Electronic News Journal on Reasoning about Actions and Change
Vol. 2: *26-47*, 31.1.1998
http://www.ida.liu.se/ext/etai/actions/nj/9801-1/

Protocol of the discussion during January, 1998 on

Ontologies of Actions and Change

Edited by: Erik Sandewall
Linköping University, Sweden

From: Hector Geffner on 21.1.1998

Pat says:

> *...Heres another illustration. I've never been very impressed by the famous Yale shooting problem, simply because it doesn't seem to me to be a problem ....*

I'm not sure I understand Pat's point well, but I think I understand the YSP. Here is the way I see it.

In system/control theory there is a principle normally called the "causality principle" that basically says that "actions cannot affect the past". If a model of a dynamic system does not comply with this principle, it's considered "faulty".

In any AI the same principle makes perfect sense when actions are *exogenous*; such actions, I think, we can agree, should never affect your beliefs about the past (indeed, as long as you cannot predict exogenous actions from your past beliefs, you shouldn't change your past beliefs when such actions occur).

What Hanks and McDermott show is that *certain models of action in AI (like simple minimization of abnormality) violate the causality principle.* In particular they show that

your beliefs at time 2, say, after LOAD AND WAIT (where you believe the gun is loaded)

are different from your beliefs at time 2, after LOAD, WAIT and SHOOT.

Namely, SHOOT at t=3 had an effect on your past beliefs (LOADED at t=2).

Most recent models of action comply with the causality principle. In some it comes for free (e.g., language $\mathcal{A}$ due to the semantic structures used (transition functions); in others (Reiter, Sandewall, etc), I'm sure it can be proved.

Regards.

- Hector Geffner

From: Murray Shanahan on 22.1.1998

Pat Hayes wrote,

> *I've never been very impressed by the famous Yale shooting problem, simply because it doesn't seem to me to be a problem.*

First I think we should distinguish between the Yale shooting scenario and the Hanks-McDermott problem. The Yale shooting is scenario is the one in which someone loads, waits and shoots, and in which those actions have certain prescribed effects. The Hanks-McDermott problem is a difficulty that arises when we take certain approaches to solving the frame problem, and is exemplified in the Yale shooting scenario. The frame problem is the problem of describing the effects of actions in logic without recourse to an excess of axioms describing their non-effects.

If you want to solve the frame problem, your solution had better be able to deal with the Yale shooting scenario. A number of early attempts at the frame problem couldn't, which is why the scenario is of interest.

Isn't that all pretty straightforward?

Murray

From: Pat Hayes on 22.1.1998

Hector Geffner:

> *... I think I understand the YSP. Here is the way I see it.*
>
> *In system/control theory there is a principle normally called the "causality principle" that basically says that "actions cannot affect the past". If a model of a dynamic system does not comply with this principle, it's considered "faulty".*
>
> *In any AI the same principle makes perfect sense when actions are* exogenous*; such actions, I think, we can agree, should never affect your beliefs about the past (indeed, as long as you cannot predict exogenous actions from your past beliefs, you shouldn't change your past beliefs when such actions occur).*

But actions - or external events - do change ones beliefs about the past. They do not change the past itself, of course: that is the causality principle. But 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. We constantly make inferences about the past on the basis of present knowledge, even in such technical areas as military planning.

I think there is a confusion here between (1) drawing conclusions about an external world and (2) simulating an external world by inferring a state from its previous state. The causality principle applies to the latter, but not the former; and even then, it is reasonable only when accompanied by a presumption of a certain kind of completeness in ones knowledge of the state. We often make predictions of the future by a kind of mental 'simulation' by inferring what is going to happen next from what is true now (as in the conventional situation calculus axiomatic approach); but in practice, such simulations are often unreliable precisely because we don't have sufficiently complete knowledge; and when this is so, we cannot cleave to the strict causality principle, but are obliged to use techniques such as nonmonotonic reasoning which allow us to recover gracefully from observed facts which contradict our predictions, which would otherwise enmesh us in contradictory beliefs. Nonmonotonicity is a good example of the need to revise ones beliefs about the past in the light of unexpected outcomes in the present, in fact, which gets us back to the YSP:

