Electronic News Journal on Reasoning about Actions and Change

Editor: Erik Sandewall

Volume 3, Number 4

Contributions during April, 1999

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

Linköping University Electronic Press
http://www.ep.liu.se/ej/enrac/1999/04/

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.

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 April, 1999. 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.

Electronic News Journal on Reasoning about Actions and Change 29
Vol. 3: *29-29*, 30.4.1999
http://www.etaij.org/rac/nj/9904/

The Month of April

Summary of News Journals Contents

An important policy issue for the ETAI came up during the month of April, namely: what is the appropriate refereeing decision for a controversial article, where opinions differ strongly and there is no definitive answer to the question about the 'quality' of an article? The case arose in the reviewing of an article by Graham White, and our decision was to accept the article for the present News Journal, together with the record of the discussion that led up to that decision, but not to accept it for the ETAI. In this way, the author has had a chance to define and to defend his position, and to put himself on the record accordingly. We then leave it to each reader to decide his or her own opinion about the reported work.

The present issue of the News Journal also contains the record of a concurrent discussion on another article, by Levesque, Pirri, and Reiter. The refereeing of that article has not been concluded during the present month of April.

Vol. 3: *30-39*, 30.4.1999
http://www.etaij.org/ra/rac/012/

Protocol of on-line discussion during April, 1999
about the following research article:

Graham White

Simulation, Ramification, and Linear Logic

A3. **Graham White** (4.4)

Anonymous Referee 1 wrote:

> 1. Are the results of the article, as specified in the summary, of significant interest for researchers in our area?
>
> *No. The main objective of the paper is to propose a solution to the Frame Problem which satisfies certain properties, which are convincingly argued for. My only but serious concern is that existing solutions seem to satisfy all of the requirements and that the author fails to argue for merits of his approach in comparison. In particular Reiter's theory [1] of generating successor state axioms from simple effect axioms fits perfectly the author's desiderata.*

No, it doesn't. One of my criteria is that the data for the problem should be what I describe as local, or, as Sandewall puts it (in a passage which I cite in the article),

> *Consider applications where every scenario describes a number of separate but interconnected objects, different scenarios involve different configurations, and each action has its immediate effects on one or a few of the objects, but indirect effects on objects which are connected to the first ones, in some sense of the word 'connected'.*
> *... action laws should only specify the "immediate" or "primary" effects of the action, and logical inference should be used for tracing how some changes "cause" other changes across the structure of the configuration at hand.*

However, Reiter's account starts from a statement of what he and Pednault describe as "effect axioms" (p. 360 of Reiter's article), which give, for each action and each fluent, conditions under which that fluent positively or negatively affects the fluent. These conditions, because they talk about all fluent-action pairs rather than just the fluents which are immediately affected by the action, are not concerned with the immediate or primary effects of actions but all of them.

Now Reiter's paper does not explicitly consider the ramification problem, and, if we do not have ramification, it is entirely reasonable to suppose that actions can be described by effect axioms: actions then have no effects but their immediate effects. And it is certainly true that, in the absence of ramification, Reiter's solution does satisfy my desiderata (or, at least, the ones that I give in Table 1, apart, of course, for the requirement that the effect axioms should be local, which has little content in the absence of ramification).

But, I am talking about a case with ramification, so that Reiter's methods do not apply; if we allow the effect axioms to talk about all effects, instead of just the immediate ones, then there is essentially no distinction between immediate and non-immediate effects.

Neither can we hope to derive the unrestricted effect axioms from the local ones using merely classical logic (this, of course, is the assumption in Reiter's article; the effect axioms are supposed to yield all of the effects, local or non-local, using classical deduction). Consider the following example:

> *I move a block to a place L (this is the immediate effect of the action; the location of the block is changed). If the block is in a wet place, it becomes wet. (This is the description of how effects propagate). Location L is wet in the initial state.*

We want, of course, to conclude that the block becomes wet. However, in order to do this, we have to conclude that location L is wet, not in the initial state, but in the successor state. And we cannot conclude this by classical logic alone; considerations of examples like this (electrical circuits are a fruitful source) shows that, if we have a chain of propagating effects, we must, in general, solve the frame problem for the propagating causes up to stage i before we can know whether to apply the rules for the propagation to stage $i + 1$.

> *... And yet this work, which is among today's most cited papers in the area of Reasoning About Actions And Change, the author does not even mention, let alone compare it to his own method.*

No, that's true, I don't, but I do cite Murray Shanahan's book Solving the Frame Problem, which does contain a very clear and accurate account, not only of Reiter's work, but also of the work of Pednault and others which led up to it.

C3-1. **Anonymous Referee 1** (13.4)

Dear author,

You wrote:

> ... And yet this work, which is among today's most cited papers in the area of Reasoning About Actions And Change, the author does not even mention, let alone compare it to his own method.

No, that's true, I don't, but I do cite Murray Shanahan's book Solving the Frame Problem, which does contain a very clear and accurate account, not only of Reiter's work, but also of the work of Pednault and others which led up to it.

If I am not mistaken then Shanahan's book is mentioned only in a footnote and only as a side remark, not for the purpose of referring to an extensive account of work related to your's.

> My only but serious concern is that existing solutions seem to satisfy all of the requirements and that the author fails to argue for merits of his approach in comparison. In particular Reiter's theory [1] of generating successor state axioms from simple effect axioms fits perfectly the author's desiderata.

