Electronic News Journal on Reasoning about Actions and Change

Editor: Erik Sandewall

Volume 2, Number 7

Contributions during July, 1998

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

Linköping University Electronic Press
http://www.ep.liu.se/ej/enrac/1998/07/

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, or from the ENRAC service page specified below. 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 ENRAC *service page,*

`http://www.etaij.org/rac/nj/`

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.

For all material in the Electronic News Journal on Reasoning about Actions and Change, the copyright belongs to the original author specified in the Journal. When no author is indicated, the copyright belongs to the Editor. Everyone submitting a text to the ENJ agrees, by doing so, that such text may be copied and used freely for all academic purposes, as long as it is not changed or misrepresented with respect to form, contents, or authorship.

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, versions as Newsletters in plaintext, HTML, and latex/postscript during the month of July, 1998. Then the contents were edited and formatted, resulting in the present, formatted version which was published on 16.7.1999, 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 25 copies, legally archived, and made available for distribution.

Vol. 2: *223-223*, 31.7.1998
http://www.etaij.org/rac/nj/9807/

The Month of July

Summary of News Journals Contents

July being a month for both vacations and undisturbed research, it is perhaps not surprising that we have seen lively discussions about four of the submitted articles, as well as in the continuing debate about ontologies of time.

Looking back at the development of the ENRAC discussions since they started a year ago, we can see a development where topic-oriented “panel” discussions dominated during the first time, but are now being replaced by discussions about submitted articles. We can also see the variation in volume that was mentioned last month: the June issue of this News Journal was quite small, but the present issue is larger than usual. The electronic format permits us the possibility of adapting the size to the needs.

Electronic News Journal on Reasoning about Actions and Change 224
Vol. 2: *224-231*, 31.7.1998
http://www.etaij.org/ra/rac/006/

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

Erik Sandewall

Logic-Based Modelling of Goal-Directed Behavior

Q2. **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.

A2. **Erik Sandewall** (17.7)

> *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.*

My understanding is that the subtitle "preliminary report" is used in conference publications in order to distinguish an article from a later, revised version of the same article that is to be published elsewhere. The present version of the article is intended to be the final one, except for the corrections that follow from the review comments.

> *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).*

See the first item in the answer to Anonymous Referee 3.

> *An essential, most basic property of any useful axiomatization is consistency. Maybe the author could add a result in that direction.*

I guess this will be a natural corollary of validation results relative to an underlying semantics.

> *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.*

Composite fluents are used in the examples in section 2.5 (although it's possible that order between those two sections ought to be reversed). Treating logically equivalent composite fluents as equal does not have any obvious advantage, and it will complicate things considerably for the occlusion predicate.

> *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?*

This is for reasons of convenience rather than logical necessity. The tentative specifications in question are easy to follow when they are written in terms of the abbreviations, but not when the abbreviations are expanded. Furthermore, verifying the important properties of the specifications becomes quite messy. All of this becomes clearer if the specifications are written without the use of the abbreviations.

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

It is used in axiom S2 on page 5, where it indicates one of the cases where an action does not start executing from the time it is invoked. The reason for making this exception is that an action of the form $a;b$ can not be allowed to start executing at a time when the action a is already executing. Therefore, it is not sufficient to say that an invoked action starts executing unless *the same action* is already in the midst of execution.

Q3. **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.

A3. **Erik Sandewall** (17.7)

> *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.

My argument goes in two steps:

1. Rational agent behavior in a robot needs a precise specification, using logic
2. Such a specification can gain considerably by being integrated with a logic of actions and change.

First, rational agent behavior is fairly complex in the case of a robot that interacts with the real world. It requires pursuing goals, choosing plans, making appropriate reactions to new information, dealing with uncertainty and temporal constraints, and so on. Even though in the end this behavior is going to be realized by software, it is clear that a high-level specification of that software will be necessary for the actual software engineering as well as for making it possible to validate that software and to prove its properties.

Furthermore, if this software is described in a purely algorithmic style, it is difficult to avoid introducing a lot of administrative machinery. This is exemplified by the article "A Formal Specification of dMARS" by Mark d'Inverno et al. (Proceedings of the ATAL-97 workshop, pages 155-176), where a variant of PRS is specified in the Z notation (which is a software specification language). In my approach, the agent's rational behavior is specified in terms of restrictions: for each scenario, the set of models for the formalization is intended to equal the set of rational behaviors, each interpretation being one possible history in the world. Such a constraint-oriented approach has chances of being much more concise and to the point.

Finally, once we decide that it is appropriate to specify rational robotic behavior in logic, we also have to deal with actions. The robotic behavior consists of actions, so this already defines a connection point. Also, when several agents are involved, the actions of one need to be observed and understood by the others. In the longer perspective, I think we will see a lot more integration between theories for reasoning about actions and for characterizing agent behavior.

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.

The text preceding the formula was:

> As another simple example, consider the case of actions which are described in terms of a precondition, a prevail condition, and a postcondition, where the postcondition is at the same time the termination condition for the action...

The formula in question therefore refers to a limited class of actions, rather than to my approach as a whole. (Besides, in the case you mention, the postcondition of the compound action ought to be the conjunction of the two conditions, that is, $\mathit{CutleryClean} \wedge \mathit{CrockeryClean}$. Each action is assumed to have *a postcondition*, in the singular, in the example).

> *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?*

Yes. This deductive process, which one may think of as a control module, is modelled as receiving formulae from two sources: both from a higher software layer of the robot itself ("decisions") and from the surrounding world via the sensors ("observations"). Both kinds of "signals" may influence actions of the agent. This is discussed in section 3.5, the last paragraph on page 12.

