Filter Preferential Entailment for the Logic of Action in Almost Continuous Worlds

Erik Sandewall

Abstract

IDA Technical Report 1989
LiTH-IDA-R-89-37
ISSN-0281-4250

Department of Computer
and Information Science,
Linköping University,
S-581 83 Linkping, Sweden

Mechanical systems, of the kinds which are of interest for qualitative reasoning, are characterized by a set of real-valued parameters, each of which is a piecewise continuous function of real-valued time. A temporal logic is introduced which allows the description of parameters, both in their continuous intervals and around their breakpoints, and which also allows the description of actions being performed in sequence or in parallel. If axioms are given which characterize physical laws, conditions and effects of actions, and observations or goals at specific points in time, one wishes to identify sets of actions ("plans") which account for the observations or obtain the goals. The paper proposes preference criteria which should determine the model set for such axioms. It is shown that conventional preferential entailment is not sufficient. A modified condition, *filter preferential entailment* is defined where preference conditions and axiom satisfaction conditions are interleaved.

This research was supported by the Swedish Board of Technical Development and the IT4 research program.

Appeared in Proceedings of the 1989 International Joint Conference on Artificial Intellligence.

1 Topic

Our ultimate research goal is to find a coherent theory for temporal reasoning, knowledge based planning, and qualitative reasoning. In that context the present paper addresses the following problem. Assume that one has obtained:

1. a description of a mechanical or other physical system
2. axioms characterizing physical laws which hold in that system
3. axioms characterizing conditions and effects of actions which can be performed by an agent in the world
4. axioms characterizing the observed or desired state of the world at certain point(s) in time

all expressed as logic formulas (wff) in a suitable logic. By what logical criteria can one then derive formulas characterizing a set of actions which together explain the observed state of the world resp. which obtain the desired state of the world?

This general problem includes both "planning" in the A.I. sense (knowledge based planning) and "temporal explanation", the difference being that in knowledge based planning the axioms in the last group express the goals, whereas in temporal explanation they express the given observations. For simplicity we shall refer to the last group of axioms as *observations* in the sequel, regardless of whether the application is planning or explanation.

2 Combining logic and differential equations for describing the physical system

In a previous paper [San89a] we have described an approach to integrating non-monotonic logic and differential equations for characterizing physical systems. The solution described there was however limited to the case in which there are no agents or actions in the world. The present paper extends the same approach to the case where actions may occur.

The key ideas in the approach of the previous paper are as follows. Object systems are assumed to be characterized by a number of parameters, or fluents, which have a real-number value at each point in time. Time is also measured as real numbers, not only as discrete time-points. All parameters are assumed to be *piecewise* continuous and differentiable, with all their derivatives. For example a scenario with a bouncing ball will be characterized by the position, velocity, acceleration etc. of the ball as functions of time. These functions are continuous except (some of them) for those moments when the ball bounces, where they are taken to be discontinuous.

The properties of, and relationships between the parameters can then be described by logic formulas, in a language which has been suitably extended to allow time derivatives, and left and right limit values of parameters at the breakpoints (i.e. the timepoints where discontinuities may occur). The following are examples of such formulas:

$$[\tau_5]Temp(b_4) > 0$$

saying that at time τ_5, the temperature of the object b_4 is greater than zero degrees Celsius,

$$\Box Temp(b_4) \leq 100$$

saying that the temperature of the object is always less than or equal to 100 degrees C,

$$\Box - 0.1 < \partial Temp(b_4) < 0.05$$

saying that the rate of change of the temperature of the objects is always between 0.1 degrees C per second decreasing, and 0.05 degrees per second increasing,

$$\Box bouncepoint(x_b, y_b) \Rightarrow \partial y_b^l = -\partial y_b^r$$

saying that whenever the current coordinates (x_b, y_b) of the object at hand, is a point where the object bounces, then the left limit value of the vertical velocity has the same magnitude, but the opposite sign, as the right limit value of the vertical velocity. In this way the logic can characterize the behavior of the parameters around their discontinuities. For additional details and for a more systematic exposition please refer to [San89a].