But, I am talking about a case with ramification, so that Reiter's methods do not apply; if we allow the effect axioms to talk about all

But Reiter's approach has been extended to dealing with state constraints and ramifications [2,3]. Several other successful approaches to the Ramification Problem emerged during the past five years or so; I mention [4,5,6]. Repeating myself, I am missing convincing arguments why your approach is substantially different and superior in comparison.

[2] F. Lin and R. Reiter, State constraints revisited, J. of Logic and Computation 4(5), 1995.

F. Lin, Embracing causality in specifying the indirect effects of actions, IJCAI 1995.

N. McCain and H. Turner, Causal theories of action and change, AAAI 1997.

M. Thielscher, Ramification and causality, AIJ 89, 1997.

M. Denecker etal, An inductive definition approach to ramifications, ETAI 3, 1999.

C3-2. **The Author** (27.4)

Here's my reply to the referee's remarks.

> But, I am talking about a case with ramification, so that Reiter's methods do not apply; if we allow the effect axioms to talk about all

> *But Reiter's approach has been extended to dealing with state constraints and ramifications [2,3]. Several other successful approaches to the Ramification Problem emerged during the past five years or so; I mention [4,5,6]. Repeating myself, I am missing convincing arguments why your approach is substantially different and superior in comparison.*
>
> *[2] F. Lin and R. Reiter, State constraints revisited, J. of Logic and Computation 4(5), 1995.*
>
> *F. Lin, Embracing causality in specifying the indirect effects of actions, IJCAI 1995.*
>
> *N. McCain and H. Turner, Causal theories of action and change, AAAI 1997.*
>
> *M. Thielscher, Ramification and causality, AIJ 89, 1997.*
>
> *M. Denecker etal, An inductive definition approach to ramifications, ETAI 3, 1999.*

Let's start with what appears to be a misconception. The referee writes "state constraints and ramifications", as if the two were synonymous. They are not, and one of the articles he refers to (by Deneker et al.) explains why not ([6], pp. 3f.):

> *we consider causal laws to be representations of how changes propagate. As such, causal laws may be but do not need to be related to state constraints. Below is an example.*
>
> *Assume a network with many input signals which can be directly modified through actions, an arbitrary internal structure, and one or more output lines. A counter on one of these output lines counts the number of times the output voltage changes from 0 to 1. [rules follow]*
>
> *The indirect effects described by the above rules are in no way related to a state constraint: there is no relation between the contents of the counter and the current state of the network."*

So there are examples which formalisms based on state constraints cannot cope with: this rules out the referee's reference [2].

Of the other approaches, [3], [4] and [6] suffer from the limitation (which Deneker et al., in [6], admit to) that multiple changes of the same fluent at the same time cannot occur. One can give well motivated examples which exhibit this behaviour: consider, for example, a modification of Deneker et al.'s counter example in which the counter is now connected to two of the output lines, and in which the input action may cause changes in either or both of them. The propagation of the effects may well result in the occurence of intermediate states of the counter when one of the output lines has had its effect but not the other.

Similarly, the electrical circuit example in Thielscher's [5] (p. 341, Fig. 5) has intermediate states of fluents (namely, the light bulb being on) which are caused during the ramification process but which do not persist into the final state. Deneker et al. do describe such examples as "not well defined", but they do not argue for this restriction.

So my approach, which correctly handles examples such as those given above, is more general than all of the cited approaches apart from Thielscher's [5], which also handles these approaches correctly.

And regardless of the success or failure of my own approach, what we have done so far is to eliminate all but one of these supposedly "successful approaches to the Ramification Problem" by referring to arguments and examples found in the other ones. This makes me a little suspicious of the rules of this refereeing game as it is played in the AI community. After all, if I say "The sky is blue" and the referee says "No, you're wrong, the sky is green", then I may at least want to verify that the sky is, in fact, blue, just to check. But if the referee says "you're wrong, the sky is green AND the sky is red", then what do I reply? And if the referee cites supposedly "succesful approaches to the ramification problem", and not all of them can be simultaneously succesful because the later ones argue against the earlier ones (so that either the earlier ones are wrong, or the arguments and examples in the later ones are wrong); well, what can be concluded from that, other than that there are papers in the literature which disagree with mine (no surprise) and also with each other (also no suprise).

What I want to insist on is the argument of my paper, and particularly the argument in Section 2.4. What I argue there is this: that the standard approach to these problems is unduly sensitive to vocabulary, because we can make formalisations, which seem to be adequate, of the same problem in vocabularies which seem to be logically equivalent (because they are equivalent in first order logic), and we find that when we apply the circumscription procedure to them, that one is successful and the other is not. And we seem to have no way of selecting one vocabulary rather than the other, other than that one gives a successful circumscription and the other does not. So this circumscription procedure does not seem to be anything that we can apply to a problem that we do not know the answer to: regardless of how adequate our formalisation might seem to be, we always have to face the possibility that there might be another formalisation (equivalent in first order logic, but using a different vocabulary) which gave different results under circumscription. We seem to have no way of differentiating between good vocabularies and bad, other than by looking at the results.