> *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.*

Such an example is forthcoming.

> *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.*

This is easily arranged (although there were in fact good reasons for the order in which this material is presently presented).

Q4. **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.

A4. **Erik Sandewall** (17.7)

> *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.*

A proof that correct conclusions are obtained for one or two toy domains does not prove that the formalism is good for any interesting purpose either. As indicated in subsection 5.1 ("Limitations of these results"), it remains to validate the proposed approach using a plausible underlying semantics. This is in line with the methodology for research in this area that I have defended at a number of occasions, beginning in my IJCAI-93 paper. Only so can one identify the range of applicability of the approach in a reliable way.

You might say that the paper ought not to be published until such a validation exists and can be included. This would be an excellent position to take, provided that it were shared by the present community. However, even a quick look at the literature in our field will show that that's not the way things work: Citations to earlier work refer almost exclusively to the *approaches* proposed by the earlier author, and quite rarely to the *theorems that were proved* in the earlier paper.

Besides the formal validation, it's the applications that will prove what the formalism is good for, but this would also take too long in the present paper.

> *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.*

I have of course done a few toy examples for my own sake; the details are too long and boring to merit going into the paper. With respect to the historical comparison: the bugs in those cases were in the entailment method rather than the axiomatization. The approach described here is not dependent on the exact formulation of the axioms, and would not collapse if some of the axioms were to require debugging. Compare subsection 5.1.

Another comparison with the work on the frame problem in the 1980's is more appropriate: those were the days of first exploration of the problem (besides for a small number of articles in the 1970's), and it is quite natural and respectable that one started by looking for approaches that were at least intuitively plausible. *After that start* it was natural to look for precise and proven properties of those approaches. Ramification and causality have followed suit, and qualification is in its early stages. The characterization of rational robotic behavior is just in its very first stage of development, that is all.

> *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.*

Well, it wasn't Rob who suggested it; you refer to my answer to his question. Anyway, the full characterization of a rational agent is bound to be fairly extensive, and the distinction between the actual and the believed value of a fluent is one of the aspects of the full problem. I believe it is best to address this complex problem by divide-and-conquer, so several other papers including references [11] and [12] "complement the work reported here", as stated in the second paragraph of section 1.2. Those articles use distinct predicates for actual and estimated feature values, so the problem you mention has indeed been addressed in the context of the present approach and we have shown how to deal with it.

The present paper works with the simplifying assumption that the agent has perfect knowledge of the fluents. This is in particular in the interest of the reader, since it makes for heavy reading to address all problems at once.

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?

By making both p and p' false; this is a standard technique. However, I don't recognize having mentioned it in this paper?

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?

No, the proof does not require that assumption.

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.

The theorem has the form "In all the models for the axioms, P. Conversely, Q". It is correct that Q follows from P even without the use of those axioms, and the second part of the theorem is merely for showing the conclusion from another point of view. Maybe it would be better to write it as a separate comment.

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.

These bugs will be corrected.

Electronic News Journal on Reasoning about Actions and Change 232
Vol. 2: *232-233*, 31.7.1998
http://www.etaij.org/ra/rac/007/

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

Iliano Cervesato, Massimo Franceschet, and Angelo Montanari

The Complexity of Model Checking in Modal Event Calculi with Quantifiers

Q3. **Rob Miller** (17.7)

Dear Iliano, Massimo and Angelo,

I've several questions about your paper "The complexity of Model Checking in Modal Event Calculi with Quantifiers" submitted to the ETAI.

Generally, I would like to understand the relationship between the modal logic event calculi you describe and refer to in your paper and the classical logic versions of the event calculus that people such as myself and Murray Shanahan use. So any remarks you have in this respect would be helpful. Here are some more specific questions related to this:

- Do you see your formalisms as what we describe as "narrative" formalisms - i.e. formalisms in which the flow of time is described independently of any actions that occur?
- Is the underlying concept of time linear? By "partial order" of events, do you mean "incomplete information about a total order", or are your techniques applicable to genuinely partially ordered time structures (such as the forward branching time structure employed by many versions of the Situation Calculus)?
- I'm confused about the syntax of your language(s). Are there any meta-predicates in your language other than "br"? Perhaps some examples would help. How roughly might you express information such as:

  "Pushing causes the door to open if it's not locked"
  "The door was pushed at 9 o'clock"

  (In the classical logic variants of the event calculus, you'd write formulas such as

  $\forall t[\mathit{Initiates}(\mathit{Push}, \mathit{Open}, t) \leftarrow \neg \mathit{Holds}(\mathit{Locked}, t)].$
  $\mathit{Happens}(\mathit{Push}, 9).$

Apologies if answers to these questions are already to be found in the other papers to which you refer. As a proponent of "the event calculus" I regret not knowing more about your previous work.

Rob

Electronic News Journal on Reasoning about Actions and Change 234
Vol. 2: *234-235*, 31.7.1998
http://www.etaij.org/ra/rac/009/

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

An Inductive Definition Approach to Ramifications

Summary of the article

This paper describes a new approach to the ramification problem. The approach consists of several contributions.

First of all we show that the standard view on causal laws as rules serving to restore the integrity of state constraints is too restrictive. We argue that causal laws essentially represent the propagation of physical or logical forces through a dynamic system, and that state constraints may be a by-product of this propagation process but that this is not necessarily the case. Therefore we introduce causal laws as pure effect propagation rules independent of state constraints, and define a transition function only dependent on the causal rules which maps each state and set of actions to a resulting state.

