From: AAAI-97 Proceedings. Copyright © 1997, AAAI (www.aaai.org). All rights reserved.

Qualified Ramifications

Michael Thielscher
FG Intellektik, TH Darmstadt
Alexanderstr. 10, 64283 Darmstadt (Germany)
mit@informatik.th-darmstadt.de

Abstract

We consider the problem of ramifications, i.e., indirect effects of actions, having exceptions. It is argued that straightforward minimization of abnormality is insufficient in this context. Taking a recent causality-based solution to the plain Ramification Problem as starting point, we develop an action theory that is shown to successfully address this amalgamation of Ramification and Qualification Problem.

Introduction

In formal systems for reasoning about actions, the Ramification Problem denotes the problem of handling indirect effects. These effects are not explicitly represented in action specifications but follow from general laws describing dependencies among the components (usually called *fluents*) of a state description. Consider, as an example, an electric circuit where a battery, a light bulb, and a switch are serially connected. Suppose the switch is currently open, then the only direct effect of closing it is that it changes its position. As an additional, indirect effect, however, we expect the light turns on. This conclusion is formally grounded on a general law, often called a *domain constraint*, stating that the light is on if and only if the switch is closed.

A fundamental assumption underlying existing approaches to the Ramification Problem is that a domain constraint is a universal truth. Consequently, it is assumed that the indirect effects it triggers always occur as expected. In any but artificially ideal environments, however, the occurrence of indirect effects often depends on many more conditions than one is usually aware of. The reason for this unawareness is that most conditions are so likely to be satisfied that they are assumed away in case there is no evidence to the contrary. With regard to our example, when we toggle the switch then, contrary to our expectations, the light may actually not turn on—due to, for instance, a broken bulb, a malfunction of the battery, or loose wiring etc. Every one of these problems renders impossible the occurrence of the expected indirect effect. Nonetheless conditions like "the bulb is not broken" should not be part of the underlying domain constraint relating the switch and the bulb. For otherwise one always has to explicitly verify all of these conditions prior to concluding that one can switch on the light. Likewise, the task of diagnosis—an important application area for the results of this paper—requires to *prima facie* assume normal circumstances.

The fact that domain constraints, and hence ramifications, may have exceptions, closely resembles the Qualification Problem, which requires to assume away unlikely disqualifications of actions to the largest possible extent (McCarthy 1977). The similarity suggests a straightforward solution to the problem of disqualified ramifications: Each domain constraint is enhanced by a *normality* condition, which restricts the constraint to all but abnormal circumstances. One could then assume normal circumstances whenever reasonable without strictly ignoring the possibility of exceptional situations. In the following section, however, we show that the straightforward approach of globally minimizing abnormalities in this context is insufficient. It turns out that the execution of actions may *cause* the fact that an indirect effect does not occur as expected. As will be illustrated, failing to distinguish caused from unmotivated abnormalities may lead to unintended conclusions.

The insufficiency of global minimization is related to difficulties, first encountered in (Lifschitz 1987), with globally assuming away abnormalities as an approach to the Qualification Problem. In the spirit of our recent solution to that problem (Thielscher 1996), we propose a method that successfully accounts for the possibility of ramifications being disqualified. Our framework solves the problem of caused vs. unmotivated abnormalities by treating abnormal disqualifications of ramifications as *fluents* which are minimized initially but may later be (indirectly) affected by the execution of actions. The method also includes the proliferation of possible explanations in case a ramification has been—unexpectedly—observed unqualified. The approach builds on our solution to the Ramification Problem (Thielscher 1997), which is based on causal

Copyright © 1997, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

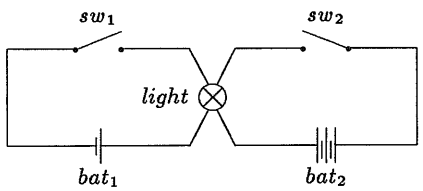

Figure 1: An example electric circuit.

propagation of indirect effects.