Almost all of the cited approaches are, at least prima facie, susceptible to this sort of problem; they all rely, to some extent, on a preferred vocabulary – explicitly in [2] and [3], because we have fluent-based circumscription; implicitly in [4], because of the definition of interpretations there as sets of fluents, and in [6] because of the use of reification in the definition of the fluent calculus. My own approach is not: I refer, in the article, to the fact that "in [the technical report] we show that we have a good notion of interpretation for theories of this sort, and that, in our example, we can change from the [one vocabulary to the other] without breaking anything". (p. 12 of the article). Furthermore, I have a semantics for my logic (also in the technical report); all of the above I am willing to expound at greater length if required (though I am still somewhat attached to the principle of separating the more philosophical and expository arguments from the mathematical details, and that was what I had hoped to accomplish in the separation between article and technical report).

One could, finally, talk about the great computational advantages of approaches based on linear logic when compared to those based on circumscription or other of the conventional non-monotonic logics.

So, finally, if you want a summary of what I think the advantages of my approach are, they would be:

- Correctness on a wider range of examples than any other approach, apart from Thielscher's [6];
- Extensionality: that is, the formalism admits a robust notion of change of vocabulary (what is technically known as interpretation), and this notion of interpretation is supported by a semantics;
- Computational efficiency.

C3-3. **Murray Shanahan** (28.4)

Here's a small contribution to the Graham White paper debate.

First, let me point out that I haven't (I'm ashamed to say) got around to reading Graham's paper, yet. So I'm not in a position to have any overall opinion on its merits. But I have followed the debate, and this has led me to the following thought.

Graham lists the following advantages of his approach.

> - *Correctness on a wider range of examples than any other approach, apart from Thielscher's [6];*
> - *Extensionality: that is, the formalism admits a robust notion of change of vocabulary (what is technically known as interpretation), and this notion of interpretation is supported by a semantics;*
> - *Computational efficiency.*

Certainly (i) on its own, if true, suggests that the significance of the paper's contribution is sufficient for a journal. (I can't comment with respect to the other usual criteria for acceptance, not having read the paper.)

But I think Graham is selling himself short. Merely to match the capabilities of existing formalisms using a radically different logic, namely linear logic, is itself a worthwhile contribution, assuming it's a non-trivial exercise. Redoing existing work in a very new way adds to our knowledge of the space of possible action formalisms, and could lead in all sorts of new directions. (Of course, the new way of doing things has to be radically different to be justified. Otherwise we end up with a proliferation of uninteresting variations on the same old theme.)

Murray

A4. **Graham White** (4.4)

Anonymous Referee 2 wrote:

> *My recommendation as regards the acceptance of this paper for ETAI is no. Accordingly, I have perhaps to explain that.*
>
> *Roughly speaking, my answer is no, because there is a very little difference between the submitted and revised versions of the paper. On the other hand, the remarks, especially those made by Andreas Herzig, were quite important, and so they should be seriously taken into account in the final version. I am not going to be very detailed, but one example should illustrate this.*
>
> *Andreas writes:*
>
>> I guess this relates to my main criticism: in order to identify actions in proofs, you must group together sequences of rules related to the same action.
>
> *Answer of Graham:*
>
>> No, not necessarily; if you look at the examples in 'Actions, Ramification and Linear Modalities', you can see that you don't need to group rule applications together like this.

Yes, in retrospect I agree that that was a little short. Here's a fuller account of my strategy with this.

I give (on p. 10 of my article or so) a logical system which is an extension of linear logic by a modal operator $\Diamond$; the important rule for $\Diamond$ is

$$\frac{\Gamma \vdash A, \Delta}{\Gamma \vdash \Diamond A', \Delta} \qquad *$$

where A' is obtained from A by a ramification transition. We must, however, add new rules to the system in order to be able to prove cut elimination: in particular, we must add the rule

$$\frac{\Gamma, A' \vdash \Diamond B, \Delta}{\Gamma, A \vdash \Diamond B, \Delta} \qquad **$$

where, again, A' is derived from A by a ramification transition.

We then prove, as I have said, cut elimination for this system; as with most such proofs, the proof of cut elimination proceeds by rather elementary stages but is extremely long-winded (28 subcases or so).

Now we can consider our representation of ramification and actions. We first describe our actions by their primary effects, namely by linear logic formulae of the form

$$a' \multimap a''$$

where a' and a'' roughly correspond to Strips-like "delete lists" and "add lists", respectively. (They do not exactly correspond, because in order to be able to perform this action, the fluents in a' must be part of the initial situation, whereas the Strips delete list is only a list of the propositions which will be deleted if they are present.) Call an action, described like this, α. Suppose also that we have a modality, $\Diamond$, which represents our ramification transitions.

Now we can use the cut elimination theorem to show that the only valid proofs of a sequent of the form

$$A, \alpha \vdash \Diamond B$$

correspond to the following sequences of rule applications:

i) Applications of (**) to A, yielding a state $a' \times A'$ (I represent the linear tensor product as *);

ii) Applications of the rules for * and $\multimap$, yielding a sequent

$$a'' \times A' \vdash \Diamond B$$

iii) Applications of either (*) or (**), corresponding in either case to a ramification transition from $a'' \times A'$ to B.

iv) The identity axiom on either $a'' \times A'$ or B, depending which rules we have applied.