Second, we argue that in order to obtain a natural and modular representation of the effect propagation process, the language of causal rules needs to allow for recursion to model mutually dependent effects, as well as for negation to model effects which propagate in the *absence* of other effects. To define a semantics for complex effect theories involving both negative and cyclic dependencies between effects, we observe that a fundamental property of the process of effect propagation is that it is *constructive*: effects do not spontaneously appear without an external cause; each effect must be caused by a preceding effect and in practice effects are generated in a bottom-up way. To adequately model this constructive nature of physical change propagation, we base the semantics of the formalism on the principle of inductive definition, the main mathematical constructive principle.

We use a generalised inductive definition principle, which in fact generalises several different approaches. For theories without recursion, the semantics we propose coincides with completion (and can as such be seen as a generalisation of explanation closure approaches). For theories without negation, the semantics coincides with positive inductive definition semantics and standard circumscription (which has also received a lot of attention in research on the frame problem). For stratified theories the semantics coincides with iterated inductive definitions and with the perfect model semantics for stratified logic programs. In the most general case the proposed operator used in the fixpoint calculation coincides with the well-founded operator in logic programming.

The proposed semantics allows to deal with any effect theories without syntactic restrictions on cycles or negations. As a consequence, syntactically correct effect theories exist which do not have a constructively well-defined model. In particular this is the case for effect theories containing actual cycles over negation, e.g. when the existence of each of two effects depends on the other effect's absence. The physical systems corresponding to such effect theories are rare, and those that exist (those deliberately constructed to yield such an effect theory) exhibit unstable behaviour. But the distinction between good and bad effect theories can not be made based on syntactical properties. Our semantics is such that it detects and marks bad effect theories while it assigns a unique transition function to good, constructive, effect theories, including some that appear syntactically problematic.

A third contribution is the introduction of so-called complex effect rules which provide a higher-level description of effect propagations. These complex effect rules allow to state that it is the change in truth value of a complex fluent formula, rather than the change of a single fluent, which triggers certain effect propagations. Complex effect rules map to sets of low-level definition rules. We have observed that one form of complex effect rule is sufficiently expressive to model all kinds of temporal reasoning examples known to us, which also suggests that the corresponding low-level rules only occur in particular combinations in practice. The high-level effect rules offer a much more natural and compact way of describing ramifications than the low-level rules. In particular they prove to be very effective for modeling cumulative or canceling effects of simultaneous actions.

An essential property of the semantics, which distinguishes it from several other approaches, is that it is deterministic: for a given effect theory, any combination of a state and set of actions yields a unique set of causations and hence a unique resulting state (sometimes an inconsistent one, in which case the set of actions is not allowed in the given starting state). In our view nondeterminism should be explicitly represented if it is intended, hence we do not allow nondeterminism to arise from special combinations of otherwise deterministic rules. We do describe an extension of effect rules which allows for a limited form of nondeterminism, in that the rules can state that some effects are allowed, but not forced, to occur. This approach has been further extended to more expressive nondeterministic rules, but this issue is beyond the scope of the current paper.

Our approach is presented independent of a specific time structure or formalism, such as Situation Calculus, $\mathcal{A}$ or Event Calculus. We briefly discuss how it can be embedded in these different time structures.

We compare our approach to related work in the literature. In particular we show a correspondence between our nondeterministic effect rules and the causal rules proposed by Thielscher. Differences are, apart from the nondeterministic aspect, especially the different view on delays in effect propagations.

Vol. 2: *236-250*, 31.7.1998
http://www.etaij.org/ra/rac/009/

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

Marc Denecker, Daniele Theseider Dupré, and Kristof Van Belleghem

An Inductive Definition Approach to Ramifications

Q1. **Murray Shanahan** (7.7)

Marc, Daniele and Kristof,

I haven't read your paper very closely yet. But I'm very interested to understand how your formalism (and indeed any other formalism) works with the circuit example of Figure 1 in your paper.

I can't find anywhere in the paper a precise claim about what your formalisation of this example yields. But presumably, when switch r is closed, switch s opens, $relay1$ is activated, and switch q opens, while $relay2$ remains not activated and switch p remains closed. (By the way it would be nice if the relays were labelled in the diagram.) Is this what your formalisation yields? I guess it is.

To relieve me of the trouble of going through your definitions and proving it myself, can you also tell me what your formalisation yields in the case that switch s is left out altogether. This is obviously more problematic, since the constraints are then contradictory. From r we get $relay2$, which gives $\neg p$, which gives $\neg relay1$ which gives $\neg q$ which gives $\neg relay2$.

Do we get inconsistency according to your formalisation? Or do we get multiple possible outcomes? On page 7, you say that "only theories that do not entail multiple changes of the same fluent at the same time are considered as well-defined". So would this example not be well-defined? Yet this is a realisable circuit. Indeed it's a simplification of the circuit you use as an example. I'm not quite sure what would actually happen - I suppose the relays would get stuck. But something would happen. So maybe inconsistency or no-well-definedness is not what we want in such examples. Or maybe the representation just needs to be a bit more elaborate. What do you think?

Murray

A1. **The Authors** (13.7)

Dear Murray,

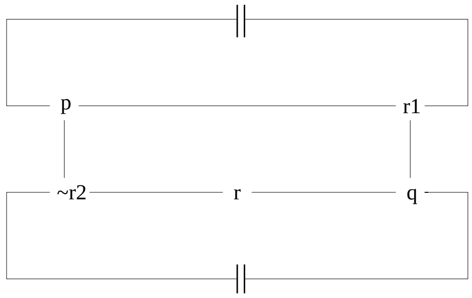

Figure 1: Simplified circuit example.