Causing Disqualifications

The ability to assume away, by default, exceptional disqualifications of ramifications requires some non-monotonic feature. For one might be forced to withdraw previous conclusions in the light of additional information. This suggests the introduction of 'normality' conditions both to domain constraints and for the computation of indirect effects triggered by these constraints. Then by suitably minimizing abnormality, we achieve the desired behavior, namely, that exceptions are assumed away to the largest reasonable extent. For example, taken as domain constraints the formulas

$$\begin{array}{l} \neg ab_1 \supset [closed(sw_1) \equiv light] \\ broken \supset ab_1 \\ broken \supset \neg light \end{array} \qquad (1)$$

state, respectively, that in any situation the light is on iff switch sw_1 is closed—but only in case there is no abnormality with this regard; that a broken bulb is such an abnormality; and that the broken bulb can never shine. Now, suppose a situation where we know nothing except that sw_1 has just been closed. Then minimizing abnormality allows us to conclude that $closed(sw_1) \equiv light$ holds. Thus any reasonable solution to the Ramification Problem should derive the indirect effect *light*. On the other hand, in a situation where we know that the bulb is broken, ab_1 is derivable, hence cannot be assumed false. Then closing the switch no longer turns on the light.

Generally, assuming away unlikely exceptions of ramifications is, however, not as easy as in this simple example. In particular, just globally minimizing abnormalities can quickly become insufficient to obtain the intended conclusions. To see why, consider the following extension of our introductory example (c.f. Figure 1). The light bulb is now involved in a second subcircuit consisting of another switch, sw_2, and another battery, bat_2. Suppose further that bat_2, without a resistor being involved, is too powerful for our light bulb so that the latter immediately gets broken when closing sw_2—provided, of course, the second battery itself does not malfunction. In conjunction with the formulas in (1), the following two domain constraints formalize this extended domain:

$$\begin{array}{l} \neg ab_2 \supset [closed(sw_2) \supset broken] \\ malfunc(bat_2) \supset ab_2 \end{array} \qquad (2)$$

Consider, now, a situation where we only know that both switches are open (c.f. Figure 1). What would be the predicted outcome of closing sw_2 followed by sw_1? Since nothing hints at bat_2 malfunctioning, we should expect that $closed(sw_2) \supset broken$ be true and, hence, *broken* is an indirect effect of closing sw_2. Consequently, closing sw_1 afterwards should fail to produce light.

But what happens if abnormality is globally minimized in this scenario? Obviously, some abnormality is inevitable. Thus one minimal model is given by $\neg ab_2$ with regard to the first action, and ab_1 with regard to the second. This corresponds to the intended conclusion. However, we can just as well assume the first ramification unqualified (i.e., ab_2), which then would avoid the necessity of assuming a disqualification of the following ramification (i.e., $\neg ab_1$). For if the bulb does not break as a consequence of toggling sw_2, then the light turns on as indirect effect of toggling sw_1 afterwards. This gives us a second, unintended model, where finally the light is on!

The reason for the second minimal model being counter-intuitive is not that we consider broken bulbs more likely than malfunctioning batteries. Rather what decisively distinguishes the two possible abnormalities is that ab_1 but not ab_2 can easily be explained from the perspective of causality. Closing sw_2 along with all of its expected indirect effects *causes* the fact that the relation $closed(sw_1) \equiv light$ no longer holds, whereas an abnormal disqualification, ab_2, of $closed(sw_2) \supset broken$ comes out of the blue in the unintended minimal model. One even tends to not call the former abnormal since being unable to turn on the light after having destroyed the bulb is, after all, what one would normally expect. This situation resembles a problem in the context of the Qualification Problem if the latter is approached without supporting the distinction between caused and unmotivated disqualifications of actions (Lifschitz 1987). The reader may also notice the similarities to the well-known Yale Shooting problem (Hanks and McDermott 1987): A gun that becomes magically unloaded while waiting deserves being called abnormal, whereas causality explains the death of the turkey if being shot at with a loaded gun.

An alternative to global minimization is minimizing chronologically, following the ideas of *chronological ignorance* (Shoham 1988). Putting off abnormalities as long as possible indeed produces the unique intended model for our example scenario. On the other hand, chronological minimization is known to have chronic problems with domains involving non-determinism. Suppose, for example, we learn that while waiting somebody non-deterministically closes sw_2, which would cause an abnormality (viz. the bulb breaks). Putting off abnormalities as long as possible, chronological minimization supports the conclusion that sw_2 never gets closed, which is obviously a conclusion too optimistic, hence unintended.

In the following, we develop a suitable account of disqualified ramifications which respects the possibility of abnormalities being caused. The approach builds on our recently proposed solution to the Ramification Problem (Thielscher 1997). A brief recapitulation of this method is next.

