Electronic News Journal on Reasoning about Actions and Change

Editor: Erik Sandewall

Volume 2, Number 1

Contributions during January, 1998

ISSN, print: 1403 - 2031
on-line: 1403 - 204X

Linköping University Electronic Press
http://www.ep.liu.se/ej/enrac/1998/01/

Publication Trail and Update Rules

The Electronic News Journals (ENJ) are a medium for exchange of scientific information and debate. In particular, they serve as the forum where articles received by the ETAI (Electronic Transactions on Artificial Intelligence) are discussed publicly for review, in separate News Journals for each ETAI area.

The Electronic News Journal on Reasoning about Actions and Change (ENRAC), in particular, uses a publication trail where contributed information is first distributed on a daily basis as *Newsletters* using electronic mail and web pages in HTML. Then, contributions during the same month are compiled into an HTML version of the monthly *News Journal*. Finally, the same information is converted via Latex to a postscript version that is suitable for printing on paper and reading off-line. It is formally published and archived by the Linköping University Electronic Press in both its electronic and its paper form, as described on page (iv).

Offprints of individual segments. Each News Journal issue consists of a sequence of *segments* relating to specific topics. "Offprints" of these can be obtained from the URL:s indicated at the head of the first page of each segment. It is intended to maintain these URL:s and their contents for the foreseeable future. Misprints of minor significance are corrected in these offprints (but not in the E-Press version since it is considered archival).

Corresponding HTML edition with dynamic links. The HTML edition contains considerable numbers of links to other pages and structures on the net: articles that are available on-line, home pages of conferences and of individual researchers, links to other part of the ETAI structure, and so on. Due to the natural limitations of the paper medium, only some of the WWW links have been retained here as footnotes. In order to retrieve articles and other information that are referenced in the present issue, it is recommended to look up the corresponding issue in HTML and to use its link. The HTML issues of the News Journals on Reasoning about Actions and Change can be found at the following URL:

`http://www.ida.liu.se/ext/etai/actions/njl/`

It is intended that the HTML issues will by updated continously to the largest extent possible, for example by replacing URL links to the home pages of authors that have changed to another site.

Table of Contents

Dates of Publication

Since the date of publication may be understood either as the date of first public appearance, or as the day of reproduction on paper in many copies, and since both of these definitions may be difficult to apply in the case of electronic publication, we make the following clarifying statement.

The contents of the present issue were put on-line in their original, HTML version during the month of January, 1998. Then the contents were edited and formatted, resulting in the present, formatted version which was published on May 22, 1998, in two concurrent editions: an on-line edition and a paper edition. The on-line edition was timestamped electronically and put on-line by Linköping University Electronic Press at the URL specified on page (i). The paper edition was obtained by printing the on-line edition on a standard computer printer. It was reproduced in 200 copies, legally archived, and made available for distribution.

Vol. 2: *1-1*, 31.1.1998
http://www.ida.liu.se/ext/etai/actions/nj/9801/

The Month of January

Summary of News Journal Contents

NRAC Panels Still Very Active

In the on-line Newsletter communications, this month has been dominated by the topic-oriented ("panel") discussions on *ontologies of actions and change* and on *methodology of research* in our area. The monthly protocols for those debates take up a major part of the present issue.

The actual discussion in the present Colloquium does not always respect the distinction betweent the two panel topics, and attempts to classify contributions according to that distinction are sometimes preempted by the way the debate then proceeds. The present editor therefore takes the liberty to both assign and reassign debate contributions, even retroactively, in such a way that the two debates are somewhat self-contained. However, it is still the case that contributions in one debate may refer back to earlier contributions in "another" debate.

ETAI Received Research Articles

There has been continued discussion about two articles that were submitted to the ETAI in earlier months. Since questions that were posed in December were answered in January, the present issue contains both the December questions and the January answers.

New Book by Ray Reiter

Ray Reiter sent the following message:

> *I have decided to make available a draft of my book on the sitcalc to ETAI subscribers. It can be downloaded from our group's web site:*
>
> `http://www.cs.toronto.edu/~cogrobo/`
>
> *Here is a brief descriptor:*
>
> *Being a draft of the first eight chapters of a book on the situation calculus and its relevance to artificial intelligence, cognitive robotics, databases, programming languages, simulation, and control theory. Permission is granted to download copies for personal, non-commercial use. Oversights, corrections and advice will be gratefully (and gracefully) received.*

Vol. 2: 2-7, 31.1.1998
http://www.ida.liu.se/ext/etai/received/actions/002/

Protocol of on-line discussion during January, 1998 about the following research article:

Michael Thielscher

A Theory of Dynamic Diagnosis

This section contains questions that were posed in December, 1997 as well as the corresponding answers from January, 1998.

Q2. **Marie-Odile Cordier** (23.12)

I have three main comments on this paper :

1) The example on page 4-6 is quite interesting and highlights very clearly what happens when dealing with interactive faults ($ab(r1)$ causes $ab(re1)$ when $s1$ is closed). But, I did not really agree with the conclusion the author is drawing from it.

What it clearly highlighted, in my opinion, is that "minimizing abnormality" cannot be used when dealing with interactive faults. Most research papers on diagnostics suppose implicitly that faults are independent and equiprobable, and in these cases, "minimizing abnormalities" is a good way of selecting the most probable diagnoses. However, as soon as you are dealing with interactive faults the preferred diagnoses have no good reason (no probabilistic foundation) to be the minimal ones. In the example, the probability of $ab(re2)$, knowing $ab(r2)$ and $closed(s2)$ is equal to 1 whatever the probability of $re2$ is of being faulty from its own. Then, $d1$ has to preferred rather than $d2$ and $d3$.

