Vol. 2: *NIL-0*, 31.5.1998
http://www.ida.liu.se/ext/etai/received/actions/002/

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

Michael Thielscher

A Theory of Dynamic Diagnosis

C2-1. **Anonymous Reviewer 2** (3.5)

I have only one complaint which has been already raised by Marie-Odile Cordier during the discussion phase (point 2): the theory assumes that the system may change its status only because of some action. However, faults cannot happen during the execution of actions. Because of this, it is reasonable to "minimize abnormalities" in the initial state. However, are the assumptions reasonable? In your reply to Marie-Odile, you say that you see no fundamental obstacle in extending the theory to "natural events aside from exogenous, volitional actions". Some more words (a section with an example?) about how to do it will be of great help.

C2-2. **Michael Thielscher** (7.5)

I hope that the following remark constitutes a satisfactory reply to your question. This is the basic idea in the paper [1] [5], which I mentioned in my answer to Cordier's question to which you refer: The happening of a (non-exogenous) event is considered a property of a world state and thus is formally treated as a fluent. Hence it can be (directly or indirectly) caused by actions or other events. In the aforementioned paper we have shown that minimization of event occurrences had better be conducted in the initial state only. Now, here is a (sketchy) idea for combining this approach with the present theory of dynamic diagnosis: Suppose we introduce binary fluents $Happens(e, t)$ indicating that event e is expected to happen at time t. If $Happens$ is minimized initially (just like Ab is in the paper under discussion), then the occurrence of natural events, too, may serve as explanation for abnormalities. These explanations would 'compete' with assumptions about initial abnormalities; additional, qualitative knowledge as to the *a priori* likelihood of both events and abnormalities may then be used to tell apart the most plausible diagnoses (along the lines of Definition 10 of the present paper).

[1] Ref: http://www.intellektik.informatik.tu-darmstadt.de/~mit/publications/conferences/ordered-abstracts.html#KR98

Q4. **Anonymous Reviewer 2** (3.5)

I would like if the theory were adjusted to cover this small "defect": According to the definition of "action law" (page 11) the set C and E of fluent literals must be consistent for any sequence of entities in the scope of the action. This simplifies the presentation of the theory. However, it does not cover simple action laws like when $Move(x, f, t)$ transforms $\{Loc(x, f)\}$ into $\{Loc(x, t), notLoc(x, f)\}$, in which the set E of effects is not consistent when you consider $t = f$.

A4. **Michael Thielscher** (7.5)

The following straightforward generalization of Definition 2 could handle your example without affecting any other part of the theory. Rather than stipulating that action laws themselves satisfy the consistency criterion, the latter shall be a requirement of *applicable instances* of these laws. In this way the statement

```
Move(x,f,t) transforms {Loc(x,f),-Loc(x,t)} into {Loc(x,t),-Loc(x,f)}
```

would satisfy the definition of an action law, and actions of the form $Move(x, f, f)$ wouldn't admit a successor state, hence wouldn't be applicable. In the paper I kept the definition of action laws as simple as I possibly could in order not to complicate matters at a point which is not an issue of the paper (namely, the specification of direct effects of actions). Since the original definition does not constitute a technical error, I hope this comment is sufficient an answer.