ELECTRONIC NEWSLETTER ON
REASONING ABOUT ACTIONS AND CHANGE

Issue 98055 Editor: Erik Sandewall 13.7.1998

Today

Sandewall article accepted

The paper submitted by the present area editor has been refereed under the coordination of Susanne Biundo, who is area editor for the ETAI area *Planning and Scheduling*. On the recommendation of the three anonymous referees, Susanne Biundo has decided that the paper shall be accepted to the ETAI, with a possibility of changes by the author.

The comments and questions by the referees follow below, except for minor comments which are handled according to the general ETAI routines:

- Minor comments of interest for the reader who studies the article in detail. These have been put in a separate "minor details" page that is accessed from the article's discussion web page.
- Typos and grammatical mistakes that each reader can easily accomodate by context. These are forwarded directly to the author, with a suggestion to correct them in the revised version of the article.

Denecker et al. answer to questions by Shanahan

The present Newsletter issue also contains the answer by Marc Denecker, Daniele Theseider Dupre, and Kristof Van Belleghem to the questions asked by Murray Shanahan earlier this month.

New ETAI area starts for Decision and Reasoning under Uncertainty

A new ETAI area has been defined for *Decision and Reasoning under Uncertainty*, with Henri Prade and Salem Benferhat of IRIT (Toulouse, France) as the area co-editors. This area covers researches on reasoning and decision under uncertainty both on the methodological and on the applicative sides. Significant papers are invited from the whole spectrum of uncertainty in Artificial Intelligence researches. Topics of interests include, but are not restricted to:

Methods:

- Probability theory (Bayesian or not),
- Belief function theory,
- Upper and lower probabilities,
- Possibility theory,
- Fuzzy sets,
- Rough sets,
- Measures of information,
- ...

Problems:

- Approximate reasoning,
- Decision-making under uncertainty,
- Planning under uncertainty,
- Uncertainty issues in learning and data mining,
- Algorithms for uncertain reasoning,
- Formal languages for representing uncertain information,
- Belief revision and plausible reasoning under uncertainty,
- Data fusion,
- Diagnosis,
- Inference under uncertainty, expert systems,
- Cognitive modelling and uncertainty,
- Practical applications,
- ...

The following are the members of the Area Editorial Committee:

- **Fahiem Bacchus**, Univ. of Waterloo, Canada
- **Bernadette Bouchon-Meunier**, LIP6, Univ. of Paris VI, France
- **Ronen I. Brafman**, Stanford, USA
- **Roger Cooke**, Tech. Univ. Delft, The Netherlands
- **Didier Dubois**, IRIT, Toulouse, France
- **Francesc Esteva**, IIIA-CSIC, Bellaterra, Spain
- **Finn V. Jensen**, Aalborg Univ., Denmark
- **Jurg Kohlas**, Univ. of Fribourg, Switzerland
- **Rudolf Kruse**, Univ. of Magdeburg, Germany
- **Serafin Moral**, Univ. of Granada, Spain
- **Prakash P. Shenoy**, Univ. of Kansas, USA
- **Philippe Smets**, IRIDIA, Free Univ. of Brussels, Belgium
- **Marek J. Druzdzel**, Univ. of Pittsburgh, USA
- **Lech Polkowski**, Warsaw Univ. of Technology, Poland

We welcome the new ETAI area and wish them very successful activities.

ETAI Publications

Discussion about received articles

Additional contributions have been received for the discussions about the following article(s). For the full context, please refer to the on-line structure which contains one article interaction page with the entire discussion for each article.

Logic-Based Modelling of Goal-Directed Behavior

by Erik Sandewall

Comment from: **Anonymous referee 1**

Generally, the paper has the flavor of a progress report rather than a well-knit article. My understanding is that this perfectly complies with the publication policy of ETAI. However, sometimes such a status of a paper is indicated by adding the qualification "preliminary report" to the title, which the author might consider in the present case as well. The author himself admits that there is still a lot of work to be done. In particular, a number of rather complex axioms are proposed, of which it is claimed that they characterize "obvious properties" (p.5) of relation symbols which have been introduced with a certain meaning in mind. Past experience taught us that it can be dangerous to propose axiomatizations which seem most natural at first glance. I therefore strongly support the author's remark on the need for an "underlying semantics, and a validation of the present approach with respect to that semantics..." (p.16). An essential, most basic property of any useful axiomatization is consistency. Maybe the author could add a result in that direction.