The key point is that "minimization" (minimizing abnormalities) is not a good preference strategy in case of interactive faults. This seems to me to be the very reason why one doesn't get the expected results in this example.

Another point concerns when this "preferring" step has to be done. The author argues that it has to be done at the starting point and uses the example as a justification. I don't contest this fact (see below), but I contest that it follows from the example.

From the following example, it can be seen that "minimizing abnormalities" is not a correct solution even if it is done at the starting state. Let us suppose that it is known that $closed(s1)$, $closed(s2)$, $closed(s3)$ and $off(light)$ are true in the initial state. Whatever the action might be, for example "open s1" or the empty action, you are going to prefer a state where $ab(re1)$ is true rather than the one where $ab(r1)$ and $ab(re1)$ are true, which is not at all justified from a probabilistic point of view. It is even problematic from a diagnostic point of view, for replacing $re1$ by an unfaulty relay (instead of replacing $r1$) will lead to breaking $re1$ again as soon as you will close $s1$. In fact, example 1 (page 20) exhibits similar results which are not satisfying by forgetting $ab(r1)$ and $ab(r2)$ as possible faults.

2) The term "dynamic diagnosis" is used throughout the paper to denote diagnosis on systems on which you are performing actions (tests). As far as I understand, the systems are supposed to be static ones; they are not supposed to evolve by their own; they don't have any proper dynamic behaviour. The only way to make them change is to perform actions. This is the reason why you can predict the resulting state by looking only to the effects of the action. Unpredictable events are not taken into account, for exemple faults (or more simply evolutions of the system) that occur during the sequence of actions.

The term "dynamic diagnosis" is then misleading, at least for the diagnostic community for which dynamic diagnosis usually means diagnosing systems evolving in time by themselves, without explicit exogeneous events making them change.

In this context, the problem which is proposed is very similar to that of postdiction : knowing some observed facts resulting from an action (or a sequence of actions), you want to infer the actual state of the system. Faults cannot happen during the sequence of actions, and then dealing with an action or a sequence of actions makes no difference.

Consequently, it is quite justified to apply the "preference step" on the initial point. You have to determine the most probable sequences of steps (or histories, scenarii, trajectories?) starting from initial states, leading to some final states in which observations are true, and corresponding to a given sequence of actions. There are no unknown events; there is no uncertainty in the actions; no uncertainty wrt their effects; the only uncertainties concern the initial states. The preference between sequences depends directly on preferences on initial states, which explains why the "preference step" concerns the initial states.

This scheme is a restricted case of a most general scheme in which you take into account the possible occurences of events (as faults) interleaved with the actions, the probability of such events, and the probability of an action to produce some effects. Selecting some of these scenarii according to preference criteria corresponds to what we called "event-based diagnosis" in [Cordier 94]. It is also close to McIlraith's approach; see [McIlraith 94], [McIlraith 97]. The main difference between these approaches is that we tried to define diagnostics independently of the mechanism used for modeling actions and changes, whereas Sheila's proposal is clearly dependent on the formalism used to model actions (situation calculus).

3) The last point concerns the ramification problem and the use of a causal model to predict the effects of an action. I realize that this point is not the main subject of this paper, since it is devoted to diagnostics. However, an important point related to this paper is to examine whether it can be applied when dealing with dynamic systems.

This theory of action based on causal relationships is very attractive as long as you are looking for the effects of an action or a sequence of actions, and as long as the concerned system has no proper evolution. Fluents which are not affected by an action are then supposed not to change, by virtue of a minimal change principle. But as soon as you are concerned with dynamical systems, (which is not really the case in the paper), such a causal model would probably not be sufficient and you will need a "transition model", describing the way things evolve along time. This happens for example if you want to model the dynamics of a system, or the possible events as faults that occur as you are monitoring a system, or the natural ageing of components.

An answer could be that there cannot be changes without causes, but most of the time you don't want to model these causal chains or you are not even able to model the primary causes of such evolutions (for example, the ageing of components or the sudden occurence of a fault), but you want nevertheless to take them into account as far as possible.

The basic idea of the proposal we made in [Cordier 95] was that a transition model (that is, a set of possible (partially ordered) transitions) is needed in order to decide what is the most plausible state after an update, or equivalently, an action. A causal model is certainly quite adequate when considering static systems reacting to actions. More than that, in my opinion, a causal model is a very nice formalism allowing to acquire the partial orderings that exist between transitions, in a natural way. There is probably a strong correspondence between your "influences" and our "partially ordered transitions" which would be worth studying more deeply. However, transitions seem to have a broader scope in that they allow to represent any changes from one world to a next one, whereas causal relations or influences are restricted to represent "causative changes" (changes for which one can exhibit the causes). This is a problem when dealing with dynamical systems.

References:

[Cordier 94] Marie-Odile Cordier and Sylvie Thiébaux. *Event-based diagnosis of evolutive systems.* Proceedings of the 1994 DX Conference.

[Cordier 95] Marie-Odile Cordier and Pierre Siégel. *Prioritized transitions for updates.* Proceedings of the 1995 ECSQARU Conference, pp. 142-151.

