerned with nonmonotonicity wrt. observations– if we view it the way John introduced it, and not oversimplify it to the marginal issue of deriving implicit preconditions from state constraints. The classical example goes as follows. By default, we conclude that the car's engine is running after turning the ignition key. Adding the observation that initially the tail pipe houses a potato, the previous conclusion gets withdrawn. Part of my paper was concerned with a Fluent Calculus axiomatization capable of dealing with (a well-defined aspect of) the Qualification

Problem. A nice feature about the Fluent Calculus is that it provides monotonic solutions to both the Frame and the Ramification Problem. But when it comes to the Qualification Problem, nonmonotonicity is inevitable per definitionem, which is why my Fluent Calculus axiomatization in the KR paper comes with a nonmonotonic feature.

References:

[c-kr-96-51] Michael Thielscher. *Causality and the Qualification Problem.* Proc. International Conf on Knowledge Representation and Reasoning, 1996, pp. 51-62.

From: Erik Sandewall

Michael,

Sorry about that, I expressed myself imprecisely. What I had in mind was in the context of Pat Hayes's contribution:

> *Nonmonotonicity is a good example of the need to revise ones beliefs about the past in the light of unexpected outcomes in the present,...*

which is why I quoted this elaboration of the stolen car scenario. You are of course quite right that for the topic of observation nonmonotonicity as such (not restricted to "backwards in time"), approaches to the qualification problem and in particular your contribution are highly relevant.

What happens in the approach of your KR-96 paper for the case of "backwards in time", such as the stolen car scenario?

Re the potato in tailpipe scenario, see also my question number 2 to Wolfgang Bibel in yesterday's newsletter.

Erik

From: Judea Pearl

On Actions vs Observations, or on Pat Hayes' reply to Geffner, Poole and me.

Well, well, and I thought my tiny correction would end with Pat just replying "Of course, I did not mean it ..." Instead, it now seems that the cleavage between the culture that Hector, David and I represent and the one represented by Pat has gotten so **deep** that we are not even sure we are talking about the same thing.

Pat does not think the "distinction between sensing and acting is either necessary or even ultimately coherent", For him, observing a surprising fact evokes the same chain of reasoning as establishing that fact by external act. In both cases, so claims Pat, the world is changing, because a world is none other but one's beliefs about the world, and these do change indeed in both cases.

I tend to suspect that Pat's position is singular, and that most readers of this newsletter share the understanding that it is **useful** to think about observations as changing beliefs about a static world, and about actions as changing the world itself. I am under the impression that this distinction has become generally accepted among AI researchers (as it has among philosophers and database people, e.g., counterfactual vs indicative conditionals, imaging vs. conditioning, belief updating vs. belief revision, etc. etc.) Otherwise, one would find it hard to understand why so much energy has been devoted in AI to **actions** vis a vis **observations**, why the title of this Newsletter is what it is. Why the Frame Problem does not exist relative to observations (e.g., need frame axions for "I saw **only** x"?), why there is no ramification problem relative to observations (surely every observation has some ramifications) and why people write so many papers on concurrent actions and not on concurrent observations. Or are these observational problems just waiting around the corner to hit us once we solve the corresponding problems with actions?

If Pat's position is singular, I would not wish to bore the rest of the readers with this issue, and I I will pursue it with Pat in the privacy of our screens. I am curious, though, if my suspicion is correct, and I will be waiting readers' feedback.

————-

Those who read my paper in TARK 1996 need not read further. Others may wish to ask when a change is considered a "world-change" and when it is merely a "belief change".

Let us use Pat's example:

> *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.)*

Question: Is flicking the switch and seeing the light come on an example of a changing world ?

- From a physical viewpoint, the world has obeyed Schrödinger's equation before and after the action, so it has not changed at all.
- From a narrow psychological viewpoint, one may argue that it did change: there was no light before, there is light now, so my beliefs about what the world is like did change, therefore the world changed.

But we are not dealing here with physics or psychology. We are dealing with various formal systems for modeling the world and our beliefs about the world. For every such system, if we are to accomplish anything more useful than trying to solve Schrödinger's equation for all the particles in the universe, we must define some limits, or boundaries, thus distinguishing the things our system will explain from those that it wont. Once the boundaries are defined, they also define a set of relationships as **invariant** namely holding true permanently for all practical purposes unless they are violated by exceptions that come from **outside** those boundaries. We say that the world "changed" if any of these invariant relationships is violated.