Causal Relationships and Ramifications

Any satisfactory solution to the Ramification Problem requires the successful treatment of two major issues. First, an appropriately weakened version of the general law of persistence needs to be developed which applies only to those parts of the world description that are unaffected by the action's direct *and* indirect effects. Second, while indirect effects derive from domain constraints, not all logical consequences of domain constraints constitute indirect effects (Lifschitz 1990). To meet these challenges, our approach described in (Thielscher 1997) takes the world description obtained through generating the direct effects of an action as a mere intermediate result. Indirect effects are then accommodated by the successive application of *causal relationships* until an overall satisfactory successor state obtains. In the following, we recall the formal definitions underlying this approach. Due to space limitations we restrict ourselves to propositional fluents; for the general case as well as for more details we refer the reader to (Thielscher 1997).

Definition 1 Let $\mathcal{F}$ be a finite set of symbols called *fluents*. A *fluent literal* is either a fluent $f \in \mathcal{F}$ or its negation, denoted by $\neg f$. A set of fluent literals is *inconsistent* iff it contains some $f \in \mathcal{F}$ along with $\neg f$, and is a *state* iff it is maximally consistent. ■

The elements of an underlying set of fluents can be considered atoms for constructing (propositional) formulas to allow for statements about states. Each fluent literal and $\top$ (*tautology*) and $\bot$ (*contradiction*) are *fluent formulas*, and if F and G are fluent formulas then so are $F \wedge G$, $F \vee G$, $F \supset G$, and $F \equiv G$. The notion of fluent formulas being *true* in a state S is based on defining a literal ℓ to be true if and only if $\ell \in S$. Fluent formulas which have to be satisfied in all states that are possible in a domain are also called *domain constraints*.

Definition 2 Let $\mathcal{A}$ be a finite set of symbols called *actions*. An *action law* is a triple $\langle C, a, E\rangle$ where C, called *condition*, and E, called effect, are consistent sets of fluent literals such that $[C] = [E]$,[1] and where $a \in \mathcal{A}$. If S is a state, then an action law $\langle C, a, E\rangle$ is *applicable* in S iff $C \subseteq S$. The *application* yields $(S \setminus C) \cup E$. ■

Notably, the resulting set $(S \setminus C) \cup E$ is a state if so is S, but it may violate the underlying domain constraints.

[1] If S is a set of fluent literals, then by $[S]$ we denote the set of fluents occurring in S. That is, $[C] = [E]$ requires C and E refer to the same fluents.

Definition 3 Let $\mathcal{F}$ be a set of fluents. A *causal relationship* is an expression of the form ε **causes** ϱ **if** Φ where Φ is a fluent formula and ε and ϱ are fluent literals. ■

The intended reading is the following: Under condition Φ, the (previously obtained, direct or indirect) effect ε triggers the indirect effect ϱ.

Causal relationships operate on pairs (S, E), where S denotes the current state and E contains all direct and indirect effects computed so far:

Definition 4 Let (S, E) be a pair consisting of a state S and a set of fluent literals E, then a causal relationship ε **causes** ϱ **if** Φ is *applicable* to (S, E) iff $\Phi \wedge \neg\varrho$ is true in S and $\varepsilon \in E$. Its application yields the pair (S', E') where $S' = (S \setminus \{\neg\varrho\}) \cup \{\varrho\}$ and $E' = (E \setminus \{\neg\varrho\}) \cup \{\varrho\}$. ■

If $\mathcal{R}$ is a set of causal relationships, then by $(S, E) \overset{*}{\rightsquigarrow}_{\mathcal{R}} (S', E')$ we indicate that there are elements in $\mathcal{R}$ whose successive application to (S, E) yields (S', E').

Now, suppose given a set of fluent literals S as the result of having computed the direct effect E of an action via Definition 2. Additional, indirect effects are then accounted for by (non-deterministically) selecting and (serially) applying causal relationships until a state satisfying the domain constraints obtains.