[McIlraith 94] Sheila A. McIlraith. *Towards a theory of diagnosis, testing and repair.* Proceedings of the 1994 DX Conference, pp. 185-192.

[McIlraith 97] Sheila A. McIlraith. *Explanatory diagnosis: conjecturing actions to explain observations.* Proceedings of the 1997 DX Conference, pp. 69-77.

A2. **Michael Thielscher** (9.1)

I want to thank Marie-Odile for the very detailed comments, the pointers to related literature, and for the interesting questions. Let me answer the latter in turn.

The main matter of concern in my paper is to point out a crucial aspect of generalizing theories of diagnosis to what I call the dynamic case: Component failures, which usually are considered abnormal, may naturally arise in the course of time as direct or indirect effects of actions (or events, for that matter). As a consequence, it is inappropriate to simply apply minimization strategies that are suitable in the static case equally to all states that arise during the diagnosis process. Rather it is necessary to employ the 'static' strategy only once, namely, initially. This allows to count as perfectly normal any component failure that is effected later on. Now, in the paper I have pursued one particular 'static' minimization strategy, namely, where it is required that if there is a difference in the a priori likelihood of component failures, then it needs to be explicitly stated so. This is not the case in the scenario and reasoning process you described: The a priori likelihood of $ab(re1)$ is higher than that of $ab(r1)$. Hence, what you suggest is a more sophisticated *static* strategy, where differences in the a priori likelihood of abnormalities are derived from state constraints. This does not affect my argument: If diagnosis concerns a sequence of states, then again this more elaborated minimization strategy should be applied only once, thus allowing for component failures to be effected in the course of time.

Concerning the second remark, I admit that for the diagnosis community the term "dynamic" might be misleading, at least at first glance. To begin with, in the area of formalizing reasoning about actions and change it has become common to call "dynamic" any system that may exhibit different states in the course of time, no matter whether state changes are brought about by exogenous action or by natural events that happen inside the system. The striving motivation for me adopting this term was to contrast the contents of the paper to Reiter's work on diagnosis [j-aij-32-57]. But apart from that, many results have been established that show how approaches to reasoning about actions can be generalized to the case where actions are performed in self-evolving systems. (See, e.g., [c-ijcai-95-1956], [c-kr-96-2], [mb-Shanahan-97], to mention a few.) The gap between dynamic systems which idle unless actions are performed, and those that are self-evolving turned out less deep than one might have expected. In particular, the theory of causal relationships (and the axiomatization on the basis of the Fluent Calculus) has very recently been extended to allow for natural events aside from exogenous, volitional actions [Thielscher 98]. Combining this extension with the diagnosis approach pursued in the paper has yet to be achieved, but there seems to be no fundamental obstacle for so doing.

This leads to my answer to your third remark. An example taken from [Thielscher 98] is the following specification of the behavior of a simple timer by means of a causal relationship:

```
(clock IS x+1) causes (timer IS t-1) if (timer=t and t>0)
```

which says that whenever the global clock ticks, then the timer counts down until it reaches zero. This shows that causal relationships do not depend on the performance of actions but just on the occurrence of effects, which might just as well be the consequence of natural events inside the system. Now, the motivation for using a causal description rather than an explicit transition model is that only the former corresponds directly to a natural and "elaboration-tolerant" (see [c-fcs-98-198]) description of actions and events, effects and change. (Notice that via the theory of causal relationships, any causal description implicitly determines a transition model.) Generally, research in the area of commonsense reasoning about actions is largely motivated by the conceptual gap between a natural description of causes and effects, and the corresponding transition model. The Ramification Problem, for instance, has to be understood in this way, and so is of course an issue in case of self-evolving dynamic systems just as well (see again [Thielscher 98]).

References:

[c-fcs-98-198] John McCarthy. *Elaboration tolerance.* Proc. Formalization of Commonsense Reasoning, 1998, pp. 198-216.

[c-ijcai-95-1956] Michael Thielscher. *The Logic of Dynamic Systems.* Proc. International Joint Conference on Artificial Intelligence, 1995, pp. 1956-1962.

[c-kr-96-2] Ray Reiter. *Natural Actions, Concurrency and Continuous Time in the Situation Calculus.* Proc. International Conf on Knowledge Representation and Reasoning, 1996, pp. 2-13.

[j-aij-32-57] Ray Reiter. *A theory of diagnosis from first principles.* Artificial Intelligence, vol. 32 (1987), pp. 57-95.

[mb-Shanahan-97] Murray Shanahan. *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia.* MIT Press, 1997.

[Thielscher 98] Michael Thielscher. *How (not) to minmize events.* Submitted to the conference KR-98.

Q3. **Wolfgang Nejdl** (27.12)

One basic remark about the title and intention of your paper:

When I first read your paper, I had some difficulties in connecting your work to the usual diagnosis literature, as you basically refer only to papers about reasoning about action and change (which is ok, considering the content, but should be changed, considering the title). Also, the current title is somewhat misleading, as the term "dynamic diagnosis" in the diagnosis community is usually reserved for systems which monitor and diagnose continuous and/or time-varying systems. I would have suggested something like "diagnosis and actions" or similar within the title.

Anyway, here are some more specific questions, which came into my mind, while I was trying to comprehend your approach.

1. It seems to me, that one main problem (chapter 2) you are considering are dependent failures like "$ab(c1)$ implies $ab(c2)$", which are usually neglected in many papers. Could you elaborate more on the advantages of your formalism when these dependent failures are not present? In such a case, what exactly do you gain by including explicit causal relationships (considering that most diagnosis systems use just ordinary state constraints)?