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. Figure 1 shows how the circuit looks. The state constraints in this example are:

$$r1 \leftrightarrow p$$
$$p \leftrightarrow \sim r2$$
$$r2 \leftrightarrow r \;\&\; q$$
$$q \leftrightarrow r1$$

If follows logically that :

$$\{p \ \& \ r1 \ \& \ q \ \& \ \sim r \ \& \ \sim r2\}$$

This corresponds to the unique stable state:

$$\{p, r1, q, \sim r, \sim r2\}$$

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.

$r1 \leftrightarrow p$ yields:

$$\begin{array}{ll} c(r1) \leftarrow c(p), \sim h(p) & \\ & (\mathit{initiating}\ p\ \mathit{causes}\ r1) \\ c(\sim r1) \leftarrow c(\sim p), h(p) & \\ & (\mathit{initiating}\ \sim p\ \mathit{causes}\ \sim r1) \end{array}$$

$r2 \leftrightarrow r \ \& \ q$ yields:

$$\begin{array}{ll} c(r2) \leftarrow c(q), \sim h(q), c(r), \sim h(r) & \\ & (\mathit{initiating}\ q\ \&\ r\ \mathit{causes}\ r2) \\ c(r2) \leftarrow c(q), \sim h(q), h(r), \sim c(\sim r) & \\ c(r2) \leftarrow h(q), \sim c(\sim q), c(r), \sim h(r) & \\ c(\sim r2) \leftarrow c(\sim q), h(q) & \\ & (\mathit{initiating}\ \sim (q\ \&\ r)\ \mathit{causes}\ \sim r2) \\ c(\sim r2) \leftarrow c(\sim r), h(r) & \end{array}$$

$q \leftrightarrow r1$ yields:

$$\begin{array}{l} c(q) \leftarrow c(r1), \sim h(r1)... \\ c(\sim q) \leftarrow c(\sim r1), h(r1) \end{array}$$

$p \leftrightarrow r2$ yields:

$$\begin{array}{l} c(p) \leftarrow c(\sim r2), h(r2) \\ c(\sim p) \leftarrow c(r2), h(r2) \end{array}$$

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

$$\begin{array}{l} c(r) \leftarrow act(\mathit{close\_r}) \\ c(\sim r) \leftarrow act(\mathit{open\_r}) \end{array}$$

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:

$\{h(p), \sim h(r2), h(r1), h(q), \sim h(r)\}$

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:

$$
\begin{aligned}
&c(r) \leftarrow \\
&c(r2) \leftarrow \sim c(\sim q), c(r) \\
&c(\sim p) \leftarrow c(r2) \\
&c(\sim r1) \leftarrow c(\sim p) \\
&c(\sim q) \leftarrow c(\sim r1) \\
&c(\sim r2) \leftarrow c(\sim q)
\end{aligned}
$$

The well-founded model of this rule set is

- true: $c(r)$
- undefined: $c(r2)$, $c(\sim q)c(\sim r1)c(\sim p)$
- false: all other literals ($c(\sim r)$, $c(p)$, ...)

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:

$c(r) \leftrightarrow true$

$c(r2) \leftrightarrow \sim c(\sim q) \leftrightarrow \sim c(\sim r1) \leftrightarrow \sim c(\sim p) \leftrightarrow \sim c(r2)$

$c(l) \leftrightarrow$ *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 $c(\sim r2)$ and contains the unique prooftrees for the atoms $c(\sim q)$, $c(\sim r1)$, $c(\sim p)$, $c(r2)$, $c(r)$

$$
\begin{aligned}
&c(\sim r2) \leftarrow c(\sim q) \leftarrow c(\sim r1) \leftarrow \\
&\quad c(\sim p) \leftarrow c(r2) \leftarrow c(r) \leftarrow true \\
&\qquad \leftarrow \sim c(\sim q)
\end{aligned}
$$

(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

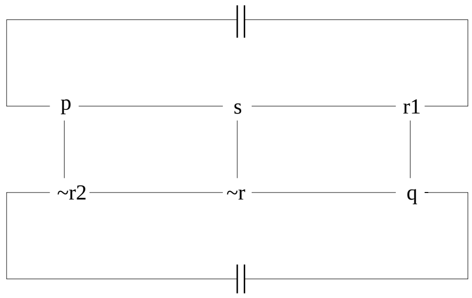

Figure 2: The original circuit example.

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, as shown in figure 2. The effect rules in the circuit are related to the following state constraints:

$r1 \leftrightarrow p \;\&\; s$

(if switches p and s are closed then relay $r1$ is active)

$r2 \leftrightarrow q \;\&\; r$

(if switches q and r are closed then relay $r2$ is active)

$r \leftrightarrow \sim s$

(switches r and s are mechanically connected this way)

$p \leftrightarrow \sim r2$

(p is closed iff $r2$ is not active)

$q \leftrightarrow r1$

(q *isclosediff* r 1 *isactive*)

It can be seen easily that this theory simplifies to

$p \;\& \sim r1 \;\&\; (s \leftrightarrow r1 \leftrightarrow q \leftrightarrow \sim r)$

Hence, the system can occur in two stable states

$\{p, \sim r1, s, r1, q, \sim r\}$

and

$\{p, \sim r1, \sim s, \sim r1, \sim q, r\}$.

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:

$caus(r) \leftarrow act(close\_r)$

$caus(\sim r) \leftarrow act(open\_r)$

$caus(s) \leftarrow act(close\_s)$

$caus(\sim s) \leftarrow act(open\_s)$

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

C1-1. **Murray Shanahan** (17.7)

Marc, Daniele and Kristof,

Many thanks for your answer to my question about your paper. I think it's an excellent paper.

I've just completed a short paper on the same topic, which you may be interested in. I've submitted it to the AIJ. Basically it shows how the event calculus from my book "Solving the Frame Problem" can be used to tackle the ramification problem, including tricky examples like Thielscher's circuit. The paper is available via the following URL.