One can also prove that, in many cases, one can apply the rules in any order one likes; one can formalise this in so-called permutability theorems, saying that the applications of certain rules can be permuted around each other. In the case of the rules applied in this proof, we are free to permute rule applications except for the obvious restriction, namely that if A does not actually contain the "delete list" a' of the action, then (**) must be applied before we apply the rule for $\multimap$.

However, the proof-theoretic restrictions here correspond to our intuitive picture of what it is to perform an action in the presence of ramification: that is, the initial situation may first undergo a sequence of ramification transitions, then the action should accomplish its effects, and finally there should be a sequence of ramification transitions leading to the final situation. And in the "standard case", as it were, of the ramification scenario, we start with a situation A which is already stable under ramification transitions, so no initial applications of (**) are possible; we have no choice but to apply the rules for * and $\multimap$ first.

So, in order to represent action in this system, we do not have to impose any artificial grouping of rules: the proof theory gives us only the proofs that correspond to realistic sequences of action and ramification.

Q6. **Area Editor** (28.4)

The outcome of this refereeing process is inconclusive.

By the conventional wisdom, journal reviewing is supposed to identify which papers are to receive the quality stamp of approval, and which are not. This works fine if the quality of an article can in fact be identified by the review reports. But what do we do if there is a considerable remaining *uncertainty* - not that the paper is "medium good", but at the end of refereeing we still *do not know* what is the case – we do not have the basis for a reliable verdict?

The present article is apparently such a case. Both the two anonymous referees were strongly critical of the article. The author has answered to their criticism, and defended his case well, but some questions do remain. A first and obvious question is - if the proposed approach (as is claimed) is superior to all previously published approaches *except one* which it does not surpass, should it then be accepted to the journal? Many may be tempted to say no, but then quite a number of articles in the literature of this field ought not to have been accepted, so precedent is not consistent with such an abstract principle. Murray Shanahan, in his statement today, argues that it is in fact worthwhile to publish more than one approach in such cases. This is in itself an important methodological question in our area, where some continued exchange of opinions may be useful.

From the point of view of the editorial decision, the debate has answered some questions but raised some new ones. In particular, consider the author's argument in his concluding remark that his approach has also two other advantages, besides epxressiveness, namely its extensionality and its computational efficiency. The problem is that neither of these advantages is very much substantiated in the article. With respect to extensionality, it would be necessary to show the advantage in concrete comparison with the most natural alternatives. The article makes an argument about problems in alternative approaches due to intensionality, but the problems he describes can possibly be explained as due to the use of domain constraints in the examples he refers to. One would have liked to see an argument why Thielscher's approach (which is otherwise mentioned as the contender) does not have the proposed extensionality property.

Similarly, the author makes a general argument that "there is a well-developed technology for implementing proof search in these systems", but again a more concrete comparison with implementations of the major alternative approach(es) would be needed in order to defend this argument.

These are objections to the effect that the article's arguments are incomplete, and not necessarily wrong. Some of these questions might be resolved by continued debate, but we also need to get to a decision with respect to the journal acceptance. The discussion can continue anyway, regardless of whether the article is accepted or not. And finally we must observe that the use of linear logic in itself often arouses strong feelings in the community, so it may not be possible to arrive at a consensus.

With this we return to the original question: what is the most appropriate decision when the quality of an article is undetermined, with a high degree of uncertainty in the estimate? If a journal is seen as a means of active communication between researchers, which presumably is how modern journals started, then controversial contributions ought to be very highly valued. On the other hand, if journals are assigned prestige levels based on the perceived quality of their articles, and other articles in the same journal are assigned a quality level based on the journal's prestige, then it is in the interest of other authors that only uncontested articles be accepted.

We may regret the contemporary practice of evaluating articles based on the ranking of journals, both because that practice institutionalizes lack of insight, and because of its side-effects on the editorial policies of the journals, but it is a fact of life and can not be ignored.

Our solution to this dilemma uses the fact that we also publish the *Electronic News Journal on Reasoning About Actions and Change*, besides the ETAI. The primary purpose of the News Journal is to publish, on a monthly basis, the debate contributions and debate notes that appeared in the Newsletters, so that they also obtain formal, published status and can be cited.

In any case, it seems appropriate that Graham White's article ought to be accompanied by the discussion that we have had here. In that way, it becomes transparent that the article is taken seriously by the community, but that there are strong differences of opinion about the proposed approach, and it is left to each reader to consider the arguments pro and con, and to make her or his own assessment.

The conclusion, therefore, is to decline the submitted article for the ETAI, but to offer the author that the article be published in the ENRAC (News Journal) for April, 1999 together with the discussion and the present concluding comment. The continued discussion will then appear in later issues of the News Journal, like for all other articles.

Electronic News Journal on Reasoning about Actions and Change 40
Vol. 3: *40-50*, 30.4.1999
http://www.etaij.org/ra/rac/017/

Protocol of on-line discussion during April, 1999
about the following reference article:

Hector Levesque, Fiora Pirri, and Ray Reiter

Foundations for the Situation Calculus

A1. **Ray Reiter** (3.4)

John McCarthy wrote:

> *I have a grumble about the title of the article by Reiter, et. al. about situation calculus. It presents a particular system of the situation calculus, and therefore can't quite serve as a reference article.*

John's grumble about our title – Foundations for the Situation Calculus – is perhaps justified. A better title might have been "Foundations for a Calculus of Situations". If Erik will permit it, maybe we can make that change in the archival copy. Hector suggested that if only we were centred in Denmark, we could have used something like "Foundations for the Copenhagen Interpretation of the Situation Calculus". Too bad.