2. A second thread which seems to emerge in chapter 4 is the integration of test actions. Have you thought about which test actions one should take, or is this only a side issue in this chapter?

3. Also, could you comment some more about the relationship of your approach to the one of Sheila McIlraith?

Best regards,
Wolfgang Nejdl

A3. **Michael Thielscher** (9.1)

Let me now give answers to the questions by Wolfgang Nejdl. One of them also concerns the use of the term "dynamic systems" and so has already been considered in the answer to the previous question. The references before refer to the bibliography of that answer.

As for the motivation for using causal relationships, this is entirely due to the dynamic aspect. The past decade or so of research in reasoning about actions has revealed that state constraints, being static in nature, may however also give rise to indirect effects of actions (or events)– yet it is usually challenging to determine exactly to which effects. In fact, state constraints often lack information needed to tell 'real' indirect effects from 'phantom' effects, which would never occur in reality. Causal relationships do provide this necessary additional information. (See the first three sections of [j-aij-89-317] for a detailed introduction to this problem and for the motivation behind causal relationships.)

As for the problem of determining suitable test actions, this is indeed a side issue in the present paper. I merely want to show that the theory does in principle support the search for such actions. A detailed account would mean to develop a sophisticated search strategy and to make precise how the relative usefulness of possible actions is to be judged. This would clearly go beyond the scope of the present paper, but is definitely of interest and could very well be the topic of a follow-up paper.

Finally, to the relation to Sheila's approach. First of all, there is a difference from a very general point of view. Her approach starts with a particular axiomatization strategy, namely, the Situation Calculus. My intention was to focus on a high-level theory of dynamic diagnosis, prior to considering the problem of a suitable first-order axiomatization. The latter I have briefly tackled in the section on a (provably correct) Fluent Calculus-based axiomatization. At least for a large part of the high-level theory, Sheila's axiomatization may very well prove correct, too. Some more specific remarks as to differences in the expressiveness have been made in the discussion part of the paper. In addition to this it is worth mentioning that much work has been done on the pros and cons of Situation Calculus in general compared to alternative axiomatization strategies (among which are the Event- and the aforementioned Fluent Calculi; see, e.g., [mb-Shanahan-97] or [j-aij-89-317]).

References:

[j-aij-89-317] Michael Thielscher. *Ramification and causality.* Artificial Intelligence, vol. 89 (1997), pp. 317-364.

[mb-Shanahan-97] Murray Shanahan. *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia.* MIT Press, 1997.

Electronic News Journal on Reasoning about Actions and Change 8
Vol. 2: *8-10*, 31.1.1998
http://www.ida.liu.se/ext/etai/received/actions/005/

Protocol of on-line discussion during January, 1998 about the following research article:

José Júlio Alferes, João Alexandre Leite, Luís Moniz Pereira, Halina Przymusinska, and Teodor Przymusinski

Dynamic Logic Programming

This section contains questions that were posed in December, 1997 as well as the corresponding answers from January, 1998.

Q1. **Erik Sandewall** (12.12)

Your paper addresses update of knowledge bases represented by logic programs - a topic which is known and understood by only a limited part of the reasoning about actions community. Because of the similarity of research goals and the difference of background, I think that a discussion between you as authors and the readers in our area is particularly important and valuable; it can hopefully facilitate very much the understanding of this specific work as well as the "school" that you represent. Towards the end of the paper, you explicitly mention "reasoning about actions" as one of the intended applications, which of course is of particular interest to us.

Let me start out with a question about the prehistory of the approach you have chosen. You refer to Marianne Winslett's 1988 article [c-aaai-88-89] as an early reference for an "interpretation update" approach. Based on an example where it does not seem to give the intended result, you propose that the principle of inertia should be applied to the rules of the initial program rather than to the individual literals in a model. However, Winslett's article was written in response to an earlier article by Ginsberg and Smith [s-Brown-87-233] where they proposed exactly this: to define update on a set of logic formulae. Winslett pointed out some examples where the approach of Ginsberg and Smith did not work as intended, which is what led her to propose interpretation update. (An even earlier reference would of course be to the work by Lewis on counterfactuals [mb-Lewis-73].

My first question, therefore, is to what extent is there a difference: does your approach avoid the problems observed by Winslett, and if it does, what is the key to this improvement?

My second question is with respect to updates in the presence of observations and action laws. One of your results is that if the initial program is just a set of facts, then program updates and model updates coincide. However, in the case of reasoning about actions, one typically deals both with facts about the world at various points in time ("observations") and with rules characterizing some of the effects of actions ("action laws", "effect laws"). If update methods are used for characterizing ramification, which is what Winslett's article was all about, then presumably one wishes to prefer changes of "facts" (that is, sign reversal of literals) over changes of the action laws, at least as a first approximation. Only in the presence of accumulated evidence is it reasonable to revise a well established action law. How would you foresee representing such cases: will action laws be written out explicitly as logic-program rules, and what updates will then be obtained on the current state? On the other hand, if action laws are not represented as rules, how are they represented and how are the results in your article to be used?

References:

[c-aaai-88-89] Marianne Winslett. *Reasoning about actions using a possible models approach.* Proc. AAAI National Conference on Artificial Intelligence, 1988, pp. 89-93.

[mb-Lewis-73] D. Lewis. *Counterfactuals.* Harvard University Press, 1973.