Interpretations for such formulas must contain the following parts:

1. a specification of a set of breakpoints i.e. those timepoints at which discontinuities may occur

2. a mapping from time points $\times$ parameter symbols to parameter values (where the set of parameter symbols is closed under prefixing of ∂)

3. appropriate mappings from constant symbols, such as τ_5 and b_4 above, to corresponding numbers, objects, etc.

Also certain general conditions must be placed on the interpretations, in particular that for each particular parameter symbol, the mapping in the second item must be continuous, as a function of time, at all time points which are not break points. (In the break points it is allowed but not required to be continuous).

3 Actions

For the purpose of the present paper, the formula language and the interpretations are extended with *actions.* Formulas are allowed such as

$$Do[t_5, Keepwarm(b_4), t_8]$$

saying that an action of the *action type* $Keepwarm(b_4)$ is performed from time t_5 to time t_8. Correspondingly interpretations are amended with one more component, namely an *action set* which should be a set of triples; each triple consisting of the time point when the action starts, the descriptor for the action, and the time point when the action ends. Action descriptors may be atomic, or may have a structure for example $\langle$ *"Keepwarm"*, $b_4^\star\rangle$ i.e. the tuple consisting of the symbol *"Keepwarm"* and the object which is being kept warm.

Additional formal details are not necessary here. The important point is that each interpretation characterizes a possible "history" of the object world in two parallel ways: partly as a set of snapshots i.e. assignments of values to parameters at each point in time, and partly as a set of actions each of which has a starting time, an ending time, and a descriptor saying what the action is. Logic formulas refer to such interpretations, and the truth value of a formula in each interpretation is defined. It should be clear how one can, for example, write axioms which specify the effects of actions i.e. the values of parameters at the time when an action ends.

The topic of the present paper is now re-phrased formally as follows. Let Γ_o be a set of formulas representing observations (or goals, in a planning problem); let Γ be a set of formulas representing all the other given information; we look for a formula ϕ which characterizes the set of actions, or plan, that accounts for or obtains the observations. Often ϕ would be a disjunction of expressions each of which characterizes an alternative explanation or plan.

4 Semantic entailment

For the purpose of the present work it was necessary to introduce a non-standard notion of semantic entailment. Let us first relate it to the traditional notions. In general a definition of whether $\Gamma \models \phi$ should consist of two parts:

1. *model set criterium:* determine the model set for Γ
2. *formula criterium:* identify which formulae ϕ one wants to conclude from the model set

where in classical logic the model set is of course simply $Mod(\Gamma)$, the set of all interpretations in which all members of Γ are true, and the formula criterium is to choose a formula which is true in all members of the model set. In *non-monotonic* logic the model set for Γ is chosen differently, using model minimization, but the formula criterium remains unchanged. For reasoning about actions, however, both criteria have to be reconsidered.

In preferential entailment, as defined by [Sho88], one assumes the existence of a preference relation $\ll$ which is a partial order on interpretations, and defines the model set for Γ in the simplest case as

$$Min(\ll, Mod(\Gamma))$$

i.e. as the set of $\ll$-minimal members of $Mod(\Gamma)$. However since there may be infinite chains of successively more preferred models, whose limit does not exist or is not a model for Γ, an alternative is to define the model set of Γ as a set of paths of interpretations,

$$Paths(\ll, Mod(\Gamma))$$

where $Paths(\ll, S)$ is the set of all maximal subsets of S within which $\ll$ is a total order. The formula criterium is then that in each preference path of the model set, there must be some member J such that ϕ is true in all elements $\ll J$ in that path.

Our previous paper proposes a definition for the preference relation $\ll$ for the logic that was outlined in the previous section. The criterium, *chronological minimization of discontinuities* (CMD) is basically that if J and J' are interpretations, then $J \ll J'$ iff there is some time-point t such that J and J' assign the same value to all parameter for all times $< t$, *and* the set of parameters in J which have a discontinuity in t is a true subset of the set of parameters in J' that have a discontinuity in t. We showed that preferential entailment using CMD, and with a reasonable set of axioms, obtains the intended set of models for a simple but prototypical example. From the example and from general considerations we concluded that CMD is a plausible choice of semantic entailment condition for piecewise continuous systems.