C1-1. **Ray Reiter** (13.4)

Erik,

In view of the tempest in a teapot that the title of our paper caused (Foundations for the situation calculus), can we change the official version to "Foundations for a calculus of situations"?

Thanks,
–Ray

Q2. **Pat Hayes** (3.4)

John McCarthy's 'grumble' in ENRAC 30.3 (99007) raises a general issue of nomenclature.

The following scenario is quite a common one. A formal term of art becomes generally used to refer to something in the literature which was invented or introduced by someone. After some time has elapsed, however, another author or group re-adopts or re-uses the term in a somewhat different sense, and this second sense is then widely used by a related community, or among a different generation of researchers. This has happened for example to the word "ontology", which is now widely used in the, er, ontology community to refer to what philosophical logicians (who originally had possession of the word) called "axiomatic theories". In our community it has happened to "situation", which is now widely taken to mean what Reiter (et. al.) defines it to be, ie a finite sequence of actions, and its cognates such as "situation calculus", which Reiter (et al) uses to refer to a particular formalization developed at Toronto in the last decade. Neither of these meanings are what these terms meant before Reiter changed the language.

The result of such alterations in meaning is pernicious only if people fail to realize that it has happened, or are unclear about which sense they intend when they use the term; but when this does happen, a great deal of confusion, not to mention emotional heat, can be generated. Once introduced into a community, terminology has a life of its own, and nobody can expect to exert a kind of eternal authority over it. But any useage of such a term should clarify which sense is meant; and any writing which purports to be a definitive *survey* of a topic must be especially careful about such ambiguities, and to make clear to a reader that the word has changed in meaning. Evidently what Reiter and his co-authors mean by "situation calculus" isn't what John McCarthy invented. Regardless of the relative merits of the calculi being referred to, a survey with that title should at least clarify and respect the way that the terminology was used for about a quarter of a century before Reiter re-defined it.

Pat Hayes

A2. **Ray Reiter** (3.4)

Pat Hayes wrote:

> *In our community it has happened to "situation", which is now widely taken to mean what Reiter (et. al.) defines it to be, ie a finite sequence of actions, and its cognates such as "situation calculus", which Reiter (et al) uses to refer to a particular formalization developed at Toronto in the last decade. Neither of these meanings are what these terms meant before Reiter changed the language. ...*
>
> *Evidently what Reiter and his co-authors mean by "situation calculus" isn't what John McCarthy invented. Regardless of the relative merits of the calculi being referred to, a survey with that title should at least clarify and respect the way that the terminology was used for about a quarter of a century before Reiter re-defined it."*

We weren't entirely unaware of the different interpretations that can be made of the sitcalc. This from our abstract:

> *This article gives the logical foundations for the situations-as-histories variant of the situation calculus, focusing on the following items:*

And this from the introduction:

> *The situation calculus (McCarthy [mccarthy63]) has long been a staple of AI research, but only recently have there been attempts to carefully axiomatize it. The variant described here has evolved in response to the needs of the Cognitive Robotics Project at the University of Toronto, and its foundations are now sufficiently developed to be gathered in one place. That is what this paper sets out to do.*

The principal intuition captured by our axioms is that situations are histories – finite sequences of primitive actions – and we provide a binary constructor $do(a, s)$ denoting the action sequence obtained from the history s by adding action a to it. Other intuitions are certainly possible about the nature of situations. McCarthy and Hayes [mccarthyhayes69] saw them as "snapshots" of a world. For the purposes of representing narratives, McCarthy and Costello [costellomccarthy] wish to express that any number of actions may occur between a situation and its "successor", and therefore they do not appeal to a constructor like *do*.

Pat and John's complaints raise some interesting issues about the importance of fully axiomatizing one's intuitions, something that the reasoning-about-actions community has been carefull to do. In their KR'98 paper, John and Tom Costello spelled out a different variant of the sitcalc to the "Toronto interpretation". To my knowledge, this is the only other fully specified sitcalc story, and now there is the possibility of looking at the pros and cons of the two approaches. It would be very interesting to see an axiomatization of the intuitions underlying the earlier "situations-as-snapshots" view. Probably other interpretations are possible of John's original sitcalc language, or extensions to it. Situations are not like the natural numbers; there seems to be no standard interpretation for them, and maybe this discussion will encourage others to step forward with alternative axiomatizations reflecting their intuitions. It remains to be seen whether, after the dust settles, something like "the Peano situation calculus" will emerge.

C2-1. **John McCarthy** (3.4)

I see that I didn't make my grumble clear. I think the Reiter, et. al. system is one situation calculus formalism. I think that a reference article should be broader.

C3-1. **Graham White** (13.4)

I'd like to make an extended comment on the debate between McCarthy, Hayes, Reiter (et al.) concerning the nature of the situation calculus, and particularly as reflected in the last exchange between McCarthy and Reiter.