[s-Brown-87-233] Matthew L. Ginsberg and David E. Smith. *Reasoning About Action I: A Possible Worlds Approach.* In: Brown (ed): The Frame Problem in Artificial Intelligence, pages 233-258. Morgan Kaufmann Publishers, Inc., 1987.

A1. **Luís Moniz Pereira** (27.1)

It's probably easiest and less time consuming that I reply personally, rather than producing a reply involving all five authors, at least at this stage, where I'm not saying anything controversial, I believe.

> *My first question, therefore, is to what extent is there a difference: does your approach avoid the problems observed by Winslett, and if it does, what is the key to this improvement?*

Section 5.1 in our paper, and Theorem 7 in particular, shows that our approach generalizes that of Winslett as it applies to logic programs. Winslett updates interpretations instead of theories. We update theories or programs, but obtain interpretation updates as a special case. In the introduction of the paper we motivate why we do so. Further elaboration on the motivation can be found in our own work, references [Lei97] and [LP97], which I'll gladly provide for those interested, or can be gotten from our home pages.

Note that we're NOT performing belief revision. We're performing updates. We are not dealing with contradiction. New information overrides old information and that takes care of conflicts between the two. If new information is by itself contradictory we do not revise it. If old information is seen to be contradictory itself upon introduction of new information we don't revise it. We do no revision at all. Ginsburg and Winslett on the other hand are preoccupied with revision.

Coupling updates and revision, for when contradiction arises, is future work. It seems to me we can do it based on our previous work in LP revision, as can be found say in [mb-Alferes-96] and references therein, or my home page for more recent work on revision of logic programs, which I've done a lot of. But I should refrain to comment any more here on belief revision since it is not the subject of our paper. I do think that updates and revision should be developed separately.

> *My second question is with respect to updates in the presence of observations and action laws. One of your results is that if the initial program is just a set of facts, then program updates and model updates coincide. However, in the case of reasoning about actions, one typically deals both with facts about the world at various points in time ("observations") and with rules characterizing some of the effects of actions ("action laws", "effect laws"). ... presumably one wishes to prefer changes of "facts" (that is, sign reversal of literals) over changes of the action laws, ... Only in the presence of accumulated evidence is it reasonable to revise a well established action law.*

Again, we're not doing revision at all. If a fact updates a "law" by going against its conclusion then the fact wins! However, if you want to have persistent laws, you simply make them part of every update. Although we mention the application to reasoning about actions in the paper, we relegate it to future work. We are working on it at the moment. To do so we use updates that connect a previous state in the premises to a future state in the conclusion of the update, which type of updates we did not introduce in the paper. Since this discussion is about the paper, I will refrain from discussing that here further.

> *How would you foresee representing such cases: will action laws be written out explicitly as logic-program rules, and what updates will then be obtained on the current state? On the other hand, if action laws are not represented as rules, how are they represented and how are the results in your article to be used?*

This is ongoing work. Basically, actions are modelled by rules whose conclusions are post-conditions, but these only come into effect in the next updated state, ie the conclusions have the form of an internal self-update (not defined in the paper...), and the premises contain the pre-conditions and the action name. By performing an (external as opposed to internal) update of the action name the self-updating rule gets triggered. One of the post conditions deletes the action name if the action is not a persistent one. This will be the subject of a future paper.

I'm sorry to sound rather evasive, but the issues you address regarding actions are not tackled in the paper, but only mentioned as future work.

What the paper does is to establish a foundational base for dealing with updates of logic programs, which was sorely lacking till now. The exploration of a more dynamical type of logic programming is thus made possible. It will, we hope, lead to the use of present and forthcoming efficient LP implementational technology, and recent semantical developments, to a wider scope of problems.

References:

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

Electronic News Journal on Reasoning about Actions and Change
Vol. 2: *11-19*, 31.1.1998
http://www.ida.liu.se/ext/etai/actions/nj/9801-3/

Protocol of the discussion during January, 1998 about the following article:

Wolfgang Bibel

Let's Plan It Deductively!

Edited by: Erik Sandewall
Linköping University, Sweden

The discussion about Wolfgang Bibel's invited lecture at IJCAI 1997 started in the September, 1997 issue of the present News Journal. After a pause of several months, the discussion continued in January as follows.

From: Erik Sandewall on 25.1.1998

Dear Wolfgang,

With respect to your IJCAI article, I have a number of questions several of which relate to the research methodology or paradigm being used. I observe that when Marc Friedman ended one of his questions with "why prefer one solution to the other", your answer was

> *To the best of my knowledge Michael Thielscher (...) was the first who gave a solution to the ramification problem which overcomes deficiencies in any previous solution (...) no previous solution would model reality in a correct way. A better solution in this sense must be preferred to a deficient one. (...)*

In this way, you appeal to the traditional method of validating approaches to common-sense reasoning by ways of counterexamples: a method is accepted until disproved by an example. Unfortunately, this research methodology can not provide any reliable conclusions. As has been recognized in core computer science since a long time, the lack of knowledge of a counterexample does not prove that a proposed solution (or a program) is correct.

On this background, I have the following questions or observations:

1. What is your perspective on current research in reasoning about actions and change where we provide *assessments* or *validation results* for new theories, instead of merely proposing new logics based on a combination of intuition and toy examples? The new approach is being used by several of us, including Shoham (who started this trend), Lifschitz, and myself. In spite of being a survey paper, your article does not mention this development at all.