Definition 5 Let $\mathcal{L}$ be a set of action laws, $\mathcal{D}$ a set of domain constraints, and $\mathcal{R}$ a set of causal relationships. Furthermore, let S be a state satisfying $\mathcal{D}$ and $a \in \mathcal{A}$. A state S' is a *successor state* of S and a iff there exists an applicable (wrt. S) action law $\langle C, a, E\rangle \in \mathcal{L}$ such that

1. $((S \setminus C) \cup E, E) \overset{*}{\rightsquigarrow}_{\mathcal{R}} (S', E')$ for some E', and
2. S' satisfies $\mathcal{D}$. ■

Example 1 Let $\mathcal{F} = \{closed(sw_1), closed(sw_2), light, broken\}$. The domain constraints

$$\begin{aligned} closed(sw_1) &\equiv light \\ closed(sw_2) &\supset broken \\ broken &\supset \neg light \end{aligned} \tag{3}$$

state what normally holds in the circuit depicted in Figure 1. All three formulas are true in the state $S = \{\neg closed(sw_1), \neg closed(sw_2), \neg light, \neg broken\}$, e.g. The following causal relationships derive from these domain constraints:[2]

$$\begin{aligned} closed(sw_1) &\ \textbf{causes}\ light &&\textbf{if}\ \top \\ \neg closed(sw_1) &\ \textbf{causes}\ \neg light &&\textbf{if}\ \top \\ closed(sw_2) &\ \textbf{causes}\ broken &&\textbf{if}\ \top \\ broken &\ \textbf{causes}\ \neg light &&\textbf{if}\ \top \end{aligned} \tag{4}$$

[2] See (Thielscher 1997) on how to automatically extract causal relationships from domain constraints given additional domain knowledge as to potential causal influences.

Let $\mathcal{A} = \{toggle(sw_1), toggle(sw_2)\}$, and let $\mathcal{L}$ consists of the four action laws

$$\begin{array}{l} \langle \{\neg closed(sw_1)\},\ toggle(sw_1),\ \{closed(sw_1)\} \rangle \\ \quad \langle \{closed(sw_1)\},\ toggle(sw_1),\ \{\neg closed(sw_1)\} \rangle \\ \langle \{\neg closed(sw_2)\},\ toggle(sw_2),\ \{closed(sw_2)\} \rangle \\ \quad \langle \{closed(sw_2)\},\ toggle(sw_2),\ \{\neg closed(sw_2)\} \rangle \end{array} \quad (5)$$

Then applying action $toggle(sw_2)$ to state S from above, for instance, yields the intermediate state $S' = \{\neg closed(sw_1), closed(sw_2), \neg light, \neg broken\}$—which violates the second domain constraint in (3). Regarding the state-effect pair $(S', \{closed(sw_2)\})$, we can apply the third one of the causal relationships depicted in (4), which results in the pair

$$(\{\neg closed(sw_1), closed(sw_2), \neg light, broken\}, \{closed(sw_2), broken\})$$

The first component satisfies (3), hence constitutes a successor of S and $toggle(sw_2)$. ■

Disqualifications of Ramifications

A fundamental assumption underlying our approach to the Ramification Problem is that domain constraints are universally valid. This assumption is carried over to the causal relationships that derive from a constraint. As a consequence, it is assumed that indirect effects always occur as expected. As argued, however, the situation might not be as ideal. Domain constraints may have exceptions, hence so do causal relationships.

Suppose given a set of domain constraints $\mathcal{D}$. In order to account for exceptions to these formulas, we first introduce, for each $d_i \in \mathcal{D}$, a unique 'abnormality' predicate ab_i. Then constraint d_i is replaced by the weaker formula $\neg ab_i \supset d_i$. This restricts the necessity of d_i being true to states in which $\neg ab_i$ holds—states which are 'normal' with respect to d_i.

The modification of domain constraints transfers to causal relationships. Each ε **causes** ϱ **if** Φ triggered by constraint d_i is replaced by ε **causes** ϱ **if** $\Phi \wedge \neg ab_i$. That is to say, effect ε causes indirect effect ϱ now only under normal circumstances—if there happens to be an exception to the underlying domain constraints, then the corresponding ramification is no longer expected.

Having admitted exceptions to domain constraints, the next step is to define the circumstances under which a particular abnormality occurs. This is accomplished by additional constraints each of which relates some ab_i to the conceivable causes. In order that observed abnormalities can be explained, it is desirable to equate an abnormality with the disjunction of all known potential reasons for its occurrence, e.g.[3]

$$ab_1 \equiv broken \vee malfunc(bat_1) \vee loose_wiring \quad (6)$$

[3] Instead of explicitly providing the "only-if" part, this could be implicitly obtained through circumscribing the predicates ab_i in a given set of domain constraints.

