# ELECTRONIC NEWSLETTER ON REASONING ABOUT ACTIONS AND CHANGE

Issue 98004 Editor: Erik Sandewall 20.1.1998




## Today

The basic ENRAC panel discussions are intended to be standing ones rather than temporary, so a lull of one or two months does not mean that the discussion is over. Their topics are not likely to be completely resolved very soon. In the present newsletter, Pat Hayes returns to the topic of the methodology panel.

## Debates

### Research methodology

#### From: Pat Hayes

Ray Reiter wrote (21.10, position statement for panel):

> *1. Erik's notion of an ontology seems odd to me, mainly because it requires "that the "frame assumption" or assumption of persistence must be built into the ontology".*

Yes, I agree. Why choose just that assumption to build in, in any case? It clearly isn't always true (for example, if we are considering temperatures, cooking, drying paint or leaky containers, or indeed any kind of process which all by itself will produce a significant change as time goes by; or when we know that our information is imperfect; or when we have reason to suppose that there may be other agents trying to frustrate us, or even just working in the same area with their own goals which might interfere with ours; or if we know that a we are dealing with an oscillating or unstable system, or one that requires constant effort to maintain its stability.) There are many other equally plausible assumptions. For example, the assumption that things have been pretty much as they are now in the recent past (back-persistence), or that nothing will make any significant difference to anything a long way away (distance-security) or on the other side of a significant barrier (the privacy assumption). All of these, and others, are equally correct and about as useful as the assumption of persistence in limiting the spread of causal contagion.

But I think I may have understood what Erik means. (Erik, can you confirm or deny?) Let me reiterate some old observations. If the world were really a very turbulent place where any action or event might have any kind of consequences, producing all kinds of changes, then there would be no 'frame problem'. While it would of course then be difficult to describe the effects and non-effects of actions, this wouldn't be surprising, and we wouldnt call it a "problem". The FP only seems to be a problem to us because we feel that we should be able to somehow take better advantage of the world's causal inertia. So, perhaps this is what Erik means by saying that the 'frame assumption' must be built-in: he wants a model-theoretic characterisation of this causal inertia, a way to exclude models which, when thought of as physical worlds, would be worlds where things happen for no reason. He wants us to somehow delineate what the 'physically normal' interpretations are.

If this is more or less right, then there seems to me to be a central question. How can we specify the relationship of the logical 'possible world' (which is just a mathematical structure of sets of subsets of ordered pairs, etc.) to the physically possible worlds about which we have intuitions to guide us? This difficulty is illustrated by the recent discussions here. For example, my bitchin' over the distinction between gof-sitcalc and R-sitcalc comes from such a difference. Both of these use a similar notation, considered as first-order axioms: they both have things which have the role of state-to-state functions but which are reified into first-order objects, and a special function which takes a state and such an object and produces a new state. In gof-sitcalc, these are supposed to represent actions taken by agents. In R-sitcalc, they correspond rather to beginnings and endings of processes which are extended through time. The difference lies not in the axiomatic signature or even in the model theory itself, but rather in the intended mapping between the (set-theoretic) model theory and the actual physical world being modelled. We have intuitions about the physical worlds, but we dont have any *physical* intuitions about formal models.

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. All the usual axioms say is that something is done, something else is done and then a third thing is done, and the outcome is unusual; and the 'problem' is supposed to be that the logic allows that it might have been the second thing that was unusual instead of the third one. When the vocabulary suggests that the third thing is a shooting and the second a mere waiting this is deemed to be a mistake, and much research has been devoted to finding principles to exclude such models. But if the vocabulary is interpreted differently (for example, if the second event had been a shooting and the third one a waiting) this wouldn't be unintuitive. But the usual situation-calculus formalisations of this problem provide no way to distinguish these cases! They say virtually nothing about the actual physical actions involved; certainly not enough about these actions to enable a sensible choice to be made between one formal interpretation and the other. To say that a shooting is 'abnormal' in that it kills someone, in the same sense that, say, a grasping which fails to grasp or which accidentally knocks over something unexpectedly is 'abnormal', is just a misuse of a formal device. Theres nothing abnormal about a shooting resulting in bodily harm, and to invoke circumscription to overcome a 'normal' case of actions being harmless seems obviously too crude a mechanism to be generally useful. (For a general strategy for defusing YSP-type examples, consider a gunfighter who always blows the smoke from his barrel after a successful fight, and ask what persistence or inertial principles are going to ensure that its his bullets that kill people, and not that final flourish.)

Pat Hayes

---

*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/`