`http://www.dcs.qmw.ac.uk/~mps/ramifications.ps.Z`

All the best, Murray

Q2. **Eugenia Ternovskaia** (11.7)

Dear Marc, Daniele and Kristof,

It was interesting to read your paper, especially because I am interested in inductive definability myself. I wrote a paper "Inductive Definability and the Situation Calculus" presented at Dynamics'97 last October. It was published in a volume of LNCS later. You can get it from
`http://www.cs.utoronto.ca/~eugenia/papers/indef.ps`

Among other things, I describe an inductive solution to the frame and ramification problems. I treat causal rules specifying direct and indirect effects of actions as rules of an inductive definition. Causal (inductive) rules specifying direct effects of actions may contain both cycles and negations. Causal rules describing indirect effects may not contain cycles. This condition can be relaxed, but in this case the correct form of successor state axioms will not be obtained. I would be grateful to you if you could read this paper and provide some comments.

Here are some comments on your paper. First of all, it's a good paper.

The main contribution, I think, is that you further elaborated the generalized inductive definition principle and applied it to the ramification problem. I consider Marc's result about the equivalence between this principle and the well-founded semantics significant, and it was good to see how it works in the context of the ramification problem. It was nice to see examples for the syntactic subclasses of the generalized inductive definition.

I think that introducing the third truth value helps you to explain the subtle details of the iterative process which was quite difficult for me since I was working in classical first order logic with fixpoints.

From my experience, it is a long way from understanding how causal (inductive) rules work and capturing it in a classical logic-based formalism such as the situation calculus. In your paper you say that your approach can be embedded in the situation calculus. Perhaps the main question to you is how the third truth value would be represented in the two-valued setting?

The following will help you to make a few little improvements, I hope.

p. 28, after Th.1.

Unless nondeterminism is explicitly introduced, our theory leaves no room for ambiguity.

I do not think this is an advantage of your theory. I think the truth value "u" should be obtained in two cases.

First, it may be a result of a "bad axiomatization". Second, it may appear due to lack of complete information about the current state of the world. (It seems that you apply the Closed World Assumption in your definition of a state as a set of (positive) fluent literals, if I understood correctly.)

I believe your definitions might be easily changed to represent the second case as well. For example, a state could be associated with a set of "known" positive and negative literals.

It also seems strange that you cannot derive anything about "holds", although I understand your motivation. Would it be difficult to introduce this feature?

p. 10 ..., where $\top$ denotes the inconsistent state.

This is a confusing notation. Usually $\perp$ is used to denote false or contradiction (as well as the bottom of a lattice) and $\top$ is used to denote true (or the top of a lattice), as you perhaps know. I noticed that you already use $\perp$ later in the text, but maybe another solution is possible. For example, in automata theory $\emptyset$ is used to denote a "junk" state without outgoing transitions.

p. 10 Footnote 8: I did not understand the second sentence at first reading. "Initial state", "presence and absence" are confusing.

p. 23, footnote 17: A more or less meaningful "real-world" example could be something like this: an object cannot be at location p and q at the same time (but must be somewhere), and actions *move-to-p* and *move-to-q* are always applicable.

p. 28, top. completely orthogonal to successor state calculation is too strong, see comments for p.38-39.

p.38-39 This approach contains several contributions: - A complete uncoupling of ramifications from state constraints...

I do not think they can be uncoupled **completely**. Even in your theory, ramifications impose some restriction on your function Trans, I believe. Trans maps states and actions to states. In this sense reachable states are restricted and state constraints are implied. Can you clarify?

The other direction, from state constraints to ramification, as described by Thielscher, is a nice idea. As far as I know, no one claims that this is *the only way* to obtain causal rules specifying indirect effects. It is a useful method to obtain some of these rules, even though it does not always work.

Also, in your summary you say "the standard view on causal laws as rules serving to restore the integrity of state constraints is too restrictive." From talking to different people, I never had the impression that this is the only and standard view on the role of causal rules in the community.

Overall, even if you think your statement is absolutely right in its current form, it's strange to see it as a contribution #1. I think you have done more interesting things. Again, it's a good paper.

I hope my comments will be helpful.

Regards,

Eugenia

A2. **The Authors** (17.7)

Dear Eugenia,

Your comments are very useful. We didn't have a close reading of your paper yet, but a first look showed that there is strong correspondence between our approaches. We have had very similar intuitions. Thanks for the careful reading and the corrections.

> *Perhaps the main question to you is how the third truth value would be represented in the two-valued setting?*

We think there is no need to represent the third truth value in a 2-valued setting. Observe that the transition function is a partial mapping from 2-valued states (and sets of actions) into 2-valued states. IF an expert designs his effect theory such that for all states which satisfy the state constraints and sets of actions which satisfy the action preconditions, the transition function is correctly defined, (i.e. the model of the rule set is 2-valued and no contradictory fluent literals are caused), and IF this effect theory is represented correctly in situation calculus, the transition of that state and those actions will be defined and will be 2-valued. So, in this case, a third truth-value is not necessary.

In your approach, you impose a syntactical restriction (no cycles) which guarantees that the effects are unique and well-defined. Our approach does not impose a syntactical restriction, but there is still the methodological constraint that the expert should design his effect theory such that it leads to well-defined transitions (at least for the above mentioned states and sets of actions). In case this methodological rule is followed, it is not necessary to introduce a third truth value in situation calculus or event calculus. This methodological rule requires from the expert that he establishes the 2-valuedness of the well-founded models of his effect theory at least in all states satisfying state constraints and actions sets satisfying action preconditions.