First, a bit of explanatory background: I'm a mathematician (I did a doctorate in algebraic geometry at Oxford with Atiyah), and that's the way I tend to think of things. I know that mathematics tends to be regarded with extreme suspicion by the AI community, but that's my background, and I can't help it.

It seems to me that one of the key assumptions behind the present stage of the debate is: (*) There is a difference between formalisms which express situations as histories and those which regard situations as snapshots. (**) Each choice of histories or snapshots expresses a different intuition about the nature of situations: e.g. Reiter writes

> *The principal intuition captured by our axioms is that situations are histories .... Other intuitions are certainly possible about the nature of situations. McCarthy and Hayes [mccarthyhayes69] saw them as "snapshots" of a world.*

I wish to take issue with these, and particularly with (**). I don't think it's possible, in the present state of the formalism, even to express these questions, much less to answer them.

The problem is this. In standard mathematical practice, one certainly does go through as stage of defining mathematical objects in some formal theory or other: this seems to correspond to what Reiter and his colleagues do when they define what situations are. But one *also* specifies, implicitly or explicitly, what it is for two such definitions to be equivalent.

For example (and this example is taken from Benacerraf's article What Numbers Cannot Be) one can define the natural numbers both as a sequence of sets thus:

$$\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\{\{\emptyset\}\}\}, ...$$

or as a sequence of sets thus:

$$\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} ...$$

Now clearly nothing hangs on this: the two definitions are clearly equivalent in any sense which matters. They can be regarded as two different implementations of the natural numbers.

How (technically) do we express this equivalence? Several ways:

i) By saying that they have "the same constructors" (i.e. we can establish a correspondence between the two sets of constructors, that the arities match up, and so on). This works under certain circumstances (that the two objects are both initial algebras for the functors defined by the constructors) which are fulfilled in this case; the constructors are, of course, zero and successor, and it's easy to see how to define them in each case.

ii) One can show that a) the two objects are isomorphic, and b) that the isomorphisms are natural. This is a more general answer, but you need, in turn, to say what you mean by isomorphism (which probably involves saying what you mean by morphism in general). In our example, we're defining a natural number object in an already defined ambient category (namely the category of sets), so we already have answers to these questions. But in a more general case there may be work to do.

So what I'm really worried about is this: do the "snapshot" and "history" definitions of situations express essentially different (i.e. non-isomorphic, or more exactly not naturally isomorphic) mathematical objects, or is the difference only notational, that is, analogous to the two definitions of the natural numbers above. And answering *this* question involves answering the prior one: what is an appropriate sense of isomorphism for these mathematical objects? More generally: what is an appropriate sense of morphism?

Now Reiter here talks of constructors, and says, quite correctly, that his approach has constructors: for each action α, we get a constructor $do(a, s)$ which maps

$$Actions \times Situations \mapsto Situations$$

But can't we also define similar things for the snapshot view? If there are a number of actions that lead from a snaphsot-situation to its successor, then we also have a map

$$Actions \times Situations \mapsto Situations$$

(and maybe it's only a partial map, but a partial map is still a morphism in some appropriately defined category, and maybe you can still establish suitable equivalences...) So I'm not convinced that there's anything in this distinction.

As far as I can see this is not a straightforward question, because you can approach it from two different directions. One direction comes from the world of process algebras, labelled transition systems, and the like, and we get a definition of morphism here which basically comes from the notion of simulation. Reiter's constructors fit into this picture.

But you can also approach it from the viewpoint of logic (or at least this is what people seem to have been trying to do all these years). And the logical viewpoint also gives you a way of mapping between theories (a mapping is, roughly speaking, an interpretation of one theory in another: but you have to be a little careful here because you actually have a two-category and not just a category). The culturally entrenched logics here are non-monotonic, which adds an extra complication because we cannot just carry over results from classical logic. (This applies particularly to circumscription-like theories, for which the input data is not just a classical theory but a classical theory *together with a set of literals to be circumscribed.* This set of literals breaks most of the interpretation relations between the underlying classical theories. For this reason the oft repeated statement that circumscription is "an extension of classical logic" is to be taken with a pinch of salt.)

So, difficult questions. But answering them seems to be necessary if we are to know what sort of mathematical objects we are talking about. Here's a parallel. Quine said "no entity without identity", by which he meant that, in order to define some class of entities A, it wasn't enough just to say which objects were As; you also have to say when two As are the same A. One could also say (and the parallel is quite exact) "no mathematical object without isomorphism and equivalence"; that is, in order to be able to define a certain class of mathematical objects, it's not enough to be able to define instances of that class, but you also have to be able to say when to instances are isomorphic, and you also have to be able to say what the essential properties are that these objects satisfy, so that you can recognise when two candidate definitions are equivalent.

Anyway, that's my .02 Euro's worth.

Graham White

C3-2. **Pat Hayes** (27.4)

Graham White's extended comments (Enrac 13.4) are interesting, but not for the reason I suspect he had in mind. White, as he says, is approaching the entire discussion from a mathematical perspective. If things like situations (whatever they are), world-states, actions and histories were indeed mathematical structures, then his remarks would be apposite: it would be good professional practice to ask what the appropriate morphisms were between them, for example, and what would count as isomorphism. But they aren't mathematical structures, and this entire game we are playing here isn't mathematics. We are using logic not to describe mathematical structures, but to describe aspects of nature.