In the following sections we address the question whether and how the model set criterium has to be revised when observations and actions also occur as axioms and in the interpretations. First however a brief remark on the formula criterium, how to go from model set to proposed conclusions. It has sometimes been debated whether one should let "goals entail plans" [Kau88] or vice versa, i.e. schematically whether one should require

$$\Gamma \cup \Gamma_o \models \phi$$

or

$$\Gamma \cup \{\phi\} \models \Gamma_o$$

In a separate paper [San89b] we argue that neither of these alternatives constrains ϕ correctly, and that both of them should be used together at least approximatively. The requirement should therefore be that the model set for Γ, Γ_o be approximatively equal to the model set for Γ, ϕ. Since the issue is somewhat complex, it merits a paper of its own. We therefore omit additional detail from the present paper, and focus on the first question of identifying the model set for Γ, Γ_o.

Of course the criterium for a proposed definition of semantic entailment is not that it in itself obtains the correct model set, since what models are obtained depends also on the axioms. For the conventional "frame problem", for example, it is perfectly possible to obtain the correct model set with standard, monotonic entailment, but the problem is that it may be very cumbersome to write out the axioms. The criterium for a definition of entailment is therefore whether it makes it easy to write axiomatizations which obtain the right model sets. This is the claim that is tentatively made for chronological minimization of discontinuities.

5 Model set criteria for axioms with observations

The extension of the logic to allowing actions, is closely tied to the use of observations. Without actions there are good reasons to study the consequences of the general axioms combined with observations at an initial point in time, or at no time at all. (The latter case corresponds essentially to the notion of "envisionment" in qualitative reasoning, [dKB85]). With actions, observations at two or more points in time are needed, for example for specifying the initial condition and the goal for the required action-plan.

We therefore first consider the case where there are observations but no actions. The set Γ_o, by the notation at the beginning of the previous section, is assumed to be non-empty, while the action set is empty. Consider the following two candidate model sets:

$$M_1 = Min(\ll, Mod(\Gamma \cup \Gamma_o))$$

$$M_2 = Filter(\Gamma_o, Min(\ll, Mod(\Gamma)))$$

where $Filter(\Gamma_o, S)$ is a subset of S consisting of those members of S which are also models for Γ_o. The definition of M_2 can therefore be written equivalently as

$$M_2 = Mod(\Gamma_o) \cap Min(\ll, Mod(\Gamma))$$

These definitions have been made in terms of minimal models; corresponding definitions in terms of preference paths are easily constructed.

The definition for M_1 is the most straightforward one from the point of view of the inference system: one just combines all the available knowledge and applies preferential entailment as usual. The definition for M_2 however has an intuitive appeal: since

$$Min(\ll, Mod(\Gamma))$$

is the set of all possible developments in the world *regardless of* any observations, it would make sense to take that whole set and "filter" it with the given observations.

It is easily seen that $M_2 \subseteq M_1$, since every member of M_2 satisfies all conditions for being a member of M_1. The opposite is however not always the case, as the following example shows. Consider a one-dimensional object system where an object (with zero size) is known to have coordinate 0 at time 0, and coordinate 2 at time 10. It is also known that the object's coordinate is continuous at all times (including breakpoints), and that the velocity of the object is always ≤ 10. Finally it is known that the object's acceleration is zero at all times. This is all one knows.

In the absence of any reason for a discontinuity in the object's velocity, one would expect to have one single member of the model set, namely where the velocity of the object is 0.2 at all times. We shall refer to it as the standard model. Consider however also a model where the velocity of the object is 0.1 from time 0 to some time-point shortly before time 10, where the velocity changes discontinously to a larger value which allows the object to be at coordinate 2 in time. Such an interpretation satisfies both the observation axiom (coordinate 2 at time 10) and the other axioms, and is therefore a member of $Mod(\Gamma \cup \Gamma_o)$[1].

Since the initial velocity in such a non-standard model is different from a standard model, the chronological minimization of discontinuities will not prefer standard over non-standard models. On the other hand for a given initial velocity (e.g. 0.1), CMD will prefer those models where the discontinuity occurs as late as possible, or in other words the model where the object uses the maximum speed after the discontinuity. Thus for every choice of initial velocity there is exactly one preferred model in M_1.