The purpose of introducing conditions of 'normality' with regard to domain constraints is to not strictly exclude the possibility of exceptions to these constraints. Since any exception is considered unlikely, we do however wish to ignore it unless there is evidence to the contrary. Abnormal circumstances should therefore be assumed away to the largest reasonable extent. As shown above, straightforward minimization of abnormality is insufficient to this end because some abnormal disqualifications of ramifications may be expected for reasons of causality. To account for this, we represent any single abnormality as a fluent. As such, abnormalities may be (indirectly) affected by the execution of actions, and otherwise are subject to the general law of persistence. The former allows us to expect an abnormality whenever it has been caused by an action. Notice that formulas such as (6) give rise to indirect effects if taken as domain constraints: Whenever some cause for an abnormality occurs as (direct or indirect) effect, then the abnormality appears through ramification. Conversely, if a cause disappears and no other cause holds, then the abnormality, too, disappears, again through ramification.

In order that abnormal circumstances are assumed away if nothing hints at their presence, the 'abnormality' fluents are considered false by default only in the *initial* state. Persistence then guarantees normal circumstances as long as no actions are performed which affect the truth-value of some fluent ab_i. Formally, we distinguish among all fluents $\mathcal{F}$ those which describe abnormalities, the set of which is denoted $\mathcal{F}_{ab}$. It is required that $ab_i \in \mathcal{F}_{ab}$ for any ab_i, but other fluents may represent abnormal circumstances too, such as, e.g., $broken$, $malfunc(bat_1)$, or $loose_wiring$. When searching for models of a scenario description, those are preferred that declare false initially as many fluents $f_{ab} \in \mathcal{F}_{ab}$ as possible according to the observations that constitute the scenario. The formal definition of model preference can be directly adopted from (Thielscher 1996):

Definition 6 An *interpretation* is a partial function Res mapping finite action sequences to states[4] such that for each $k \geq 0$ and each action sequence $a^* = [a_1, \ldots, a_k, a_{k+1}]$,

1. $Res([])$ is defined and satisfies the domain constraints.
2. $Res(a^*)$ is defined iff so is $Res([a_1, \ldots, a_k])$ and there is a successor of $Res([a_1, \ldots, a_k])$ and a_{k+1}.
3. If $Res(a^*)$ is defined, then it is a successor of $Res([a_1, \ldots, a_k])$ and a_{k+1}.

An *observation* is an expression F **after** $[a_1, \ldots, a_n]$ where F is a fluent formula and $a_1, \ldots, a_n$ are actions ($n \geq 0$). The observation *holds* in an interpretation Res iff $Res([a_1, \ldots, a_n])$ is defined and F is true in that state.

[4] We consider a branching time structure.

A *model* for a set of observation $\mathcal{O}$ is an interpretation in which all observations hold. A model *Res* is *preferred* iff there is no model *Res'* such that $Res'([]) \cap \mathcal{F}_{ab} \subsetneq Res([]) \cap \mathcal{F}_{ab}$. ■

In the remainder of this section we illustrate how our framework successfully addresses basic issues when dealing with disqualifications of ramifications. Although no strictly formal claim is made, the following discussion is meant to convince the reader that our theory is suitable as regards a variety of fundamental aspects in this context.

Assuming qualification by default. Let us extend the set of fluents $\mathcal{F}$ used in Example 1 by ab_1, ab_2, $malfunc(bat_1)$, $malfunc(bat_2)$, and *loose_wiring*. All of these plus fluent *broken* shall belong to $\mathcal{F}_{ab}$. We define the following domain constraints (see Figure 1):

$$\neg ab_1 \supset [closed(sw_1) \equiv light] \quad (7)$$
$$ab_1 \equiv broken \vee malfunc(bat_1) \vee loose_wiring \quad (8)$$
$$\neg ab_2 \supset [closed(sw_2) \supset broken] \quad (9)$$
$$ab_2 \equiv malfunc(bat_2) \vee loose_wiring$$
$$\neg ab_3 \supset [broken \supset \neg light]$$
$$ab_3 \equiv \bot$$

The very last formula states that a broken bulb shining would be inexplicable. Due to space restrictions we only mention four out of all the causal relationships determined by these constraints:

$$closed(sw_1) \texttt{ causes } light \texttt{ if } \neg ab_1 \quad (10)$$
$$broken \texttt{ causes } ab_1 \texttt{ if } \top \quad (11)$$
$$\neg broken \texttt{ causes } \neg ab_1 \texttt{ if } \neg malfunc(bat_1) \wedge \neg loose_wiring \quad (12)$$
$$closed(sw_2) \texttt{ causes } broken \texttt{ if } \neg ab_2 \quad (13)$$

deriving from (7), (8), (8), and (9), respectively.

