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

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

Michael Thielscher

A Theory of Dynamic Diagnosis

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

Q2. Marie-Odile Cordier (23.12)

I have three main comments on this paper :

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

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

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

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

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

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

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

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

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

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

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

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

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

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

References:

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

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

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

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

A2. **Michael Thielscher** (9.1)

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

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

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

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

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

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

References:

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

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

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

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

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

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

Q3. **Wolfgang Nejdl** (27.12)

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

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

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

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

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

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

Best regards,
Wolfgang Nejdl

A3. **Michael Thielscher** (9.1)

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

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

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

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

References:

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

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