This is a point that deserved more attention.

> "Unless nondeterminism is explicitly introduced, our theory leaves no room for ambiguity."

> *I do not think this is an advantage of your theory. I think the truth value "u" should be obtained in two cases.*
>
> *First, it may be a result of a "bad axiomatization". Second, it may appear due to lack of complete information about the current state of the world.*
>
> *(It seems that you apply the Closed World Assumption in your definition of a state as a set of (positive) fluent literals, if I understood correctly.)*
>
> *I believe your definitions might be easily changed to represent the second case as well. For example, a state could be associated with a set of "known" positive and negative literals.*

You propose to use the truth value "u" for two different purposes: as a pointer to ill-defined cause literals and as a pointer to unknown atoms. This may be problematic. When a cause literal is 3-valued, how does the expert know whether to interpret it in one way or the other?

In our view, "u" is reservated for pointing to ill-defined causal literals. Uncertainty caused by nondeterministic actions must be modeled explicitly through nondeterministic effect rules and is modeled by having different transition states (see the section about this topic). We do not model uncertainty on the input state; we do not need to do so. We expect that in any approach which integrates our effect theories in a full temporal language (e.g. EC or SC), uncertainty on the state at a certain time point is represented by having different models with different states at that moment. In each single model, the state is well-defined and can be represented as a 2-valued set on which our transition function can be applied. In our experience, this is a better way to model uncertainty than through belief sets. We can't see in what way we would be applying the Closed World Assumption.

> *It also seems strange that you cannot derive anything about "holds", although I understand your motivation. Would it be difficult to introduce this feature?*

This is not really true. The result of the transition function defines the holds predicate in the next state. One can verify that the next state is derived from the previous state and the causal literals as described by the normal situation calculus persistence axiom.

> *p. 23, footnote 17: A more or less meaningful "real-world" example could be something like this: an object cannot be at location p and q at the same time (but must be somewhere), and actions move_to_p and move_to_q are always applicable.*

We agree that the footnote may not be correct, but it's not clear to us what you mean by that example. Another example of a theory with negative loops is given in the Murray Shanahan's variant of the circuit example. Though we consider the semantics of the rule set as ill-defined, the rules are certainly meaningful.

> p.38-39 This approach contains several contributions: - A complete uncoupling of ramifications from state constraints...

> *I do not think they can be uncoupled* completely*. Even in your theory, ramifications impose some restriction on your function Trans, I believe. Trans maps states and actions to states. In this sense reachable states are restricted and state constraints are implied. Can you clarify?*

This remark is related to your other remark:

> *Also, in your summary you say "the standard view on causal laws as rules serving to restore the integrity of state constraints is too restrictive." From talking to different people, I never had the impression that this is the only and standard view on the role of causal rules in the community.*

First, let us clarify what we mean by uncoupling: we define the semantics of the effect rule theory without taking state constraints into consideration. In many other approaches, this is not the case: in fact in most approaches causal rules directly imply a state constraint, they are a state constraint augmented with some directionality-of-propagation information; an exception to this is Michael Thielscher's approach, but there state constraints are used in the calculation of the possible successor states, in that effect propagations may stop as soon as a state satisfying the constraints is reached. In our approach the successor state calculation is in no way dependent on the constraints. It is always carried out to its end and yields a unique successor state. However, it is true that the end result of the calculation is checked against the state constraints. These constraints may cause rejection of the found successor state, but will in that case not calculate another (valid) successor state.

Observe moreover that ramifications need not imply state constraints even if they restrict Trans. Of course a restriction of Trans rules out certain transitions, but that does not need to mean that any states are ruled out: a state is ruled out only if all transitions leading to it are invalid. The "counter" example in our paper gives an example where this is not the case: whenever an output changes to true, the counter must be augmented. But a resulting state in which the counter were not updated (but all other effects applied) need not be invalid in itself. This state might well have existed at an earlier time in the history of the network. It can only not be reached from the currently given state. So ramifications need not imply any state constraints.

It is absolutely true that in many cases, there is a relation between state constraints and ramifications; in that case the method of Michael Thielscher is applicable and works fine in our approach as well (we applied it in the circuit example; see our answer to Murray Shanahans question). In this case, the set of effect rules should satisfy the following correctness criterion: that any state satisfying the state constraints and any set of actions satisfying the action preconditions and occurring in this state lead to a transition state in which the state constraints are satisfied. This is not automatically satisfied but should be proven in our approach.

Also we discussed with a number of people about the issue of the relationship between state constraints and effect rules; had anybody contradicted us on this point, we would not have put it the way we did. But we have no problem in weakening the point. Still, it seems true that in most approaches, state constraints are underlying or are directly involved in describing the semantics of causal rules. Maybe this is not a deliberate, conscious choice, but in the existing formalisations it is apparently a fact.

Marc, Daniele, Kristof

C2-1. **Eugenia Ternovskaia** (27.7)

Dear Marc, Daniele, Kristof,

Thank you for your answers to my questions. These answers clarify an important difference between our works.

Using generalized inductive definitions to detect ill-defined literals is the advantage of your work.

However, after 2-valuedness is shown, all reasoning involving multiple models have to be performed on the metamathematical level. In contrast, first order logic augmented with inductive definitions allows us to work with all models at once therefore allowing for incomplete information. Perhaps your approach could be developed in this direction, but you do not address this problem in the paper. Notice I do not talk about non-determinism here.

I completely agree with your point of view about non-deterministic actions.

Therefore it seems that we mostly developed different parts of the problem.

