To be published in:
Electronic News Journal on Reasoning about Actions and Change
Vol. 2: [*pages tbd*], 31.10.1997
http://www.ida.liu.se/ext/etai/received/actions/002/

Michael Thielscher:

A Theory of Dynamic Diagnosis

Summary of the article

The original version of the full article has been published by Linköping University Electronic Press, and is permanently available at
`http://www.ep.liu.se/ea/cis/1997/011/`

Diagnosis in general requires more than just passively observing the behavior of a faulty system. Often it is necessary to actively produce observations by performing actions. Diagnosing then amounts to reasoning about more than a single state of the system to be examined. We propose to capture this dynamic aspect by appealing to Action Theory. A formal system description consists of a *static* and a *dynamic* part. The former introduces the system components and their static relations in form of so-called state constraints, like, for instance,

$$\texttt{active}(\texttt{relay}_1) \equiv \texttt{closed}(\texttt{switch}_1)$$

stating that a particular relay is active if and only if a corresponding switch is closed. The dynamic part of a system description specifies the actions which can be used to manipulate the system's state. These definitions are accompanied by so-called action laws, which focus on the direct effects. State constraints like the above then give rise to additional, indirect effects of actions, which we accommodate according to the theory of causal relationships [Thielscher, 1997b]. E.g., this causal relationship is a consequence of our example state constraint:

$$\texttt{closed}(\texttt{switch}_1) \ \underline{\text{causes}} \ \texttt{active}(\texttt{relay}_1)$$

Informally speaking, it means that whenever $\texttt{closed}(\texttt{switch}_1)$ occurs as direct or indirect effect of an action, then this has the additional, indirect effect that $\texttt{active}(\texttt{relay}_1)$. Generally, causal relationships are successively applied subsequent to the generation of the direct effects of an action until a satisfactory successor state obtains.

In this way, the reactions of a system under healthy condition are modeled as indirect effects, so-called *ramifications*, of actions. Under abnormal circumstances—i.e., if certain aspects or components of the system are faulty—one or more of these ramifications fail to materialize. We introduce an abnormality fluent $\mathtt{ab}$ by which we account for such exceptions to both state constraints and the ramifications they trigger. Thus our example constraint from above, for instance, may read weaker—e.g., to the effect that

$$\neg\mathtt{ab}(\mathtt{resistor}_1)\wedge\neg\mathtt{ab}(\mathtt{relay}_1) \supset [\mathtt{active}(\mathtt{relay}_1) \equiv \mathtt{closed}(\mathtt{switch}_1)]$$

where $\mathtt{ab}(\mathtt{resistor}_1)$ and $\mathtt{ab}(\mathtt{relay}_1)$ represent an abnormal failure of a corresponding resistor and the relay itself, respectively. This weakening transfers to our expectations regarding indirect effects: The aforementioned causal relationship becomes

$$\mathtt{closed}(\mathtt{switch}_1)\ \underline{\text{causes}}\ \mathtt{active}(\mathtt{relay}_1)\ \underline{\text{if}}\ \neg\mathtt{ab}(\mathtt{resistor}_1)\wedge\neg\mathtt{ab}(\mathtt{relay}_1)$$

An important contribution of this paper, now, is a proof that due to the phenomenon of causality straightforward global minimization of abnormality—which is suitable for static diagnosis—is problematic in case of dynamic diagnosis. This raises a challenge much like the one raised by the famous Yale Shooting counter-example in the context of the Frame Problem. Meeting this challenge is inevitable when searching for 'good' diagnoses.

As a solution, we adapt from a recent causality-based solution to the Qualification Problem the key principle of *initial minimization*. All instances of the abnormality fluent are assumed false initially but may be indirectly affected by the execution of actions. In this way, our theory of dynamic diagnosis suitably exploits causal information when generating diagnoses. Our theory moreover respects available knowledge of the *a priori* likelihood of component failures. Since it is often difficult if not impossible to provide precise numerical knowledge of probabilities, we deal with qualitative rather than quantitative information, and we do not rely on complete knowledge. Such possibly incomplete information as to different degrees of abnormality is formally represented by a partial ordering among the instances of the abnormality fluent.

For the entire theory there exists a provably correct axiomatization based on the Fluent Calculus paradigm and which uses Default Logic to accommodate the nonmonotonic aspect of the diagnostic problem.