> *What Hanks and McDermott show is that* certain models of action in AI (like simple minimization of abnormality) violate the causality principle. *In particular they show that*
>
> *your beliefs at time 2, say, after LOAD AND WAIT (where you believe the gun is loaded)*

But why should you believe the gun is loaded at this time? Why is this considered so obvious? Remember, all the axioms say about WAIT is ...well, nothing at all. That's the point of the example: if you say nothing about an action, the logic is supposed to assume that nothing much happened. But if what we are talking about is a *description* of the world, saying nothing doesn't assert blankness: it just fails to give any information. If one has no substantial information about this action, the right conclusion should be that anything could happen. Maybe WAIT is one of those actions that routinely unloads guns, for all I know about it from an axiomatic description that fails to say anything about it. So the 'problem' interpretation about which all the fuss is made seems to me to be a perfectly reasonable one. If I see a gun loaded, then taken behind a curtain for a while, and then the trigger pulled and nothing happened, I would conclude that the gun had been unloaded behind the curtain. So would you, I suspect. If I am told that a gun is loaded, then something unspecified happens to it, I would be suspicious that maybe the 'something' had interfered with the gun; at the very least, that seems to be a possibility one should consider. This is a more accurate intuitive rendering of the YSS axioms than talking about 'waiting'.

We all know that waiting definitely does not alter loadedness, as a matter of fact: but this isn't dependent on some kind of universal background default 'normality' assumption, but follows from what we know about what 'waiting' means. It is about as secure a piece of positive commonsense knowledge as one could wish to find. Just imagine it: there's the gun, sitting on the table, in full view, and you can *see* that *nothing* happens to it. Of course it's still loaded. How could the bullet have gotten out all by itself? But this follows from knowledge that we have about the way things work - that solid objects can't just evaporate or pass through solid boundaries, that things don't move or change their physical constitution unless acted on somehow, that guns are made of metal, and so on. And the firmness of our intuition about the gun still being loaded depends on that knowledge. (To see why, imagine the gun is a cup and the loading is filling it with solid carbon dioxide, or that the gun is made of paper and the bullet is made of ice, and ask what the effects would be of 'waiting'.) So if we want to appeal to those intuitions, we ought to be prepared to try to represent that knowledge and use it in our reasoners, instead of looking for simplistic 'principles' of minimising changes or temporal extension, etc., which will magically solve our problems for us without needing to get down to the actual facts of the matter. (Part of my frustration with the sitcalc is that it seems to provide no way to express or use such knowledge.)

I know how the usual story goes, as Murray Shanahan deftly outlines it. Theres a proposed solution to the frame problem - minimising abnormality -

which has the nice sideeffect that when you say nothing about an action, the default conclusion is that nothing happened. The Yale- shooting- scenario- Hanks- McDermott problem is that this gives this 'unintuitive' consequence, when we insert gratuitous 'waitings', that these blank actions might be the abnormal ones. My point is that this is not a problem: this is exactly what one would expect such a logic to say, given the semantic insights which motivated it in the first place; and moreover, it is a perfectly reasonable conclusion, one which a human thinker might also come up with, given that amount of information.

Murray says:

> *If you want to solve the frame problem, your solution had better be able to deal with the Yale shooting scenario.*

This is crucially ambiguous. The conclusion I drew from this example when it first appeared was that it showed very vividly that this style of axiomatisation simply couldnt be made to work properly. So if "the Yale shooting scenario" refers to some typical set of axioms, I disagree. If it refers to something involving guns, bullets and time, then I agree, but think that a lot more needs to be said about solidity, containment, velocity, impact, etc., before one can even begin to ask whether a formalisation is adequate to describing this business of slow murder at Yale. Certainly your solution had better be able to describe what it means to just wait, doing nothing, for a while, and maybe (at least here in the USA) it had better be capable of describing guns and the effects of loading and shooting them. But thats not the same as saying that it has to be able to deal with the way this is conventionally axiomatised in the situation calculus.