In M_2 on the other hand, only the standard model is obtained. This is because $Min(\ll, Mod(\Gamma))$ will not contain any model with a discontinuity: for every model with a discontinuity there is a corresponding model without the discontinuity which is preferred according to $\ll$, and which also satisfies all the axioms in Γ. The set $Min(\ll, Mod(\Gamma))$ will contain one member for every possible value of the object's velocity, but only one of them will remain after filtering with the observation axiom.

Based on this discussion and example, we suggest that the definition of M_2 is the one which should be used for identifying model sets in piecewise continuous worlds, and as the first step in the definition of semantic entailment there. The term *filter preferential entailment* is proposed for semantic entailment using the definition of M_2 as its model set criterium.

Parenthetically, it is interesting to note that if the maximal speed condition is dropped (or is changed from a $\leq$ condition to a $<$ condition), *and* the model minimization is performed in terms of minimal models rather than preference

[1]It may seem strange that the velocity may have a discontinuity while at the same time the acceleration is constantly zero. It is OK, however; see [San89a] for the explanation.

paths, then M_1 as well as M_2 contains only the standard model. This is because there is an infinite progression of non-standard models where the break point occurs later and later, and the velocity after the break point is greater and greater. In the interpretation at the limit of that progression, the object's coordinate has a discontinuity at time 10, which means that the limit interpretation is not a model of the given axioms in Γ. Therefore non-standard models are de-selected. However we can not see this as a reason for reconsidering the M_1 definition - it would be too *ad hoc* as a method for de-selecting unintended models.

6 Model set criteria for interpretations with action sets

We proceed now to the case where interpretations contain not only the "snapshots" i.e. the state of the world or the parameter values at each point in time, but also a set of actions each specified by starting time, ending time, and action type, as described in section 3. This means in particular that $Mod(\Gamma)$ is going to contain interpretations containing no, one, or several actions in their action sets.

An additional preference condition must be involved when action sets occur in interpretations, namely a preference relation between interpretation which is due to preference between their respective action sets. Suppose for given Γ and given Γ_o we have a model set with in particular two interpretations J and J', where J contains an action set which is in fact necessary for explaining or achieving the observations, and J' is essentially the same as J except its action set also contains a redundant action which does not influence the value obtained for the observations. In such a case one would obviously prefer J over J', both as an explanation of the observations and as a plan for the agent's own actions.

There may however also be other preferences which are not so obvious, all the way to the criteria based on a cost function on plans. The question is then, how shall the model set be modified to account for our preference between action sets, or plans?

The choice in CMD to apply observation enforcement "after" the (chronological) minimization condition, generalizes naturally to plan minimization. We therefore propose the following model set criterium:

$$Min(\prec, Filter(\Gamma_o, Min(\ll, Mod(\Gamma))))$$

where $\ll$ only compares interpretations with *the same action set*, and $\prec$ compares interpretations according to the preference of their respective action sets, so that at least if the set of actions in the interpretation J is $\subset$ the set of actions in J' then $J \prec J'$. Other action-set preferences may of course also be added.

The first preference relation $\ll$ was defined in the previous paper and described above. However it now has to be slightly revised, since one should not do chronological minimization of those discontinuities which are caused by actions. Only discontinuities which occur "spontaneously" as consequences of the laws of nature, for example when a ball falls over an edge, or heated water arrives to the boiling point, should be chronologically minimized. Otherwise the preference according to $\ll$ will prefer all actions to take place as late as possible!

Formally this modification can be achieved by introducing two *masking relations* X_l and X_r on

$$timepoints \times properties$$

where $X_l(t,u)$ waives the CMD preference for interpretations where $u^l(t) = u(t)$, and similarly for X_r. (Notice if $u^l(t) = u(t) = u^r(t)$ then u is continuous at time t). The relations X_l and X_r are themselves minimized by the relation $\ll$. Also modes (propositional fluents) are dealt with in the same way. The resulting formal semantics is as follows.