p.7: The purpose of Section 2.6 is not clear. Composite fluents aren't used elsewhere in the paper, and so one wonders, for instance, why logically equivalent composite fluents cannot be treated as equal.

p.8: The operators G and D_s have been introduced as abbreviations for formulas, e.g., $G(s,a) \leftrightarrow H(s, inv(a))$. Why is it, then, that they cannot serve as part of axioms, provided the axiomatization includes the respective definitional formulas?

p.8: The predicate *Composite* is introduced but (as far as I could see) not used elsewhere in the paper.

Comment from: **Anonymous referee 2**

Overall

This is a well-motivated and timely contribution to the field, and, perhaps with some presentational improvements, should be accepted for publication in ETAI. The paper's main original contribution is an axiom-based account of control, focussing on the issue of action failure, retry and replanning. It also addresses the issue of how these axioms can be integrated with logical theories of action.

There follow a number of general questions I would like to raise, some presentational suggestions, and a list of typos.

General questions

Although I'm attracted to the author's work, I am tempted to ask why we should want an axiomatic account of goal-directed behaviour. For example, Axioms G1 to G4 really constitute an algorithm, and perhaps make better sense presented as an algorithm. From my point-of-view, the role of logic in cognitive robotics is to represent the world, not to represent the internal workings of the agent. Perhaps the author would like to address this question, in an electronic discussion and/or in the final version of the paper.

In Section 2.5, the third formula on page 7 entails that the postcondition of an action is false until the action is finished. But consider the example of a compound action *WashDishes*, which has the effect of making *CutleryClean* true and *CrockeryClean* true. Now, it's bound to be the case, assuming one item is washed at a time, that either *CutleryClean* becomes true before *CrockeryClean* or the other way around. Either way, one of the action's postconditions becomes true before the whole action is finished.

It's not clear to me what it means for the robot to "state" a formula on page 9. Does this mean it passes that formula to a control module?

Presentational suggestions

I have one serious presentational objection to the paper. I think it would be very much improved if there were some concrete examples showing what inferences can be drawn from the axioms. (By a concrete example, I mean one with meaningful fluent names.) Even a single concrete example illustrating some of the main ideas would be of enormous benefit to the reader's intuition. Otherwise, a hostile reader might think this is all just empty formalism.

Here are some more minor presentational suggestions. At the start of Section 2.6, the author introduces logical connectives for composing fluents. But these have already been used in Section 2.5. Perhaps this material could be reshuffled. Similarly, on page 10, the author declares how the variable symbols a and g will be used, although they have already appeared many times in the paper.

Comment from: **Anonymous referee 3**

The paper relates the concept of a "deliberated retry" to the author's earlier work on the logic of actions and change. It is intended for use in an applied project related to controlling an intelligent airborne vehicle. The problems discussed in the paper are important, and its ideas are interesting and original.

On the negative side, the paper does not really prove that its formalism is good for any specific purpose. It does not even make any mathematically precise claim of this kind. It would be good to include a description of goal-directed behavior in a toy domain and prove that some intuitively expected conclusions abour goal-directedness do indeed follow from the given axioms using the entailment methods proposed by the author. This would be more convincing than the "hand-waving" arguments in favor of the proposed approach given in Sec. 4. In the absence of such an example, the paper is reminiscent of the work on actions done in the early days of AI, such as the "frame default" in Reiter's 1980 paper on default logic, or the formalization of the blocks world in McCarthy's 1986 paper on circumscription. The ideas were interesting, but their authors were unable to prove anything about them. As the familiar shooting scenario demonstrated, a nonmonotonic formalization that looks plausible may turn out to be unsatisfactory after all. If the author of this paper tries to check that his theory works for one or two toy examples, he may very well discover bugs that need to be fixed.

It seems to me that Rob Miller was right when he suggested in his message to the author that $test(p)$ is the action of testing the agent's knowledge rather the real environment, and that otherwise the axioms for testing are not convincing. The notation proposed in the paper uses the same symbol p to represent the fluent itself and the agent's knowledge about it, which looks peculiar. Regarding the author's suggestion that the incompleteness of knowledge be represented by distinguishing between p and p' "where the lack of information is represented by making both of them false". I have this question: How would you represent the assertion that p is true but this fact is not known to the agent?

