Electronic News Journal on Reasoning about Actions and Change 19
Vol. 2: *19–24*, 31.1.1998
http://www.ida.liu.se/ext/etai/actions/nj/9801-2/

Protocol of the discussion during January, 1998 on

Methodology of Research on Reasoning about Actions and Change

Edited by: Erik Sandewall
Linköping University, Sweden

From: Pat Hayes on 20.1.1998

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

From: Erik Sandewall on 21.1.1998

Pat, citing Ray Reiter's earlier contribution, you wrote:

> 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? ...

Well, Ray and you are bringing up two different issues here. Ray's objection was with respect to classification: he argued that the frame assumption (when one uses it) ought to be considered as epistemological rather ontological. (In the position statement that he referred to, I had proposed a definition of ontology and suggested that the situation calculus does not represent one, since the frame assumption is represented by separate axioms rather than being built into the underlying ontology). On the other hand, the question that you bring up is what kind or kinds of persistence we ought to prefer: temporal forward in time, temporal backwards, geometrical, etc.

Let me address your letter first. I certainly agree with the analysis in the second paragraph of your message: the world is not entirely chaotic, some of its regularities can be characterized in terms of persistence (= restrictions on change, or on discontinuities in the case of piecewise continuous change) and all those exceptions to persistence that are now well-known: ramifications, interactions due to concurrency, causality with delays, surprises, and so on.

For quite some time now, research in our field has used a direct method in trying to find a logic that is capable of dealing correctly with all these phenomena, that is, by considering a number of "typical" examples of commonsense reasoning and looking for a logic that does those examples right. My concern is that this is a misguided approach, for two reasons:

- Single examples do not provide enough insight into the essential structure of the problem. We will keep finding new counterexamples.
- Not all applications *require* a logic that is able to deal with all the complications. However, if we are going to use logics that only deal with some of the known phenomena, we ought to know what the logic does correctly and what it doesn't.

What I proposed, therefore (in particular in the book "Features and Fluents") was to subdivide this complex problem into the following loop consisting of managable parts (the "systematic methodology"):

1. Define an ontology, that is, a "model" of what the world is like. States and assignment-like state transitions is a very simple such ontology. Ramifications, concurrency, and so on are phenomena that call for more complex ontologies. Make up your mind about which of them you want to allow, and leave the others aside for the time being.
2. Define an appropriate logical language for describing phenomena in the ontology, including actions. Each combination of ontology and language defines a mapping from a set of formulae to the set of *intended models* for those formulae.
3. Define entailment methods, that is, mappings from the set of classical models to a modified set called the *selected models*. Usually, the selected models are a subset of the classical models.
4. Identify the range of applicability for each entailment method, that is, the conditions which guarantee that the selected models are exactly the intended ones.

5. Define "implementations" of entailment methods by expressing them e.g. in circumscription, or using modified tableau techniques. If the implementation is done right, then the range of applicability for the entailment method is also the range of applicability of its implementations.
6. When one has obtained sufficient understanding of points 2-5 for a given ontology, then define a richer one (allowing for additional phenomena of interest), and go back to item 2.

This agenda certainly aims to address all the intricacies that you mention in the first paragraph of your message, *but only in due time.* We can not do everything at once; if we try doing that then we'll just run around in circles.

In the Features and Fluents approach we have iterated this loop a few times, starting with strict inertia and then adding concurrency and ramification, doing assessments in each case. What about the other major current approaches? Early action languages, in particular $\mathcal{A}$, fits nicely into this paradigm, except that whereas above we use one single language and two semantics (classical models and intended models), $\mathcal{A}$ uses two different languages each with its own semantics. However, later action languages, such as $\mathcal{AR}$, do not qualify since they *define* the models of the action language (intended models, in the above terms) using a minimization rule. To me, minimization techniques belong in the entailment methods which are to be assessed according to the paradigm, but the gold standard that we assess them *against* should not use such an obscure concept as minimization.

On similar grounds, I argued that a situation-calculus approach where a frame assumption is realized by a recipe for adding more axioms to a given axiomatization does not really define an ontology. It can be measured against an ontology, of course, but it does not constitute one.

Ray's argument against that was that the frame assumption is inherently epistemological, or maybe metaphysical. Since most people would probably interpret "metaphysical" as "unreal" rather than in the technical sense used by philosophers, we couldn't really use that term. With respect to the term epistemological, I just notice that some entailment methods have been observed to have problems e.g. with postdiction: prediction works fine but postdiction doesn't. This means that when we specify the range of applicability of an entailment method, we can not restrict ourselves to ontological restrictions, such as "does this method work if the world behaves nondeterministically?"; we must also take into account those restrictions that refer to the properties of what is known and what is asked, and to their relationship. The restriction to only work for prediction is then for me an epistemological restriction.

On this background, Ray then questioned whether the frame assumption itself is ontological or epistemological in nature. I'd say that in a systematic methodology (as in items 1-6 above), the ontology that is defined in step 1 and revised in step 6 must specify the persistence properties of the world, otherwise there isn't much one can say with respect to assessments. This technical argument I think is more useful than the purely philosophical question of what it "really is".

You then address the following 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...*

Yes, exactly! There are two different ontologies at work here; my argument would be that each of them should be articulated in terms that are not only precise but also concise, and which facilitate comparison with other approaches both within and outside KR.

But your question at the beginning of this quotation is a fundamental one: how do we choose the additional ontological structures as we iterate over the systematic methodology loop, and how do we motivate our choices?

In some cases the choice is fairly obvious, at least if you have decided to base the ontology on a combination of structured states and linear metric time (integers or reals). Concurrency, chains of transitions, immediate (delay-free) dependencies, and surprise changes can then be formalized in a straight-forward manner. Also, we can and should borrow structures from neighboring fields, such as automata theory, theory of real-time systems, and Markov chain theory.

However, there are also cases where the choice is less than obvious. What about the representation of actions by an invocation event and a termination event, which is what R-sitcalc is about? What about the recent proposal by Karlsson and Gustafsson [f-cis.linep.se-97-014] to use a concept of "influences" (vaguely similar to what is used in qualitative reasoning), so that if you try to light a fire and I drip water on the firewood at the same time, then your action has a light-fire-influence and my action has an extinguish-fire-influence, where the latter dominates? (If there is only a light-fire-influence for a sufficient period of time, then a fire results). These are nontrivial choices of ontology; how can we motivate them, assess them, and put them to use?

To my mind, this ties in with what Bob Kowalski said in the panel discussion at the recent workshop on Formalization of Common Sense: these are pre-logical issues. It is not meaningful to begin writing formulae in logic at once and to ask what variant of circumscription is going to be needed. Instead, one ought to work out an application area of non-trivial size with the proposed ontology, probably also using a tentative syntax that matches the ontology, but without committing to anything else. Only then, as one knows what ontology is needed, is it meaningful to look for entailment methods and their implementations which may be appropriate for the ontology one needs.

The bottom line is: let's use the ontology, or the underlying semantics, as an intermediate step on the way from application to implemented system.

Going from application to ontology requires one kind of activity; going from ontology to implementation requires another kind. Such a decomposition has all the obvious advantages: it allows one to address simpler problems before proceeding to more difficult ones, it provides a way of characterizing and comparing results, and it facilitates reuse of earlier results.

References:

[f-cis.linep.se-97-014] Lars Karlsson and Joakim Gustafsson. *Reasoning about actions in a multi-agent environment.* Linköping University Electronic Press, 1997, Nr. 14.