Definition. An interpretation is a tuple

$$\langle H, M, U, S, X_l, X_r, P, R, Q, W \rangle$$

where H is a set of action type symbols, M is a set of mode symbols (for truth valued parameters), U is a set of parameter symbols closed under prefixing by ∂, $S \subset \mathbf{R}$ is a "sparse" set of time-points namely the set of breakpoints, $X_l \subset \mathbf{R} \times (M \cup U)$ and X_r with the same type, are the masks on where modes and parameters are "allowed" to be discontinuous, $P \subset S \times H \times S$ is the "plan" i.e. the set of actions, R is a mapping

$$(\mathbf{R} \times M) \rightarrow \{T, F\}$$

which gives the (truth-)value of a mode at each point in time, Q is a mapping

$$(\mathbf{R} \times U) \rightarrow \mathbf{R}$$

which similarly gives the (real-)value of a parameter at each point in time, and W is a mapping from temporal constant symbols (such as τ_1) to the domain $\mathbf{R}$ of real numbers understood as time-points.

Interpretations are subject to the *continuity requirement* that for every $t \notin S$, $R(t,m)$ and $Q(t,u)$ shall be continuous as functions of t. For t not in S it is also required

$$d/dt\ Q(t,u) = Q(t, \partial u)$$

Logic formulas which can be evaluated in such interpretations are defined as desired.

Definition. A parameter u is *essentially continuous* at time t, and we write $ec(t,u)$, iff

$$(X_l(t,u) \vee Q^l(t,u) = Q(t,u)) \wedge$$
$$(X_r(t,u) \vee Q^r(t,u) = Q(t,u))$$

where the index ^{l} and ^{r} on Q represent the left and right limit values, as used in section 2 above.

Essential continuity for modes is defined similarly.

Definition. For every interpretation J and timepoint $t \in \mathbf{R}$ we define $breakset(J, t)$ as

$$\{u \mid \neg ec(t, u)\} \cup \{m \mid \neg ec(t, m)\}$$

Definition. If J and J' are interpretations, whose elements are $H, M, \ldots$ and $H', M', \ldots$ respectively, then $J \ll J'$ iff $P = P'$, $X_l \subseteq X'_l$, $X_r \subseteq X'_r$, and either of the following applies: either $X_l \subset X'_l$, or $X_r \subset X'_r$, or there is some time-point t_0 such that

1. for all $t < t_0$, all $m \in M$, and all $u \in U$, $R(t, m) = R'(t, m)$ and $Q(t, u) = Q'(t, u)$
2. $breakset(J, t) \subset breakset(J', t)$

Definition. If J and J' are interpretations whose elements are like in the previous definition, then $J \prec J'$ iff $P \subset P'$. (Additional conditions may be added).

Definition. The *filter preferential model set* of the general axioms Γ and the observation axioms Γ_o, using the preference relations $\ll$ and $\prec$, is

$$Min(\prec, Filter(\Gamma_o, Min(\ll, Mod(\Gamma))))$$

The hypothesis is that for prototypical applications one can conveniently write axiomatizations for which the filter preferential model set, with CMD as the first preference relation, obtains the intended model set.

7 Example: the covered shaft

The following example illustrates the proposed entailment criterium and the issues involved in choosing it. Two small objects called h and k are bouncing indefinitely back and forth at constant speed in a fixed horizontal range. In figure 1 the x dimension is "horizontal", the y dimension is "vertical", and the range is between the points (a, e) and (d, e) in the (x, y) plane. Also the part of the range which is between the x coordinates b and c (where $a < b < c < d$) is the lid of a shaft which extends indefinitely downwards. At each point in time the lid is either *on* or *off*. If the lid is off at a time when either of the objects is between (x coordinate) b and c, then the object starts falling into the shaft, with constant vertical acceleration -10 during the fall. The horizontal coordinate of the object during the fall is considered irrelevant, so we omit axioms that constrain it.

In the absence of any action the lid is on. There is an action *liftlid* which always takes two seconds, and which has the effect that the lid starts being *off* at some time between 0 and 0.5, and starts being *on* again at some time between 1.5 and 2, counted from the beginning of the *liftlid* action. Also of course it is not possible to do two *liftlid* actions during overlapping time intervals.

The scenario starts at time 0, with the two objects at opposite ends of the range, and moving towards each other each with velocity 1. There is no interference when the objects meet along the way. (They may e.g. be thought of as having different but constant coordinates in a third dimension z).

The scenario is illustrated in figure 1, and can be used for a number of deduction exercises, such as:

* what are the possible interpretations in the scenario world, each interpretation being a possible course of events there from time 0 and onwards? (envisioning)

* given additional observations also at some time(s) > 0, what remaining courses of events are compatible with those observations but without assuming any actions?