Re Theorem 1: There seems to be an implicit assumption here that the set of intervals in question is finite. Shouldn't it be included in the statement of the theorem?

Re Theorem 2: I am puzzled by the use of the word "conversely" here. It seems to me that both parts say more or less the same thing.

In the 3rd displayed formula on p. 7, conjunction is applied to fluents, which is only explained in the next section, and an interval is used as the first argument of H which, as far as I can see, is not defined at all.

My recommendation is that the paper be accepted for publication in the ETAI after the author makes the changes that he deems appropriate.

An Inductive Definition Approach to Ramifications

by Marc Denecker, Daniele Theseider Dupré, and Kristof Van Belleghem

Comment from: authors

To:**Murray Shanahan**
Dear Murray,

Thanks for your comments. You are right that we should have made more precise claims about what our formalisation yields in the circuit example.

Below, we try to answer your questions. There is brief answer and an extended answer.

First the brief answer.

In our circuit example, there are two stable states. In case a suitable primitive action of opening or closing switches r or s happens, then the physical system will switch from one state to the other. Our claim is that that our formalisation models these state transitions correctly. (Despite the fact that the effect rules contain positive and negative cycles, are not stratifiable and simple circumscription and completion both fail to assign the right semantics.)

In your simplified example, there is one stable state; in this state the switch r is open. The physical system will start to oscillate when r is closed in this stable state. As we mention in the paper, our approach is not intended to model such non-stable behaviour; actually no approach in which ramifications have "immediate" effects without delay can be expected to model an oscillation process. This is a case where delays in the effect propagation are relevant and should be explicitated. However, as we hoped for and predicted in the paper, our formalisation detects this non-stable behaviour by returning a 3-valued model. As a consequence, the transition function is undefined for the case that a primitive action of closing r is performed in the stable state.

Now the extended answer with a more technical discussion. First we make a detailed analysis of your example.

In your example, there is one circuit containing a switch p which if closed, activates relay $r1$; in the other circuit, there are switches r and q, and if they are both closed, relay $r2$ is active. An active $r1$ CLOSES q; an active $r2$ OPENS p. This is how the circuit looks:

```
--------||----------
|       |
- p ----------- r1 -
  |             |
-~r2----- r --- q --
|       |
--------||----------
```

The state constraints in this example are:

```
r1 <-> p
p  <-> ~r2
r2 <-> r & q
q <-> r1
```

If follows logically that :

```
{ p & r1 & q & ~r & ~r2}
```

This corresponds to the unique stable state:

```
{p, r1, q, ~r, ~r2}
```

If r gets closed in this stable state, the system is trapped in an oscillation: $r2$ is activated, p is opened, $r1$ is disactivated, q is opened, $r2$ is disactivated, which closes p, activates $r1$, closes q, activates $r2$, etc.. Simple electromechanical bells are constructed in a similar way and show a similar behaviour.

The formalisation in our language is as follows ($c(p)$ stands for $caus(p)$; $h(p)$ stands for $holds(p)$). The effect rules can be derived from the above state constraints with Thielscher's method using the influence information that fluents at the right can influence fluents at the left.

```
r1 <-> p yields:

c(r1) <- c(p), ~h(p)   (initiating p causes r1)
c(~r1) <- c(~p), h(p)    (initiating ~p causes ~r1)

r2 <-> r &  q  yields:

c(r2) <-  c(q), ~h(q), c(r), ~ h(r) (initiating q&r causes r2)
c(r2) <-  c(q), ~h(q), h(r), ~ c(~r)
c(r2) <-  h(q), ~c(~q), c(r), ~ h(r)

c(~r2) <-  c(~q), h(q) (initiating ~(q&r) causes ~r2)
c(~r2) <-  c(~r), h(r)

q <-> r1 yields:

c(q) <- c(r1), ~h(r1)  ...
c(~q) <- c(~r1), h(r1)

p <-> r2 yields:

c(p) <-  c(~r2), h(r2)
c(~p) <-  c(r2), h(r2)
```

Assume that in addition, there are primitive actions *close_r* , *open_r* :

```
c(r) <-  act(close_r)
c(~r) <- act(open_r)
```

Consider this definition in the case that a *close_r* action happens in the stable state. The stable state is given by the set of literals:

```
{h(p), ~h(r2), h(r1), h(q), ~h(r)}
```

The set of actions is represented by the set

```
{act(close_r)}
```

The well-founded model of the above rule set, given that *act* and *h* are interpreted by the above sets, is obtained as follows.