> *In your approach, you impose a syntactical restriction (no cycles) which guarantees that the effects are unique and well-defined.*

The syntactic requirement of acyclicity is by no means inherent to my approach. It simplifies the presentation in the LNCS paper. Otherwise I would have to consider the cases where the correct syntactic form of successor state axioms is not obtained, thus making the paper even longer.

To verify my method, I worked out the Gear wheels example. The approach works just fine, moreover, it produces successor state axioms in the correct form. Starting from your set of rules, I obtain the following successor state axiom:

$$turn\_1(do(a,s)) \leftrightarrow (a = start\_1 \lor a = start\_2 \lor \\ turn\_1(s) \land a \neq stop\_1 \land a \neq stop\_2)$$

And similarly for $turn\_2(do(a,s))$. To avoid the cyclic regression problem, regression has to be done with respect to the causal rules for direct effects rather than with respect to successor state axioms. This is the way I do it in the paper Causality via Inductive Definitions. See

> *http://www.cs.utoronto.ca/~eugenia/publications.html.*

> *We can't see in what way we would be applying the Closed World Assumption.*

I use this term in a sloppy way here to say that a state is represented by a set of positive fluent literal, and the literals not in the set are false in this state.

> It also seems strange that you cannot derive anything about "holds", although I understand your motivation. Would it be difficult to introduce this feature?

> *This is not really true. The result of the transition function defines the holds predicate in the next state. One can verify that the next state is derived from the previous state and the causal literals as described by the normal situation calculus persistence axiom.*

Perhaps I should have chosen a different wording. I used "derive" which I understand as "derive in a formal proof system". I did not mean "makes true in a particular model".

> *Observe moreover that ramifications need not imply state constraints even if they restrict Trans. Of course a restriction of Trans rules out certain transitions, but that does not need to mean that any states are ruled out: a state is ruled out only if all transitions leading to it are invalid. The "counter" example in our paper gives an example where this is not the case: whenever an output changes to true, the counter must be augmented. But a resulting state in which the counter were not updated (but all other effects applied) need not be invalid in itself. This state might well have existed at an earlier time in the history of the network. It can only not be reached from the currently given state. So ramifications need not imply any state constraints.*

I do not understand neither this comment nor the “Counter” example. The rules imply a state constraint relating any two consecutive states. This constraint is perfectly valid.

Finally, a question about Definition 1. You say “and body B a nonempty set of positive and negative literals of D”. Do you mean “and body B, a nonempty set of symbols from D, including $\{t\}$ and $\{f\}$“?

Regards,

Eugenia

Q3. **Michael Thielscher** (14.7)

Dear Marc, Daniele and Kristof,

I have a question concerning the notion of negative cycles. If I understand correctly, they arise whenever the absence of some causation is the cause of some effect. Now, on page 23 you consider a construction like

$$caus(p) \leftarrow \neg caus(q)$$
$$caus(q) \leftarrow \neg caus(p)$$

ambiguous or nonsensical. But what if you specified, say, the effect of tossing a coin by "if heads isn't caused, then tail is" and "if tail isn't caused, then heads is." The expected model would be that either of heads or tail is caused, which is what the stable models semantics proposes. Of course it is questionable whether the above is a good specification of tossing a coin, but I would argue that this is due to the notion of "absence of causation as cause" being rather odd from an intuitive point of view in general. I therefore wonder if this notion is really needed? The suitcase scenario, for example, can be handled with Lin's approach (and others) without this notion.

Also I would like to make a remark on your discussion on zero vs. non-zero delays (Section 6). This is a very interesting question, and I have recently discovered a related problem in connection with my method of using causal relationships: If effects with zero delay are not distinguished from effects with real delay, then some desirable conclusions may not follow. Since your approach has much in common with mine, I wonder if the distinction between so-called steady and stabilizing state constraints, which solves the problem in my approach, is needed here as well? (C.f. my paper entitled "Steady vs. Stabilizing State Constraints" at this year's CommonSense workshop; see

`http://www.ida.liu.se/ext/etai/nj/fcs-98/listing.html`

Michael

A3. **The Authors** (18.7)

Michael,

Here is an answer to your comments.

First of all, we agree with you that it is unnatural to have just the absence of an effect as a cause for anything. In fact as you can see in the paper, our high-level effect rules (which syntactically look a lot like Lin's) contain only the presence of an effect (on a complex formula) as possible trigger for further changes. But at a low level, we need to introduce negative cause literals to define initiation of complex formulas.

For example, when is the formula $sw1 \wedge sw2$ initiated? Is it initated when $sw1$ is true in the initial state and is made false and $sw2$ is made true simultaneously? No: it is initiated when both are made true, or when one of them is made true and the other is already true and is not made false. To represent this, a negative causal literal is needed. Observe however that any syntactically correct high-level rule corresponds to a set of low-level rules in which each rule contains at least one positive cause literal. There is always one actual effect at the basis of any propagation, but the propagation can be prevented by other effects - hence the negative cause literals.

One could argue that one should define that $sw1 \wedge sw2$ is initiated by comparing two successive states (and so there would be no need for resorting to low-level rules), which is what you do in your approach. However, our approach is constructive in that the next state is computed using the previous one and the computed causal literals. This definition of the next state is of course only mathematically sound if we can compute the caused literals without knowing the next state already.