Imagine a gun which requires a wick to be freshly soaked in acetone, so that just waiting too long can cause it to become unready to shoot. This satisfies the usual YSS axioms perfectly: when loaded, it is (normally) ready to fire, when fired, it (normally) kills, etc.. But if you wait a while, this gun (normally) unloads itself. Now, what is missing from the usual axiomatisation which would rule out such a gun? Notice, one doesnt want to make such a device *logically* impossible, since it obviously could be constructed, and indeed some mechanisms are time-critical in this way (hand grenades, for example). So one wants to be able to write an axiom which would say that the gun in question isn't time-critical: it has what one might call non-evaporative loading. Maybe its something to do with the fact that the bullets are securely located inside the gun, and that they don't change their state until fired...or whatever. My point is only that there is no way to avoid getting into this level of detail; formalisations which try to get intuitive results with very sketchy information cannot hope to succeed except in domains which are severely restricted.

(Response to Erik Sandewall in later message.)

Pat Hayes

From: Hector Geffner on 23.1.1998

Some brief comments about Pat's last comments.

> *In any AI the same (causality) principle makes perfect sense when actions are* exogenous*; such actions, I think, we can agree, should never affect your beliefs about the past ..*

> *But actions - or external events - do change ones beliefs about the past. They do not change the past itself, of course: that is the causality principle. But 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.*

It's important to distinguish *observations* from *actions*. 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 $j \leq i$ (the past) Basically the effect of an observation is to prune state trajectories (this is explicit in Sandewall's filtered entailment, in $\mathcal{A}$, etc)

On the other hand, what I'm saying (and of course, many other people) is that *exogenous actions* at time i, unlike observations, should *not* have an effect on your beliefs at times $j \leq i$.

You may say that if you drop the cup to the floor and it breaks, then you should change your beliefs about the past, infering things like that the cup *was* made of glass, etc.

Yet, it's not the *action* that is having an influence on your beliefs about the past; it is the *observation* that it breaks (you may say that "breaks" is not an observation but an action; yet in that case, it's definitely *not* an *exogenous* action as it depends on variables in the model)

If you remove that observation, and simply drop the cup, you will learn nothing about the past.

BTW, the causality principle is *not* about physics; I believe it's about *models of physics*. Whether such (dynamic) models are specified by means of mathematical equations or rules in a logical language, I don't think it's relevant (for compliance with the principle).

A final point before this gets to long. Pat says

> *.... such simulations are often unreliable precisely because we don't have sufficiently complete knowledge; and when this is so, we cannot cleave to the strict causality principle, but are obliged to use techniques such as nonmonotonic reasoning which allow us to recover gracefully from observed facts which contradict our predictions, which would otherwise enmesh us in contradictory beliefs. Nonmonotonicity is a good example of the need to revise ones beliefs about the past in the light of unexpected outcomes in the present, in fact, which gets us back to the YSP:*

Actually, I think *none* of the received models of actions in AI (say $\mathcal{A}$, Toronto Sit Calc, Basic Features and Fluents, etc) does that.

I believe they are all *monotonic* in the set of observations. In other words, if they predict F at time i, nothing that they observe is going to affect that prediction. At most they can make the theories inconsistent. (one of the few exceptions that I'm aware of, is a proposal in a paper of mine [c-aaai-94-245] in which state trajectories are ordered by a plausibility measure)

They are *non-monotonic* however in the set of *actions*. That is, you can affect the prediction "F at time i" by performing an action *before* i.

It's basically like in the standard dynamic models either deterministic or probabilistic.