Suppose given the observation

$$\neg closed(sw_1) \wedge \neg closed(sw_2) \texttt{ after } [] \quad (14)$$

It is consistent with this observation to assume false initially each $f_{ab} \in \mathcal{F}_{ab}$. Any preferred model *Res* therefore satisfies $\neg ab_2 \in Res([])$. Consequently, $broken \in Res([toggle(sw_2)])$ given action laws (5) and causal relationship (13). In words, in any preferred model the bulb is broken after toggling sw_2 —which is the intended conclusion: Abnormal disqualification of constraint (9) and of the ramification it triggers, c.f. (13), is assumed away by default.

Causing disqualifications. We just saw that any preferred model *Res* of observation (14) satisfies $broken \in Res([toggle(sw_2)])$. According to causal relationship (11), *broken* becoming true determines another ramification, namely, ab_1 becoming true in $Res([toggle(sw_2)])$. That is, the ramification becomes disqualified which normally turns on the light if sw_1 gets closed. Thus any preferred model *Res* of (14) satisfies $\neg light \in Res([toggle(sw_2), toggle(sw_1)])$, which is the intended conclusion as argued at the beginning: The disqualification of domain constraint (7) has been obtained as a side-effect of, hence has been caused by, performing $toggle(sw_2)$.

Explaining disqualifications. Suppose given

$$\neg broken \wedge \neg loose_wiring \texttt{ after } [] \quad (15)$$
$$\neg light \texttt{ after } [toggle(sw_1)] \quad (16)$$

in addition to observation (14). Any model *Res* must satisfy $ab_1 \in Res([])$ to account for (16). Observation (15) implies $\neg broken, \neg loose_wiring \in Res([])$. Definition 6 requires that $Res([])$ satisfy the underlying domain constraints, in particular (8); thus, $malfunc(bat_1) \in Res([])$. That is to say, each preferred model determines a malfunction of bat_1 as explanation for being unable to switch on the light via closing sw_1.

Minimizing explanations. Suppose we learn that $malfunc(bat_1)$ `after` [], then $ab_1 \in Res([])$ for any model *Res* according to (8). Nonetheless it is consistent to assume both *broken* and *loose_wiring* be false initially. These, too, being abnormality fluents, we have $\neg broken, \neg loose_wiring \in Res([])$ in any preferred model. Thus, although we know there must be an abnormality with the sub-circuit involving bat_1, we still conclude, by default, that bulb and wiring are ok.

Revoking qualification. Suppose we introduce the action of replacing the broken bulb, specified by the action law $\langle\{broken\}, replace, \{\neg broken\}\rangle$. As we have seen, any preferred model *Res* of observation (14) satisfies $broken, ab_1 \in Res([toggle(sw_2)])$. According to (5), again toggling sw_2 opens this switch while nothing else changes, that is, *broken* and ab_1 remain true in $Res([toggle(sw_2), toggle(sw_2)])$. Consequently, the *replace* action is applicable to this state with the effect that *broken* is false in the next state. Then causal relationship (12) determines the ramification $\neg ab_1$ (notice that both $\neg malfunc(bat_1)$ and $\neg loose_wiring$ hold initially in preferred models and persist throughout the entire process). This implies that the light could again be switched on in the state $Res([toggle(sw_2), toggle(sw_2), replace])$. This in turn shows how a qualification gets revoked as soon as its cause (here: the broken bulb) disappears.

Non-deterministic actions. Recall the scenario discussed earlier where switch sw_2 may or may not get closed while waiting. Action *wait* being non-deterministic, it is specified by two action laws whose conditions are not mutually exclusive, viz.

$$\begin{array}{c} \langle\{\}, wait, \{\}\rangle \\ \langle\{\neg closed(sw_2)\}, wait, \{closed(sw_2)\}\rangle \end{array} \quad (17)$$

Suppose again given observation (14). Since the initial values of the only non-abnormality fluents in

our domain are specified via this observation, there is a unique preferred initial state $Res([])$, namely, $\{\neg closed(sw_1), \neg closed(sw_2)\} \cup \{\neg f_{ab} : f_{ab} \in \mathcal{F}_{ab}\}$. Among others, this state entails the (default) assumption $\neg ab_1$, that is, the light could be switched on. However, performing a *wait* action in $Res([])$ determines two possible successor states according to (17): one that is identical to $Res([])$, and one where sw_2 gets closed and, hence, the bulb breaks and ab_1 becomes true (c.f. (13) and (11)). Thus there are two different preferred models Res_1 and Res_2 such that $\neg ab_1 \in Res_1([wait])$ whereas $ab_1 \in Res_2([wait])$. Due to the latter, it cannot be concluded that the light can be switched on after waiting, which is the intended conclusion. This shows that our framework, in contrast to chronological minimization, treats non-deterministic information appropriately, namely, the cautious way.