* given additional observations like in the previous case, but in such a way that some action(s) must be assumed, what formulas characterizing the set of actions are entailed?

For example, the formula saying that there is a time > 0 at which the y coordinate of k is $< e$, can only be true in interpretations where there is an *openlid* action which has been chosen in such a way as to let through k.

The following is a partial set of axioms characterizing the scenario. Let (x_h, y_h) be the coordinates of h as a function of time, let (x_k, y_k) be the same for k, and let j range over the two objects h and k. Most of the axioms are similar to the axioms which were defined and discussed in [San89a].

1. $\Box a < b < c < d$
2. a, b, c, d, e are constants over time
3. $[0](x_h = a \wedge y_h = e \wedge \partial x_h = 1)$
4. $[0](x_k = d \wedge y_k = e \wedge \partial x_k = -1)$
5. $\Box(x_j = a \vee x_j = d) \wedge y_j = e \Rightarrow (\partial x_j^l = -\partial x_j^r)$
6. $\Box \partial^2 x_j = 0$
7. $\Box y_j = e \wedge \neg(b < x_j < c \wedge \mathit{lidoff}) \Rightarrow \partial y_j = 0$
8. $\Box y_j \neq e \vee (b < x_j < c \wedge \mathit{lidoff}) \Rightarrow \partial y_j = -10$
9. $Do[t_1, \mathit{openlid}, t_2] \Rightarrow t_2 = t_1 + 2 \wedge$

. $\exists t_3 \exists t_4 [t_1 < t_3 \leq t_1 + 0.5 \wedge$

. $t_1 + 1.5 \leq t_4 \leq t_2 \wedge$

. $[t_1, t_3\rceil \neg \mathit{lidoff} \wedge$

. $[t_3, t_4\rceil\, \mathit{lidoff} \wedge$

. $[t_4, t_2] \neg \mathit{lidoff} \wedge$

. $\lceil t_1, t_2] X_l(\mathit{lidoff}) \wedge [t_1, t_2\rceil X_r(\mathit{lidoff})]$

10. $\Box C(X_j)$
11. $\Box a < x_j < d \Rightarrow C(\partial x_j)$

The following notation is used for intervals: $[t, v]$ represents the closed interval from t to v inclusive, and $\lceil t, v \rceil$ represents the open interval not containing t or v. Mixed intervals (left open, right closed and vice versa) are written in the obvious way, and $[t]$ represents the interval whose only member t is.

Let us now add the observation axiom saying that object k has fallen into the shaft and object h has not, nor is it just going to:

12. $[\tau_1](y_h = e \wedge y_k < e \wedge \neg \mathit{lidoff})$

The third conjunct is needed in order to exclude interpretations where the lid opens for both objects, but k falls in first. Notice that the t_i are temporal variables and τ_i are temporal constant symbols. Furthermore we add axioms for the exact range and lid size in the scenario:

13. $\Box a = 1.5 \wedge b = 2 \wedge c = 3 \wedge d = 5$

This set of axioms is still somewhat incomplete, but not in any way that matters for the continued argument.

8 Discussion of example

With the sizes stated in axiom 13 there are two primary good "strategies" for achieving the "goal" expressed in axiom 12. One strategy is to perform the *openlid* action so that it catches the object k as it is travelling right, and with a timing that makes sure that one does not trap h as well. This is satisfied if the action is initiated between 0 and 0.5 seconds before k bounces at coordinate a. The other strategy is to catch k as it is travelling left, and around the time that the objects pass each other. The timing may be e.g. to initiate the action exactly when the objects meet, or more generelly when the x coordinate of k is between 3.5 and 2.5.

These two alternative strategies can be characterized by the following "strategy formula"

$$\exists t_1[1.5 < x_k < 2 \wedge \partial x_k = -1 \wedge Do(t_1, \mathit{openlid}, t_1 + 2)] \vee$$

$$\exists t_1[2.5 < x_k < 3.5 \wedge \partial x_k = -1 \wedge Do(t_1, \mathit{openlid}, t_1 + 2)]$$

It would be ideal if the model set for the axioms were exactly the set of interpretations where either of these two strategies is performed, while also the "laws of nature" in axioms 1 through 11 are also satisfied. If this were the case then the given axioms would entail the strategy formula (the disjunction of the possible plans) using the filter preferential model set criterium, and the conventional formula criterium in the sense of section 4.