This isn't a matter of looking on mathematics with 'extreme suspicion', by the way. (I have a mathematical training myself, and find the logical security of mathematics comforting, if anything.) The point is that what we are wanting to do is more like science than mathematics. There is a real world out there, and our job is to try to describe it adequately. Many people in action reasoning, in fact, see their task as rigorously empirical, and advocate testing their logical theories by incorporating them into actual physical robots with measurable behaviour. Having made that analogy, however, let me immediately modify it. What we often find ourselves doing isn't science so much as a kind of pre-science: the kind of conceptual analysis that precedes an exact science. We are still sometimes arguing about what our words mean. But whether it is scientific or pre-scientific, it is always directed towards, and inspired by, something in the actual world, not by mathematical criteria.

To see the difference, take a mathematical concept such as 'group', and ask how one could know a group if one met it. Something is a group just when it satisfies the group axioms. The axioms are all there is: *anything* that satisfies the group axioms is a group, by definition. One can find examples of many famous finite groups all over the place: the Klein group, for example, is revealed in the possible orientations of a rectangular card. Now, contrast this with a concept like 'banana'. Even an optimistic mathematician would surely not expect the be able to reconstruct bananas within set theory (even up to isomorphism), or give a sensible account of banana-morphisms. Actions and world-states are more like bananas than groups.

White's examples make the contrast very clear. He uses the example of the natural numbers:

> *For example (and this example is taken from Benacerraf's article What Numbers Cannot Be) one can define the natural numbers both*

> *as a sequence of sets thus:*
>
> $\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\{\{\emptyset\}\}\}, \ldots$
>
> *or as a sequence of sets thus:*
>
> $\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} \ldots$
>
> *Now clearly nothing hangs on this: the two definitions are clearly equivalent in any sense which matters. They can be regarded as two different implementations of the natural numbers.*

First, note that the idea of natural number preceded both of these set-theoretical constructions. In fact, it preceded the idea of 'set' by several centuries, if not millenia. (Its easy to forget that the great success of logicism in modern foundations of mathematics tends to produce a historical revision of ideas that go back a very long way into human history; modern accounts of mathematical structure are written from a perspective which *identifies* mathematical domains with their modern reconstructions in set theory from the 1920s, or (still more sophisticated) in category theory from the 1950s.) So even in mathematics, I think that the formalisations tend to be driven by pre-formal intuitions. But why would we expect that actions would form a category, or be definable (even up to isomorphism) in set theory? I see no reason to suppose they would, and several good reasons to suppose not.

But even if they did, there is still a further question to ask, which is how that mathematical structure fits onto the actual world we are aiming to describe.

An example. Consider a ball bouncing on a table. We could say that each period between two bounces is a 'state' of the ball, and that each bounce is a state-transition (one where it might lose a little energy, say, producing a lower bounce next time.) Or, we could say that each moment of contact with the table was a 'state', and each trajectory between bounces a state-transition. These might well be isomorphic structures (finite total orders) in White's mathematical sense, but the kinds of axioms one will write about the 'states' and 'actions' will be very different in the two cases; and if someone is thinking in one way when talking to someone else who is thinking the other way, mutual incomprehension may well result. (Ray Reiter and I are conducting a discussion offline to see if we are misunderstanding each other in exactly this way.)

> *... So what I'm really worried about is this: do the "snapshot" and "history" definitions of situations express essentially different (i.e. non-isomorphic, or more exactly not naturally isomorphic) mathematical objects, or is the difference only notational, that is, analogous to the two definitions of the natural numbers above.*

Answer: neither of these. The difference between a state of a world and a sequence of actions is an ontological distinction, like that between a banana and an orange (or, a better analogy, between a positions and a velocity.)

Pat Hayes

PS a minor note:

> *Now Reiter here talks of constructors, and says, quite correctly, that his approach has constructors: for each action α, we get a constructor $do(a, s)$ which maps*
>
> $Actions \times Situations \mapsto Situations$
>
> *But can't we also define similar things for the snapshot view? If there are a number of actions that lead from a snaphsot-situation to its successor, then we also have a map*
>
> $Actions \times Situations \mapsto Situations$
>
> *(and maybe it's only a partial map, but a partial map is still a morphism in some appropriately defined category, and maybe you can still establish suitable equivalences...) So I'm not convinced that there's anything in this distinction.*

Of course there is nothing in this as a *distinction*. The $do(a, s)$ notation was used to refer to the world-state which results from doing the action a in the world-state s since the very first formulation of the situation calculus ontology.

C3-3. **Graham White** (5.5)

Here's my reply to Pat Hayes. First to correct his misapprehension of where I'm coming from: he writes

> *I have a mathematical training myself, and find the logical security of mathematics comforting, if anything.*

Now I am a mathematician, not a logician, and for my taste I find that there's far too much logic in AI, and far too little mathematics. My background is in category theory, but I also have a great love of the mathematics of the nineteenth century; I'll be expressing myself in terms of category theory, but much of what I say would, I would imagine, be comprehensible to people like Felix Klein or Elie Cartan. Hayes' historical perspective seems to involve an inexorable growth in sophistication ("mathematical domains with their modern reconstructions in set theory from the 1920s, or (still more sophisticated) in category theory from the 1950s") whereas as far as I'm concerned things got very boring from Hilbert till the end of the second world war and picked up again after that.