These concepts were developed for systems with static boundaries. Humans on the other hand, show remarkable flexibility in expanding and shrinking these boundaries as the needs arise, and most of Pat's examples draw on this flexibility. This may account for his statement:

> *I dont think this sharp distinction .... is either necessary or even ultimately coherent, in fact. We are constantly....*

Yes. **We** are constantly expanding (shrinking) our model of the world as we bring in more (less) background knowlege into the working space of our inference engine. But most formal systems today are not **we**, they do work with fixed and finite number of invariants, and the laws of astrophysics are **not** among them.

Flicking the switch would be a case of a changing world if our formal system is ignorant of the causal connection between the switch and the light. But for a formal system that includes causal rules saying

```
"switch flicked causes light on"
```

and (this is important)

```
"switch unflicked causes light off"
```

seeing "light on" (even without flicking the switch) need not be considered a change of world, because "light on" can be explained within the system (in terms of the switch being flicked by some agent) and there is no need to invoke such dramatic phrases as "changing worlds" for things that we know how to handle by standard inference methods. (e.g., adding a proposition "light on" to the system and let a classical theorem prover draw the consequences. [and if our current beliefs contain "light off" then the contradiction can be handled by either temporal precedence or minimal-change belief-revision]).

What cannot be explained within such a system is the flicking of the switch (unless we model the agents' motivations too). Thus, we ask: Should the flicking of the switch be considered a "change of world" in such system? Yes, it **can** be considered such. But if our model is confined to the objects and the lighting in the room, and does not contain knowledge about switch-flicking agents, then it wont make any difference if we consider it on equal footing with the observation: "the switch is reported to be flicked". All inferences triggered by a changing world semantics will also be triggered by an observational semantic, according to which the world was static and we just opened our eyes and learned a new fact: "switched is flicked".

So when **must** we resort to the changing-world semantics? Pat answered this question:

> *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.)*

Here we have it! a violation of one of the system's invariant – the causal rules (stated above), which specify the influence of the switch on the light. Why it is special? Consider now the action "take the bulb out" with its immediate consequence "light out". Had the causal rule (above) been qualified with "flick causes light **unless** someone takes the bulb out" there would be no violation, and again we can absorb the new information as before, without invoking world-change semantics. But without such qualification, a violation occurs in a rule that was meant to be **permanent** (Vladimir even uses the predicate **always** to distinguish causal rules from less permanent relationships) Well, now we have no choice BUT to invoke changing world semantics. The new semantics is similar to the one for changing-observations, with one important difference: the suppression of certain explanations. Before the action "take the bulb out" was enacted we had the license (excuse the religious overtones, Pat) to use our causal rule and infer and explanation: "light out, therefore switch must be unflicked" But after that action, the license is revoked, and this explanation should not be infered. (even though the causal rule still resides somewhere in the database, ready to be used again in a new situation.)

And this is where the difference between actions and observations come in. An action must be equipped with a pointer to mark some causal rule "violated", to suppress its use in abduction (or explanation). Such marking is needed **only** when some of the predicates in the action's add-list are consequences by some causal rule. Predicates that are not consequences of any causal rules are called "exogenous", and it does not matter if we treat them as actions or observations – we will never attempt to explain them anyway.

Looking back, it appears that I did manage to bore you after all, despite my determination to do it to Pat only.

Time to stop.

Judea

From: Erik Sandewall

The topic of "what's in an action?" is important and quite worth some attention. In the previous Newsletter issues, David Poole wrote and Pat Hayes answered as follows:

> 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??)*

However, earlier in the same contribution Pat had written:

> *... 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.)*

Pat, I can't make sense out of your position: at one point you seem to argue that low-level and high-level descriptions of actions can't ever be separated; at another point you seem to say that they are best treated in complete separation.

My own preference is to take both into account, but to be precise about having two distinct levels of descriptions with distinct roles. In particular, this allows for dealing *both* with an action as a prescription for motor controls *and* as the expectation for what state changes will be obtained as a result. It also allows one to relate those two levels to each other.