Unfortunately there will also be other interpretations which satisfy all the axioms, but which rely on "coincidence" for obtaining their results. For example if the *openlid* action starts when h is in coordinate 2.75 and travelling in the positive direction, then k will certainly go into the shaft, and h may or may not go into the shaft depending on how quickly the lid opens. There will be some interpretations where the lid opens late enough for all the axioms to be satisfied.

The question of how to deal with such *coincident models* is discussed in [San89b], in the context of how to choose the formula criterium. Our concern here is only to make sure that the filter preferential model set for axioms such as those given above, obtains exactly those interpretations where the course of events in the world has been decoded correctly. In particular it must be required that all interpretations which are in accordance with either of the two primary strategies, and which are in accordance with the general laws, remain in the filter preferential model set.

It is easily seen that all interpretations of that kind, which have only one *openlid* action in their P component, satisfy all the axioms. It is also clear that they are minimal with respect to $\prec$ since no interpretation with an empty P component could satisfy all the axioms. Furthermore if the interpretations only have discontinuities where "necessary" i.e. where the objects bounce against the ends of the range, and when k starts falling, they will be minimal with respect to $\ll$.

Interpretations with more than one action in their P components can not be members of the model set. Even if they satisfy all the axioms, they are still not $\prec$-minimal since one can remove the redundant actions and obtain another interpretation which also satisfies all the axioms.

Interpretations where the objects are allowed to bounce several times before the lid is opened, will also be members of the model set and rightfully so.

Other possible forms of entailment conditions or model set criteria exhibit various kinds of interesting bugs. For example, one concern in the choice of criteria is to not introduce unnecessary actions which may account for natural discontinuities. Suppose there is some type of discontinuity, for example an object falling over an edge, which may both be the direct result of an action, and be the natural effect of previous movements. One does not wish an interpretation where the discontinuity occurs as a natural effect, to be dominated in the sense of the preference relations, by another interpretation containing an extra action which has the discontinuity as an effect. The filter preferential model set does not have that bug. However if the definition of $\ll$ is changed by omitting the condition $P = P'$, then exactly this bug is obtained.

The filter preferential model set has one property which is probably not desired, namely that it "prefers" a coincident model with fewer actions, to a non-coincident model with a larger set of actions. In other words, if certain observations of the world can be understood either as a coincidence due to a particular combination of parameter values etc., or as the necessary result of an action independently of coincidence, then the filter preferential model set will only contain interpretations of the former kind. It remains to find a good correction to this anomaly.

Chronological minimization of discontinuities is a natural preference criterium; it captures the one-way, non-symmetric character of time in the real world.

However one must make an exception from chronological minimization within the time-span of actions. The masking "X" relations are of course the technical device for realizing those exceptions. This was illustrated also by the example above: without the exceptions, CMD would prefer interpretations where the lid opens and closes as late as possible.

In our particular example the action involved discontinuities, at arbitrarily chosen times, of a propositional fluent ("mode"). Actions involving quantitative feedback, which proliferate in many real-world applications, are seen in our system as introducing discontinuities for quantitative parameters at arbitrary times.

References

[dKB85] Johan de Kleer and John Seeley Brown. A qualitative physics based on confluences. In Daniel G. Bobrow, editor, *Qualitative Reasoning about Physical Systems*, pages 7–83. MIT press, 1985.

[Kau88] Henry Kautz. Formal theories of plan recognition. position statement. In Jay Weber, Josh Tenenberg, and James Allen, editors, *From Formal Systems to Practical Systems. Advance Proceedings from the Rochester Planning Workshop*, pages 64–70, 1988.

[San89a] Erik Sandewall. Combining logic and differential equations for describing real-world systems. In *Proc. International Conference on Knowledge Representation, Toronto, Canada*, 1989.

[San89b] Erik Sandewall. On the variety of minimization related conditions for reasoning about actions and plans. Technical Report LAIC-IDA-89-TR9, IDA, 1989. Preliminary version.

[Sho88] Yoav Shoham. *Reasoning about Change.* MIT Press, 1988.