Now to business. Hayes writes:

> *To see the difference, take a mathematical concept such as 'group', and ask how one could know a group if one met it. Something is a group just when it satisfies the group axioms. The axioms are all there is: anything that satisfies the group axioms is a group, by definition."*

"how one could know a group if one met it"? Well, there are two questions here:

i) What sort of things are groups?

Hayes says “Something is a group just when it satisfies the group axioms. The axioms are all there is”.

Well, no.

Here’s Hayes’ definition (maybe):

> *a group is a set with a composition law such that ...*

and here’s my definition:

> *a group is a category with one object in which every morphism is an isomorphism*

and there are, of course, innumerable others.

Part of the reason for this multiplicity is that we want to use some definitions in some circumstances and some in others: for example, my definition works for things such as algebraic groups, or formal groups, where there is no underlying set.

The other question is, of course,

ii) What is this particular group, for example, the Klein four-group?

Also in this case, it’s not simply the case that something “just is” the Klein four-group or not: Hayes describes it as “the possible orientations of a rectangular card”, but you can also describe it with a multiplication table, or a Cayley diagram, or a matrix representation, and so on. And in each case it’s a non-trivial question whether these definitions are equivalent, or whether some item, given in some way or other, fulfils some particular one of these definitions.

In either case, there is a difference between how things are presented and how they are: we can present things in all sorts of different ways, but they can turn out to be essentially the same. We cannot go by the mere appearance of the definitions.

Now this is entirely similar to real life: I can think of hardly any area in which one does not have to make any distinction between how things appear and how they are. There is, it seems to me, no subject (say “ontology”) with enough authority to say “It is simply an ontological distinction that X is not Y”. (If you think that ontology will deliver you stuff like this, then you ought to come to terms with Quine.)

If we want to talk about the sort of empirical investigation that Hayes talks about – of comparison between formalism and reality (”testing their logical theories by incorporating them into actual physical robots with measurable behaviour”) – then we are relying on implicit ideas about identity, on both sides. If we are testing whether formal theory A fits phenomenon X, then we had better be able to say what it is, on the side of reality, for X to be the same as Y (despite appearances), and, on the level of formalism, for A to be the same as B (despite appearances, or notation, or whatever). Otherwise we cannot claim to be doing science. All the notation, and all the measurement, in the world will not save us if we cannot answer these questions.

So why category theory? Nothing special hangs upon it: but two reasons why it is useful.

The first is that it provides a treatment, within mathematics, of the difference between how things are presented and how things are. This is why category theory is useful: people use it in subjects (such as algebraic topology) where things can be presented in many different ways, but maintain an essential unity despite that. We don't need to use category theory for this: other areas of mathematics play similar roles (for example, the spectral theory of operators plays an entirely similar role within quantum theory), but category theory has the advantage of being fairly topic-neutral and adaptable.

The second is this. Category theory is behind several of the most successful areas of modern mathematics (algebraic geometry, differential geometry, algebraic topology), and what these applications of category theory have in common is this: they are all mathematical treatments of qualitative phenomena. They are very successful: the proof of Fermat's last theorem (though the final result is about numbers) is an offshoot of work in these qualitative areas. Many of the problems of artificial intelligence are problems of formalising qualitative phenomena, and AI seems to lack conceptual tools for dealing with the qualitative (Hayes, for example, takes it for granted that experiment involves measurement, and seems in the field to be a general tendency to confuse the qualitative with the approximate). Category theory provides a range of technical tools which are qualitative, successful, and richly structured.

You might, of course, say – and Hayes seems to be tending in this direction – that this is all mathematics, and Real Life is somewhat different, and that AI is a study of Real Life. To which I would like to reply:
i) in that case, why are AI publications so full of ugly formulae? and
ii) my argument relies on one respect in which mathematics tends to be similar to Real Life, namely that, in both cases, you have to be careful about the distinction between how things are presented and how things really are. In neither case can you jump to conclusions about essential identity, or difference, merely because of notational, or presentational, differences; and this was the point that got me started in the first place.

Anyway, that's that.

Graham White

Q4. **John McCarthy** (5.5)

I have no grumble about the new title. Here are some comments on the substance of the article.

The paper presents a big bang interpretation of the situation calculus. Everything started with $S0$.

1. It is not elaboration tolerant w/r to what might have happened before $S0$.

2. The induction scheme can obtained relative to any situation. We introduce $Can\text{-}future(s0, s)$, where the variable $s0$ can be any initial situation. The axiom schema is

$$\phi(s0) \wedge forall(a, s)(\phi(s) \rightarrow \phi(result(a, s))) \rightarrow \\ \quad forall(s)(Can\text{-}future(s0, s) \rightarrow \phi(s))$$

3. We need to compare a situation with situations that it isn't convenient to consider as having arisen from $S0$. One example comes from counterfactuals. "If another car had come over the hill while you were passing, there would have been a head-on collision." There is no need to consider the hypothetical situation as arising from $S0$.

4. We can use the context mechanism proposed in citeMcC93 to put the big bang interpretation in its place. We have

$$c0 : Ist(Bigbang, Is\text{-}situation(s)) \leftrightarrow Can\text{-}future(s0, s)$$

When the context *Bigbang* is entered, the LPR big bang induction schema holds.

John McCarthy