Discussion

We have argued that both domain constraints and the ramifications they trigger may have exceptions, which, however, need to be assumed away by default. We have illustrated that simple global minimization of abnormalities is insufficient to this end because it fails to distinguish caused from unmotivated disqualifications. We then have developed an action theory where abnormal disqualifications of ramifications are taken as fluents which are assumed false initially but may be indirectly affected by the execution of actions. It has been illustrated how this allows to assume away exceptional circumstances to a reasonable extent—including a proper distinction of caused vs. unmotivated abnormalities. In addition, our framework supports the proliferation of explanations for observed disqualifications of ramifications. These explanations, too, are minimized, and if a cause for a disqualification disappears then qualification gets revoked. Finally, our approach to ramifications with exceptions has been shown not to interfere with handling non-deterministic information.

The work reported here essentially relies on our causality-based solution to the Ramification Problem described in (Thielscher 1997). Arguments in favor of this solution, linked with a through comparison to other approaches, can be found in that paper. The problem of ramifications having exceptions has received little attention in literature up to now, probably because satisfactory solutions to the Ramification Problem itself have not emerged until very recently. To the best of the author's knowledge, the only existing papers dealing with disqualifications of ramifications are (Baral and Lobo 1996; Zhang 1996). In both of them expressions resembling causal relationships are allowed to be defeasible. Neither of the approaches, however, goes beyond defining a notion of successor state based on minimizing abnormality. Therefore, if applied as they stand the two approaches, by producing the unintended model, immediately get caught in the causality trap illustrated with our key example depicted in Figure 1.

The action theory presented in this paper was inspired by a solution to the problem of abnormal disqualifications of actions proposed in (Thielscher 1996). The latter also describes a provably correct Fluent Calculus (Hölldobler and Schneeberger 1990) realization of that theory, which uses default rules to encode the assumptions of 'normality'. This Fluent Calculus encoding can straightforwardly be adopted to the theory proposed in the present paper. (Thielscher 1996) also describes a way to deal with so-called *miraculous* abnormalities. This idea, too, can be adapted to the problem of disqualified ramifications. A miracle occurs whenever an abnormality cannot be explained from the (necessarily restricted) knowledge provided by domain constraints such as (6). See (Thielscher 1996) for a formal discussion on this topic.

References

Baral, C., and Lobo, J. 1996. Formalizing defeasible causality in action theories. (Unpublished manuscript).

Hanks, S., and McDermott, D. 1987. Nonmonotonic logic and temporal projection. *Artificial Intelligence* 33(3):379–412.

Hölldobler, S., and Schneeberger, J. 1990. A new deductive approach to planning. *New Generation Computing* 8:225–244.

Lifschitz, V. 1987. Formal theories of action (preliminary report). In McDermott, J., ed., *Proc. of IJCAI*, 966–972.

Lifschitz, V. 1990. Frames in the space of situations. *Artificial Intelligence* 46:365–376.

McCarthy, J. 1977. Epistemological problems of artificial intelligence. In *Proc. of IJCAI*, 1038–1044.

Shoham, Y. 1988. Chronological ignorance: experiments in nonmonotonic temporal reasoning. *Artificial Intelligence* 36:279–331.

Thielscher, M. 1996. Causality and the qualification problem. In Aiello, L. C.; Doyle, J.; and Shapiro, S., eds., *Proc. of the Int.'l Conf. on Principles of Knowledge Representation and Reasoning*, 51–62, Cambridge, MA. Morgan Kaufmann.

Thielscher, M. 1997. Ramification and causality. *Artificial Intelligence* 89(1–2):317–364.

Zhang, Y. 1996. Compiling causality into action theories. In *Proc. of the Symposium on Logical Formalizations of Commonsense Reasoning*, 263–270. Stanford, CA.