2. In the section on qualification, you describe a solution using transition logic: "TL opens a new way to deal with this problem". The solution is exemplified using McCarthy's old "potato in the tailpipe" problem, but not

motivated in any other way. In particular, there is no statement of when the method is correct or is not correct, nor any proof or reference to a proof of its correctness under some precisely stated assumptions.

Unfortunately, however, it is well known that simple solutions to the qualification problem will easily fail if the naive potato and tailpipe example is modified ever so slightly. For example, suppose there are two cars, A and B, it is known that a potato is put into the tailpipe of one of them, and one asks whether car B will start properly. In such a case, the absence of positive knowledge that the tailpipe of car B has been plugged, does not allow one to draw the default conclusion. Or, suppose one does not know whether the plugging of the pipe preceded or succeeded the attempt to start the car; the same difficulty arises again.

Given that previously proposed solutions have failed so easily, one would like to have some evidence that the method proposed in your article is not going to encounter those problems, or other ones which may come up. In other words, here is a case where the shortcomings of the older, example-driven methodology are evident.

My question is what can be said in general about the method you propose for qualification: when is it known to work, and when not?

3. In section 4, you write "LCM has been the first method which actually solved these aspects of the frame problem and did so in the optimally possible way". Since there is no obvious definition for optimality in this respect, I wonder which quality measure do you use, and what is the proof that the LCM method is optimal with respect to it?

4. Another example of an incompletely substantiated claim occurs in your section on ramification. You write "Lifschitz' categorization of fluents does *not* work in this example. We need to *categorize the actions* into primary and secondary ones (rather than the fluents) as done in the solution presented in this section". However, the second sentence does not follow from the first one. The fact that the particular variety of fluent categorization that was proposed by Lifschitz doesn't work for the example, does not prove that **all** fluent categorization methods fail for it, or for some reasonable class of examples. So in what sense do we "need to" categorize the actions?

5. Also with respect to your formalization of Thielscher's example with three switches and a relay, it is remarkable that the electric circuit in question can easily be understood in terms of dependencies and persistence, but the proposed formalization requires the axioms to represent the propagation of *changes*: "if this fluent changes in such-and-such a way, then that fluent also changes in such a way". This seems clumsy and counterintuitive. Do you claim that it is the best possible representation in the present state of the art?

6. As you correctly observe, a rational agent must be capable of reasoning about the timing of actions and about changes within the duration of an action. In section 6.4 of the paper you describe in outline how to introduce timing of actions into TL. However, the most obvious way of arranging this is by approaches that use explicit, metric time, as in Shoham's work in the 1980's, the Features and fluents approach, and the modern event calculus. The PMON(RCs) logic presented by Doherty and Gustavsson at KR-96 is an example of such a logic. To summarize,

- PMON(RCs) uses explicit time, which allows one to characterize the duration of actions and the intermediate changes within its duration;

- PMON(RCs) also allows one to characterize the timing of concurrent actions with overlapping duration periods;
- PMON(RCs) allows delayed effects of actions;
- The circumscription policy used in PMON(RCs) has been shown to be reducible to the first-order case, which insures that standard theorem-proving techniques and their optimizations can be used to compute entailment.

In other words, it does in concrete detail what you only describe in gross outline. My questions are:

1. What are the advantages of your approach over the one I just referred to? (This is a valid question in view of your statement "A better solution in this sense must be preferred to a deficient one").
2. In PMON(RCs) and other approaches that use explicit metric time, it is straightforward to make statements about durations, comparing durations of actions, and so on, basically because each interpretation represent an entire history of the world. How can this be done in a logic like TL where the => operator takes one from one state to the next, and updating the current time in the process?

7. You write "section 6 shows how the various aspects involved in reasoning about actions and causality can be taken into account within TL". However, nothing in section 6 or elsewhere in the article presents any concrete results about how spontaneous change in the world can be represented - it is as though the world were entirely static when no actions are taken. The resulting concept of causality is quite meagre.

Here again, there *exists* in fact a body of results in this area, ranging from the Situated Action Theory of Morgenstern and Stein and my own early work on integrating differential equations into logic (presented at KR-89 and IJCAI-89) to Murray Shanahan's earlier and recent work which uses a very similar approach, and which has also been used for implementing a simulated robot. When you write "another issue concerns the integration of differential equations and their computation within a logic such as TL", the uninformed reader will not easily guess that the integration occurred nine years ago, although not for TL.

8. In section 4, on the topic of how to obtain a specific plan as a solution to a planning problem, you write "[Bibel 1986a] introduced state literals, *S(x)*, which keep track of the states passed through while executing a plan. (...) By unification the variable denoting the goal situation will then along with a successful proof always provide a term that expresses the linearized sequence of actions." Yes, but how is this different from the use of the Answer predicate which was proposed and used by Cordell Green in the late 1960's?

9. In subsection 6.3 you refer to an "easy solution" for allowing action laws for special cases to take precedence over a more general law for the same action. However, in the example you quote, the specialized case only contains an *additional* effect besides the effect in the general case. Does the same solution also apply if some of the effects of the general case *do not* arise in the special case?

10. If the answer to the previous question is positive, does the planning method that you propose using unification against the state literal (answer literal) still work when such specificity is allowed? The question arises since unification against answer literals results in planning backwards from goals, whereas the solution for specificity is defined in terms of forward simulation of the plan.