In Lin's approach (the IJCAI95 paper), absence of causation is not present because the formalisation is entirely based on the values of fluents, not on their changes. One might say that all (present or absent) causations are implicit because it is the resulting state that is considered rather than the transition, as apparent in the rule

$up(L1, s) \wedge up(L2, s) \rightarrow Caused(open, true, s)$

Wrt our approach, things are somewhat mixed since no concurrent actions are considered in that paper, but it's true that the approach would also work for the concurrent flipping of the two switches. The fluent-triggered causal statements are similar to our complex initiation formulae. One difference is of course that Lin's rules necessarily imply the corresponding state constraint. Related to this is the fact that a number of "useless" (even if correct) instances of "Caused" can be derived for any situation where $L1 \wedge L2$ is true but not initiated. "Useless" here just means that it does not give rise to any changes, but just maintains a reason for *open* being true. For cases where the causal rules have a corresponding state constraint, it is probably a matter of taste if our formalisation or Lin's one is preferable. In our version

initiating $L1open \wedge L2open$ causes $open$

it is the change of $L1open \wedge L2open$ that causes a change in *open*. In Lin's view, it is the state of $L1open \wedge L2open$ that causes the state of *open*.

This different view allows our approach to be more general: in cases where causal rules do not correspond to state constraints, such as example 1 in our paper (with the electronic counter), the relation is inherently dependent on the value change: it's the change of "out" from true to false that makes "count" change. This is not a case where absence of initiation is needed, since the dependency is on the single fluent "out" but how would this be modeled in Lin's causal rules?

We had another example (in our NAIC'97 paper and in Kristof's PhD thesis): an alarm system that detects if somehow people enter a building. While the system is active, anyone entering the building triggers the alarm: if *in* (someone in the building) becomes true when *active* is already true, *ring* becomes true. Under one interpretation of that example, the alarm does not ring if it is deactivated at the same time where someone enters.

Then we would have the rule:

$caus(ring) \leftarrow init(in), holds(active), \neg init(\neg active)$

Could the same be achieved in Lin's causal rules? A rule

$in(s) \wedge active(s) \rightarrow Caused(ring, true, s)$

is not correct precisely because the state constraint

$in \wedge active \rightarrow ring$

is not true: we suppose in fact that the bell does not ring if the system is activated where someone is already in.

Regarding your interpretation of

$caus(p) \leftarrow \neg caus(q)$
$caus(q) \leftarrow \neg caus(p)$

as a representation of $caus(p)$ xor $caus(q)$.

As indicated above, we agree that having only "not caus" literal as the body of a causal rule is counterintuitive, precisely because such rules are not constructive. However negative cycles can also occur if there are positive caus literals present (just add $caus(r)$ to both rules).

We wanted the constructive principle of inductive definition to model the propagation of effects. It is hard to interpret the above two rules as a well-defined inductive definition. As you say, it could be interpreted as a cause for p or for q, but it seems (as you also say) to be a very peculiar way of modeling this. We prefer to model that there is either a cause for p or for q explicitly by using a nondeterministic effect rule (something like *toss* causes *heads* xor *tails*). This may be a matter of taste, but we do not want unintended nondeterministic effects, therefore, we force the expert to make nondeterministic effects explicit.

It is true that in stable semantics the above clauses do represent the exclusive disjunction p xor q. But we do not regard stable semantics as the right sort of semantics for modeling definitions. For more discussion on this, see the paper by Marc which appeared at NM-98. In Kristof's PhD thesis a version of our approach is presented in a narrative-based (i.e. Event-Calculus-like) temporal structure, also including actions with alternative effects like tossing a coin. Only a more restricted form of nondeterminism is in the ETAI paper, but the idea in the semantics is the same.

As regards your "Steady vs stabilizing state constraints". There, you say that mixing the two types of ramifications is responsible for the unintended conclusion in the example. From our point of view, the problem in the specific example would be that the causal rules are not correct: in our formalisation we would say that $up(lhs)$ causes *stain* if $notup(rhs)$ is *and remains* true, i.e. $\neg caus(up(rhs))$. Again, absence of causation solves the problem, but it again appears in combination with a positive causation, due to an interaction of causes. So on the basis of this example we can not give a definite answer.

Nevertheless, the issue of delays is intriguing, and your distinction between zero and non-zero delays is an interesting contribution, even if there could be a more general view on it. First, it could be possible that even what you consider zero delays (steady constraints and ramifications) are an abstraction wrt reality. Maybe in quantum physics it is not true that an object (a sub-atomic particle, in particular) cannot be in two locations at the same time? (is any expert of quantum physics reading ETAI?)

More importantly, you argue that the important distinction in qualitative reasoning is between zero and non-zero delays, but still small enough to be considered virtually instantaneous by common sense, so to be distinguished by actually "delayed" effects. That makes sense, even if one could also view a continuum between zero delays (or delays that are too small for classical physics), non-zero but "commonsense" instantaneous delays, and delayed effects. Knowledge about different delays could be still qualitative or imprecise, but still allow to conclude that a delay is smaller than another and then it is not correct to apply the (non-zero-delay) causal rules in any order.

Marc, Daniele and Kristof

Vol. 2: *251*, 31.7.1998
http://www.etaij.org/ra/rac/011/

Chitta Baral and Son Cao Tran:

Relating Theories of Actions and Reactive Control

Abstract of the article

This paper has been submitted to the Electronic Transactions on Artificial Intelligence. For the purpose of the review discussion, it has first been published and posted by Linkping University Electronic Press; it is available at http://www.ep.liu.se/ea/cis/1998/009/

In this paper we give a formal characterization of reactive control using action theories. In the process we formalize the notion of a reactive control program being correct with respect to a given goal, a set of initial states, and action theories about the agent/robot and about the exogenous actions. We give sufficiency conditions that guarantee correctness and use it to give an automatic method for constructing provenly correct control modules. We then extend our approach to action theories and control modules that have specialized sensing actions and that encode conditional plans. Finally we briefly relate our theory with the implementation of our mobile robot Diablo, which successfully competed in AAAI 96 and 97 mobile robot contests.