So I think our models are not so different from more standard models of action. Of course, they are very *different* in the description languages; but that's the type of difference that you have between Strips, and transition-functions. The first is much more convenient, but it's basically a "front-end". It's not a *new* model of action; it's a new language for describing the most basic one (deterministic models). Of course, I'm completely convinced that this is very important, and I think it's precisely there where KR/action fits in.

- Hector Geffner

References:

[c-aaai-94-245] H. Geffner. *Causal Default Reasoning: Principles and Algorithms.* Proc. AAAI National Conference on Artificial Intelligence, 1994, pp. 245-250.

From: Judea Pearl on 23.1.1998

Correcting a statement by Pat Hayes.

Hector Geffner said, about the principle of causality:

> In any AI the same principle makes perfect sense when actions are *exogenous*; such actions, I think, we can agree, should never affect your beliefs about the past (indeed, as long as you cannot predict exogenous actions from your past beliefs, you shouldn't change your past beliefs when such actions occur).

To which Pat Hayes replied:

> *But actions - or external events - do change ones beliefs about the past. They do not change the past itself, of course: that is the causality principle. But 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 someone else has been present there recently. We constantly make inferences about the past on the basis of present knowledge, even in such technical areas as military planning.*

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. *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.) 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.

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?.)

Put another way, the peculiarities of actions vis a vis observations stem from the boundaries we put around our models and our audacity to call our models "worlds".

B. Russel (1910) noted (my words, not his) that if we model the entire universe, there is no meaning to causality or to actions, because the manipulator and the manipulated lose their distinction.

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. This suppression is easily enforced in systems such as logic programming (Lifschitz et al) in which inference has preferred directionality, and where abduction is not a built-in feature (e.g., contraposition is not sanctioned) but must be crafted as a special feature. However, the suppression is problematic in standard logic and probability where inference is symmetric, and no differentiation is made between abduction and prediction. Such differentiation is one role played by causally ordered relations.

Well, enough said for one tiny correction.

From: David Poole on 23.1.1998

Pat Hayes wrote:

> *But actions - or external events - do change ones beliefs about the past. They do not change the past itself, of course: that is the causality principle. But 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 someone else has been present there recently. We constantly make inferences about the past on the basis of present knowledge, even in such technical areas as military planning.*

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. The action of coming into the room didn't change your beliefs about the past (unless you include the observation that the action succeeded).

To make everything clearer it is crucial to distinguish sensing (making observations) from motor control (doing/attempting actions). The only role of sensing is to update your beliefs. Sensing in this sense doesn't change the world (of course, acting in order that you can sense can, and often does, change the world). Acting (i.e., sending a command to your motor controllers) doesn't change your beliefs about your present or your past, but only about the future. Acting can change the world.

If we fail to distinguish these, confusion will reign supreme. Surely in order to make some progress we could agree on such a distinction, then get back to the argument with at least one confusion cleared up!

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

An alternative explanation of 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.

From: Erik Sandewall on 23.1.1998

Do all current approaches behave monotonically with respect to observations? On 23.1, Hector Geffner wrote:

> *Actually, I think* none *of the received models of actions in AI (say A, Toronto Sit Calc, Basic Features and Fluents, etc) does that.*
>
> *I believe they are all* monotonic *in the set of observations. In other words, if they predict F at time i, nothing that they observe is going to affect that prediction. At most they can make the theories inconsistent.*

With respect to features and fluents, this is true for those cases where proven assessments exist at present, but not in general. The full catalogue of ontological characteristics includes such phenomena as "surprises" and "normality", both of which are nonmonotonic with respect to observations. A simple example is for the stolen car scenario: the car is left in the parking lot, three nights pass, it is for sure that the car is not removed during the day, the default is for it to stay where it is at night as well. The default conclusion is that it's still there after the three nights. Then add the observation that it's gone at the end of the period. Without admitting surprises, this scenario is inconsistent, as Hector writes. If surprises are allowed for, then you may conclude that it was removed during one of those three nights.