Regardless of level, I understand an action as an instantaneous *invocation* that starts a *process*, just like in the ontology of Reiter-sitcalc. The invocation is treated similarly across levels. The two levels of description apply for the process; I've called them the *material* level and the *deliberative* level (previously in [mb-Sandewall-94], p. 3, but the deliberative level was called "image level"). An illustrative example of an action description on the material level is as follows for the action of going from the current location A to a new location B: turn around until you're looking towards B, then accelerate by 0.3 m/s/s until you reach the velocity of 1 m/s, from then on keep a constant velocity until you are at B. A lot of details aside, such as defining what it means to be "at B", the point I want to illustrate is that we use a continuous level description of the action, or more precisely, at least piecewise continuous for all the relevant state variables, *but with a possibility to change the control mode due to interventions by some higher control level.*

From a software engineering point of view, this is more like a specification than like a program. It is also "above" the level of control engineering in the sense that it provides no details about the feedback control loop between velocity sensor and motor controls. At the same time, it is much closer to the control level than a conventional description in some of our favorite logics in KR.

The deliberative level is the one that Pat alludes to, where actions are characterized by discrete properties at a small number of timepoints: possibly only the beginning and the end of the action, possibly a few more, possibly a sequence of partial world states at integer timepoints (as in basic Features and fluents). From the point of view of the deliberative level, it may be convenient to think of the material level as a "program" for performing the action, but then in a very general sense of program.

Here are some of the benefits from the explicit treatment of the action's process on both the material level and the deliberative level:

- Certain distinctions on the deliberative level can best be understood by appealing to the material level and its underlying software base, for example the *inapplicability* and the *failure* of actions. Inapplicability is when the action's program is not even able to start, for example if the robot is asked to turn on its headlights and it doesnot have any. The failure of an action is if the program executes for a while, but the expected new state of the world is not obtained. Then in turn, there are some cases where failure can be predicted in advance, and some other cases where failure comes up unexpectedly. In either case, the postcondition specifies the outcome when the action succeeds; action failure has to be described or analyzed separately.

  The precondition in an deliberative-level action description should at least guarantee that the action is applicable. It is desirable that it also excludes those cases where the action can be predicted to fail.

- The concept of action failure may be grounded on the material description level, but it is crucial for the deliberative system level as well. In particular, a simple approach to planning is to assume that for all actions we know their preconditions (that is, when they are applicable and have a good chance of not failing) and the result state when they do succeed. The plan is made for the case that actions succeed; if they fail then we have to replan. (Ray and Mikhail, I understand this is the way you do it as well?)

  A recent article [f-cis.linep.se-97-019] showed how the goal-directed process of pursuing a goal even though some of the planned actions fail, can be characterized in logic. In this way, the same concepts of failure and applicability are at work across several system levels.

- As an approach to the *validation* of the deliberative-level description of the action, including both its outcome when it succeeds and for various cases where it fails. Ideally, we'd like to *prove* that the deliberative-level description of the action follows from its material-level description. A method for doing this is described in [j-aicom-9-214].

- In order for the cogrobot to *visualize* in concrete terms what it's like to execute an action, it may instantiate its present description of the world to a particular case on the material level, and then do a simulation of the material-level description of the action on that instance.

Furthermore, David Poole wrote and Pat Hayes answered:

> 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. ... 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 doesn't 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.*

I'll agree with David if the last sentence is changed to go "...it is only sensible to attempt to carry out the action when the preconditions *are believed to* hold". This belief may be based on correct observations, on defaults, or even on incorrect observations, but if one does have bona fide belief that the preconditions hold, with a high degree of assurance (and one has no particular reason to believe that a great disaster is to ensue if one happens to be mistaken, etc. etc), then of course it is *sensible* to attempt the action.

Therefore, Pat, I don't understand what you are objecting against in this respect. (And what does it have to do with the original need for non-monotonic logic?)

Erik

References:

[f-cis.linep.se-97-019] Erik Sandewall. *Logic-Based Modelling of Goal-Directed Behavior.* Linköping University Electronic Press, 1997, Nr. 19.

[j-aicom-9-214] Erik Sandewall. *Towards the validation of high-level action descriptions from their low-level definitions.* AI Communications, vol. 9 (1996), pp. 214-224.

[mb-Sandewall-94] Erik Sandewall. *Features and Fluents. The Representation of Knowledge about Dynamical Systems.* Oxford University Press, 1994.

This newsletter is issued whenever there is new news, and is sent by automatic 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/`