11. A final observation: The title and the introduction of your article focusses on planning, as shown when you write: "In this paper I review the state of the art in deductive planning...", but the major part of the contents (after introducing the TL logic as such) deals with problems in reasoning about actions and change: the frame problem, ramification, and so on. May I suggest that some additional coverage of modern results in the latter area would be appropriate, in particular since this is claimed to be a survey article.

Sincerely
Erik Sandewall

From: Wolfgang Bibel on 27.1.1998

Dear Erik,

Thank you for your continuing interest in the paper underlying my invited IJCAI-97 talk. Since the main focus of your questions concerns "the research methodology or paradigm being used" I need first to clarify the nature of this paper.

As you correctly state the paper is meant to be "a survey paper"; however you failed to add (as clearly stated in the paper in the same sentence) that it is meant to be one "with an emphasis on the contributions from research groups influenced by the author's own work" (p2). As an invited talk this is an absolutely appropriate focus. For instance, your own invited talk at the ECAI conference in Budapest was of exactly the same nature. In fact you may recall that I then for fun complained about YOUR ignorance of all of OUR work (although NOT in public as you now decided to do). Reiter's Chambery IJCAI address did again the same representing the Toronto school. And so forth.

In consequence of this purposely chosen and clearly stated nature of the paper there are of course substantial omissions wrt planning, action and causality approaches in general. In fact I would not even consider myself capable of giving a fair survey of the current state of the art in these areas (perhaps not even in deduction any more where I feel myself still better "at home"). And of course in such a paper there is no need to justify each of the contributions from scratch (even for invited papers there are at least unwritten space limits). If such a need would be, one should not have invited me to give this talk in the first place. 9 out of 11 of your questions are essentially answered by these statements, as I will briefly show you shortly, and have been disappointing in this regard.

Apart from the (in more than one respect incomplete) survey character there is one novel contribution in this paper which is the extension of reasoning about change and causality within the LCM (the action part of TL) frame to include (or to be embedded into) also classical reasoning. In other words the paper marries two complementary areas which belong together. Given this fact the focus of any questions wrt this particular paper should actually be directed to this particular contribution rather than to the contributions which have been done many years ago (as far back as 1985) and to a large extent by students and close colleagues of mine rather than to me. Now to your questions in (boring but time-consuming) detail. As to the paper's section numbering I refer to the version in the web (not to the one in the proceedings).

ad 1.

> *provide* assessments *or* validation results *for new theories*

Since there is no new theory I do not think this particular paper is in lack of such an addition, nor is this the case for the old theories surveyed in the paper. Classical deduction is so established that I am sure you agree wrt this part. As to the action part I know from you personally that you (as well as others) appreciate Thielscher's work especially for his contributions in this respect (relation of the fluent calculus with Lifschitz' $\mathcal{A}$ etc). In my Section 6.3 I survey the equivalence results between the fluent calculus and the action part of TL. On the basis of these theorems his results apply to TL in the same way.

ad 2.

> *My question is what can be said in general about the method you propose for qualification: when is it known to work, and when not?*

In line with the paper's survey character I properly state in Section 7.2 to which you refer here the source of this approach: "The solution is again adapted to TL from [Thi96]'s FC using the example discussed there in great detail." In that paper Thielscher indeed gives theorems and proofs addressing your question. He is the right person to answer specific questions in this regard. But I can add a general comment to issues like

> *For example, suppose there are two cars, A and B, it is known that a potato is put into the tailpipe of one of them, and one asks whether car B will start properly. In such a case, the absence of positive knowledge that the tailpipe of car B has been plugged, does not allow one to draw the default conclusion.*

The advantage of staying within the framework of logic is that you have a long experience on your side (and this is my research methodology which I followed for more than 3 decades). TL can of course handle your example (and also the other one you mention) easily since it features (a resource sensitive) disjunction.

ad 3.

> *Since there is no obvious definition for optimality in this respect, I wonder which quality measure do you use, and what is the proof that the LCM method is optimal with respect to it?*

Assuming the method is semantically correct (discussed in 1) there is of course an "obvious definition of optimality" namely the ones familiar from AD. Any explicit frame axiom increases the search space for the deductive mechanism (and even the proof lengths). Since LCM, TL, FC, LL have no frame axioms at all they are optimal in this respect.

ad 4.

> *So in what sense do we "need to" categorize the actions?*

Your question refers to the paper's Section 7.1 in which I again clearly stated: "The discussion in this section closely follows [Thi97b] ...". So, given the equivalence results mentioned above, your question should again better be addressed to Thielscher directly who in his paper and his habilitation thesis (also cited in this part) provides excellent answers to your question.

ad 5.

> *Also with respect to your formalization of Thielscher's example with three switches and a relay, it is remarkable that the electric circuit in question can easily be understood in terms of dependencies and persistence, but the proposed formalization requires the axioms to represent the propagation of* changes*: "if this fluent changes in such-and-such a way, then that fluent also changes in such a way". This seems clumsy and counterintuitive. Do you claim that it is the best possible representation in the present state of the art?*

If you know a less "clumsy and counterintuitive" formalization of the example then use it. TL as any logic is a neutral formalism which does not bother about the way particular scenarios are represented in them (as Bob Kowalsky convincingly argued in the memorable Crete workshop in 1985 where, in your presence and bombarded by your criticisms, I first presented LCM - as you can still read in the proceedings transcript of the discussions then).

ad 6.