If you add in addition the fact that the parking lot was supervised the first and the third night, making removal impossible, then one is entitled to infer that it was stolen during the second night. Thus, the initial default conclusion regarding the presence of the car in the parking lot during the day after the second night goes from "yes" via "don't know" to "no" as these successive observations are added.

From the point of view of diagnostic reasoning these are familiar problems, 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. Except, without knowing the details, I would imagine that the people who do probabilistic or possibilistic approaches might have something to say about this. Judea, or the folks in Toulouse, any input?

Do all current approaches comply with the causality principle? On 21.1, Hector Geffner wrote:

> *Most recent models of action comply with the causality principle. In some it comes for free (e.g., language $\mathcal{A}$ due to the semantic structures used (transition functions); in others (Reiter, Sandewall, etc), I'm sure it can be proved.*

Yes, with respect to the features and fluents approach, it also "comes for free". The underlying semantics that's used there is essentially a simulation of the world, using non-deterministic transition from state to state or from state to state sequence (the latter in order to account for actions with extended duration). In either case the simulation proceeds forward in time, so it complies with the causality principle. The same applies for the extensions of the approach to deal with concurrency, ramification, and continuous change.

Erik Sandewall

From: Pat Hayes on 25.1.1998

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??)

From: Michael Thielscher on 26.1.1998

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 on 26.1.1998

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 26.1.1998

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 on 26.1.1998

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.

From: Michael Thielscher on 28.1.1998

Dear Erik,

On 26.1., you wrote:

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.

The approach is not restricted to projection, so observations may very well give cause to revising one's belief about the qualifications of an action in the past. As for the stolen car scenario: The only abnormality I consider in the paper is that of being unable to perform an action, in which case none of its effects materializes. Your Stolen Car scenario requires to consider abnormalities as to the surprising production of a single effect (or the failure of producing an expected effect). However, I can give you a straightforward formalization of your example in the Fluent Calculus, including default rules, along the lines of my KR-paper. The resulting axiomatization supports precisely the intended conclusions which you mentioned. My approach also works with non-deterministic action, so if an action has the (nondeterministic) effect that the tail pipe of either of two cars A and B gets clogged, then two preferred models result, one of which denies that we can start car B–as intended.

Michael

From: Michael Gelfond on 29.1.1998

I would like to better understand the following comment by Hector Geffner:

> *I believe they (models of actions in AI) are all* monotonic *in the set of observations. In other words, if they predict F at time i, nothing that they observe is going to affect that prediction.*

If I understood Hector correctly, the following may be a counter example. Consider the following domain description D0 in the language $\mathcal{L}$ from [j-jlp-31-201].

The language of D0 contains names for two actions, A and B, and two fluents, F and P. D0 consists of two causal laws and two statements describing the initial situation $S0$:

$$A \text{ causes } F \text{ if } P.$$
$$B \text{ causes } neg(P).$$
$$true\_at\ (P, S0).$$
$$true\_at\ (neg(F), S0).$$

The first statement says that F will be true after execution of A in any situation in which P is true. The third one means that P is true in the initial situation $S0$. $neg(P)$ stands for negation of P.

(Domain descriptions in $\mathcal{L}$ allow two other types of statements: $occurs(A, S)$ - action A occurred at situation S, and $S1 < S2$. We use them later)

Here we are interested in queries of the type

$$holds(F, [A1, ...An])$$

which can be read as "If sequence $A1...An$ were executed starting in the current situation then fluent F would be true afterwards". This seems to correspond to Hector's prediction of F at time i. We can also ask about occurrences of actions, truth of fluents in actual situations, etc).

The entailment relation on $\mathcal{L}$ between domain descriptions and queries formalizes the following informal assumptions:

- changes in the values of fluents can only be caused by execution of actions;

- there are no actions except those from the language of the domain description;
- there are no effects of actions except those specified by the causal laws of the domain;
- reasoner assumes that if there is no reason to believe that action A occured in situation S then it did not (In other words he normally observes all occurrences of actions).