First, we replace all *h* and *act* literals by their truth value in all effect rules. Then we delete all rules with *false* in the body and delete all literals *true* in the body. This yields:

```
c(r)  <-
c(r2) <- ~c(~q), c(r)
c(~p) <- c(r2)
c(~r1)<- c(~p)
c(~q) <- c(~r1)
c(~r2)<- c(~q)
```

The well-founded model of this rule set is

```
true :     c(r)
undefined : c(r2),  c(~q) c(~r1) c(~p)
false :     all other literals (c(~r), c(p), ...)
```

There are many ways to verify this. Any way to construct the well-founded model is ok. For example, the well-founded model is known to be a model of the 3-valued completion semantics. Verify that the (3-)valued completion entails the following equivalences:

```
c(r) <-> true
c(r2) <-> ~c(~q) <-> ~c(~r1) <-> ~c(~p) <-> ~c(r2)
c(l) <-> false  for the other fluent literals l
```

As a consequence, the 3-valued completion has a unique model which must be the well-founded model and which corresponds to the above model.

Another way is to use our formalisation with prooftrees. The prooftree belows is the unique prooftree of c(r2) and contains the unique prooftrees for the atoms c(q), c(r1), c(p), c(r2), c(r)

```
c(~r2) <- c(~q) <- c(~r1) <- c(~p) <- c(r2) <- c(r)  <- true
                                          <- ~c(~q)
```

(So the node $c(r2)$ has 2 daughter nodes $c(r)$ and $\neg c(\neg q)$)

Our semantics first associates u to $c(\neg q)$. $c(r)$ has a true prooftree, all other atoms in the above list have a weak prooftree with only true and literals with truth value u in the leafs. So, the well-founded model is reached after one step already. It is a fixpoint because $\neg c(\neg q)$ is still undefined.

Note the close correspondence between this prooftree and the actual effect propagation. This correspondence between effect propagation and the constructiveness of inductive definitions was our main motivation for our work.

Our approach is not suitable to model this oscillation. To model oscillation, a formalism with evolving time and effect rules with delays seems necessary. However, in the paper we claim that our formalisation "detects" non-stable behaviour when it produces a 3-valued model for the above state transition and also when it derives contradictory causal literals $caus(f)$ and $caus(\neg f)$. The transition function is not defined in such cases.

So, the behaviour of our formalisation for this case is exactly what we hoped for and had predicted in the paper.

Finally, consider the original circuit example. The circuit in the paper is slightly more complex:

```
 --------||----------
 |      |
 - p ----- s --- r1 -
   |       |     |
 -~r2-----~r --- q --
 |      |
 --------||----------
```

The effect rules in the circuit are related to the following state constraints:

```
r1 <-> p & s  (if switches p and s are closed then relay r1 is active)
r2 <-> q & r  (if switches q and r are closed then relay r2 is active)
r  <-> ~s     (switches r and s are mechanically connected this way)
p  <-> ~r2    (p is closed iff r2 is not active)
q  <-> r1     (q is closed iff r1 is active)
```

It can be seen easily that this theory simplifies to

```
p & ~r1 & (s <-> r1 <-> q <-> ~r)
```

Hence, the system can occur in two stable states

```
{p, ~r1, s, r1, q, ~r}
```

and

```
{p, ~r1, ~s, ~r1, ~q, r}.
```

It can be verified also that any applicable primitive action of opening or closing r or s will yield a transition from one stable state to the other stable state.

To formalise this, we could add primitive actions *close_r* , *open_r* , *close_s* and *open_s* with the following effects:

```
caus(r) <- act(close_r)
caus(~r) <- act(open_r)
caus(s) <- act(close_s)
caus(~s) <- act(open_s)
```

The other effect rules in the paper can be derived from the first set of state constraints using the influence information that fluents at the right can influence fluents at the left and r can influence s.

Our formalisation yields the correct state transitions starting from the stable states and performing any of the primitive actions. We have a proof of this which we will add in a next version of the paper.

Thanks again.

Marc, Kristof, Daniele

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/