> *My questions are:*
>
> - *What are the advantages of your approach over the one I just referred to? (This is a valid question in view of your statement "A better solution in this sense must be preferred to a deficient one").*

Again this four paragraph Section 7.4 summarizes the work reported [BT97] and [Gro96] so that the answers to your question are better to be looked up in the original sources rather than in my paper and addressed to the authors rather than to me.

> - *In PMON(RCs) and other approaches that use explicit metric time, it is straightforward to make statements about durations, comparing durations of actions, and so on, basically because each interpretation represents an entire history of the world. How can this be done in a logic like TL where the => operator takes one from one state to the next, and updating the current time in the process?*

It is obvious how to state metric time durations in TL (and in fact illustrated in the lifting example of that section) by having Tt before => and $T(t+d)$, d being the duration, after it (the example specializes to $d = 1$). I am not sure at this point whether Bornscheuer and the other people at Dresden working on time did already consider this more detail. My survey is definitely not a complete one.

ad 7.

You write "section 6 shows how the various aspects involved in reasoning about actions and causality can be taken into account within TL". However, nothing in section 6 or elsewhere in the article presents any concrete results about how spontaneous change in the world can be represented - it is as though the world were entirely static when no actions are taken. The resulting concept of causality is quite meagre.

Again, my survey is definitely not (meant to be) a complete one even within the LCM family of approaches let alone all the rest. Spontaneous change has been formalized again by Thielscher in the references given in Section 7. I will include a pointer to that issue to avoid the (false) impression of "meagreness". Your other point concerning differential equations does not tell me anything new and my remark in the paper remains anyway correct.
ad 8.

Yes, but how is this different from the use of the Answer predicate which was proposed and used by Cordell Green in the late 1960's?

The technique is of course the same and well established in logic. So what is the point?
ad 9.

Does the same solution also apply if some of the effects of the general case do not *arise in the special case?*

No, it does not. It only applies in "occasions of a similar nature" as stated there. Again Thielscher is the expert on these issues.
ad 10.

If the answer to the previous question is positive ...

Since it is negative I skip this point.
ad 11.

I believe that some additional coverage of modern results in the latter area would be appropriate

I disagree as explained at the outset.
Best regards, Wolfgang

From: Erik Sandewall on 28.1.1998

Dear Wolfgang,
A question can serve as a means of obtaining information, but it can also be the syntactic form that's used for questioning a proposition. For those of my questions that were of the former kind, you have provided valuable answers - thank you. However, for those that are of the latter kind, I want to pursue our discussion a bit further on two specific issues: methodology of research in our area, and citation principles. Both of them have a general interest, I believe, and for simplicity let's address them one at a time.
I'll start with the citation principles. The following are two cases containing a quotation from your article, my question or comment, and your answer:

1. Concerning the answer literals or state predicate:

> *For that purpose, [Bibel, 1986a] introduced state literals,* S(x)*, which keep trace of the states passed through while executing a plan. ... By unification the variable denoting the goal situation will then ... always provide a term which expresses the linearized sequence of actions.*

>> Yes, but how is this different from the use of the Answer predicate which was proposed by Cordell Green in the late 1960's?

> *The technique is of course the same and well established in logic. So what is the point?*

2. Concerning the combination of differential equations and logic:

> *Modelling continuous processes within a logic has become an active area of research though. ... [Herrmann and Thielscher, 1996]. Another issue concerns the integration of differential equations and their computation within a logic such as TL.*

>> When you write "another issue concerns the integration of differential equations and their computation within a logic such as TL", the uninformed reader will not easily guess that the integration occurred nine years ago, although not for TL.

> *Your other point concerning differential equations does not tell me anything new and my remark in the paper remains anyway correct.*

I must say that I find your reactions exceptional. It is true that it's always difficult to produce correct and complete references to earlier work; it is difficult for example to give adequate coverage of, and comparisons with approaches different from one's own. In the beginning of your answer, you gave an eloquent description of the problem that we all encounter: you are frustrated by a lack of reference to your own work (even in a case where it was not relevant, actually), and you don't have the space to give an adequate presentation of the work of others.

There is no easy solution to this problem, but would it be possible to agree on a few elementary principles, such as:

- Don't quote a well established result as your own innovation
- Don't try to make earlier work invisible, for example by quoting a topic as a current area of investigation and entirely omitting references to earlier work on that topic that you do know about
- Most importantly, if by accident you get out of line wrt principles such as these, agree to make a correction.

These points are fairly obvious, I would think, although I'm afraid they're not always practiced. There may be a number of other principles that ought to be added as well; the present ones resulted from reflecting on the two concrete cases that I quoted above.

Besides a clarification of policy, I think we need additional mechanisms for feedback and correction of citations. The traditional mechanism, namely, the confidential peer reviews, is clearly not sufficient. One obvious way of proceeding is to use ETAI's review mechanism using public discussion, which allows anyone to contribute questions or suggestions to the author in a transparent but undramatical way. This may be used for pointing out additional references that are relevant for the article.

Using a public forum for this feedback to the author has the additional advantage that it can help to resolve disagreements. If an author is unwilling to accept a suggestion for additional citations, it becomes possible for others to listen to the arguments and to form their own opinion. Then at least, if there is a citation fault it's less likely to propagate to later articles written by others.

I am sure that you have also thought about these issues, in particular as the Research Notes Editor of the Artificial Intelligence Journal, and I look forward to your comments on this important topic.

- Erik