As expected, we have that

$$D0 \text{ entails } holds(F, [A])$$

Now assume that the agent observed (or performed) B. This will be recorded in his description of the domain. New domain description D1 is D0 plus the statement

$$occurs(B, S0).$$

Now we have that D1 entails $neg(holds(F, [A]))$. It seems to me that the observation changed the prediction.

The second example shows how observations can change beliefs about the past. Consider a domain description D3

$$A \text{ causes } neg(F).$$
$$F \text{ at } S0.$$

This description entails $neg(occurs(A, S0))$. Now the reasoner observed that in some situation $S1$, F is false. This is recorded by adding to D3

$$S0 < S1$$
$$neg(F) \text{ at } S1.$$

The new description entails $occurs(A, S0)$. Again, observations changed the belief (this time about the past).

Hector, is this really a counter example or you meant something else?

Reference.

C. Baral, M. Gelfond, A. Provetti, "Representing Actions: Laws, Observations and Hypotheses", Journal of Logic Programming, vol. 31, Num. 1,2 and 3, pp. 201-245, 1997.

References:

[j-jlp-31-201] Chitta Baral, Michael Gelfond, and Alessandro Provetti. *Representing Action: Laws, Obervations and Hypotheses.* Journal of Logic Programming, vol. 31 (1997), pp. 201-244.

From: Luís Moniz Pereira on 29.1.1998

Dear Erik,

I noticed in the discussion that you said:

> *From the point of view of diagnostic reasoning these are familiar problems, 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.*

I have tackled the issue of nonmonotonicty with respect to observations. Cf my home page, the AAAI-96, ECAI-96, AIMSA-96, LPKR97, JANCL97, AI&MATH98 papers. Using a LP approach I perform abuction to explain observations. The abductive explanations may be: non-inertiality of some fluent with respect to some action; occurrence of some erstwhile unsuspected foreign concurrent action along with some action of mine; or opting for a definite initial state of the world up till then given only by a disjunction of possibilities.

You're right, the techniques I and my co-author, Renwei Li, use were first developed by me and others in the context of diagnosis using LP! In fact we haven't yet used them all up yet in actions. For a view of LP and diagnosis, as well as representing actions in LP, see our book [mb-Alferes-96].

Best, Luís

References:

[mb-Alferes-96] José Júlio Alferes and Luís Moniz Pereira. *Reasoning with Logic Programs.* Springer Verlag, 1996.

From: Hector Geffner on 30.1.1998

This is in reply to Michael's message.

As I see it, in the "basic" approaches to the frame problem (Reiter's completion, Sandewall form of chrnological minimization, language $\mathcal{L}$, etc), in one way or another, action rules are compiled into transition functions of the form $f(a, s)$ - where s is a state and a is an action - that describe the set of possible state trajectories. Observations in such models just prune some of the trajectories and hence have a "monotonic" effect (i.e., predictions are not retracted, at most they are made inconsistent).

In other models, Michael's and Luís's included, actions are represented in the state (in one way or the other) and hence abdution to both fluents and actions are supported. Such models *are* non-monotonic in the set of observations. Actually the only apparent difference in such models between actions and fluents is that the former are assumed to be "false by default" while the latter are assumed "to persist by default".

Interestingly, there seems to be another difference between actions and fluents in those models, and that is, that actions, unlike fluents, are not allowed to appear in the head of rules. Actually I'm not sure whether this is the case in all such models, but it is true at least in my AAAI-96 paper. The reason I excluded actions from rule heads was precisely "to prevent actions from having an effect on beliefs about the past" (just think of a rule like **hungry then eat**). Again "the causality principle".

I wonder now what's the reason the same constraint is in force in most other models of such type (if that's so).

Any ideas?

Hector Geffner