The range of applicability of nonmonotonic logics for the inertia problem

Erik Sandewall
Department of Computer and Information Science
Linköping University
58183 Linköping, Sweden
E-mail ejs@ida.liu.se

Abstract

We introduce and use a new methodology for the study of logics for action and change. The methodology allows one to define a taxonomy of reasoning problems, based in particular on the properties of the actions in those worlds where the actions take place, and on the autoepistemic assumptions that are being made. For each of a number of previously proposed logics, we have identified a corresponding class in the taxonomy, and proved that for reasoning problems within that class, the logic is guaranteed to obtain exactly the intended set of conclusions.

1 Logics of action and change: a problem of verification

There has been much research in recent years on methods for reasoning about actions and change, and on finding solutions to the so-called "frame problems". New variants of nonmonotonic logics for common-sense reasoning have been proposed, only to be quickly refuted by counterexamples. Unfortunately the results that have been obtained in this fashion are notoriously unreliable. According to the standard research methodology in the area, the evidence in favor of a proposed logic should consist of intuitive plausibility arguments and a small number of scenario examples for which the logic is proven (or claimed) to give the intended conclusions and no others.

Clearly there is a need for more systematic results, where a proposed logic is verified for a whole class of reasoning problems and not only for single examples. Recently, Lin and Shoham[LS91], Lifschitz[Lif91], and Reiter[Rei91] have reported such correctness results for several nonmonotonic logics which are based on the situation calculus. Their approach has been to relate a nonmonotonic theory over a certain set of "common-sense" axioms, to a monotonic theory over a larger set of axioms.

In this paper I present another approach which differs from these previous authors in three ways. It addresses logics with *explicit time*, such as the integers, and not only the situation calculus. This allows one to deal with actions with extended duration, and to analyze plans where the order of the actions is indetermined. Also it is based on an *underlying semantics* which captures basic notions of intelligent agents. This is hoped to facilitate the use of these results for the design of practical autonomous agents. Finally, rather than starting with a fixed class of reasoning problems and analyzing whether a single, proposed logic is correct or not for that class, I start by defining a *taxonomy* of reasoning problems. For each of several different logics I can then identify some class of reasoning problems within the taxonomy wherein the logic is provably correct.

The use of the taxonomy allows us to compare the range of applicability of different proposed logics. It is not clear that a broader-range logic will always be preferred, since a logic with a more narrow range of applicability maybe allows a more efficient implementation. However in order to make use of it one must have precise knowledge of whether it is correctly applicable for the application one has at hand.

The underlying semantics captures the basic A.I. intuitions, similar to the "agent model" of Genesereth and Nilsson [GN87]. In particular the notion of inertia is built into the underlying semantics. The semantics is used both for defining the taxonomy of reasoning problems, and as the basis for the assessments of applicability.

The definition of correctness for a logic is that for a specified class of reasoning problems, the set of intended models (as defined by the underlying semantics) equals the set of selected or preferred models. It is therefore a soundness-and-completeness condition and not only a soundness condition.

The present paper summarizes the current results in this research, and is by necessity quite brief. For the full account please refer to the much more detailed presentation in a forthcoming book hereafter referred to as "the book". A preliminary version is available as a departmental technical report[San92].

2 Surface logics

Before proceeding to the underlying semantics we shall outline the logic as such, with respect to its syntax and surface semantics. The logic is essentially a two-sorted first-order logic, with "time-points" and "physical objects" as the two sorts. In addition there is a type of "features" and one or more types for the value domains of features. These latter types are second class in the sense that it is not possible to quantify over them, and their domains are held fixed across interpretations.

A statement such as "the color of box 5 is green at time 620" would be expressed as

$$[620]\, color(\#5) \doteq \mathtt{G}$$

which can be understood as a syntactically sugared variant of

$$Holds(color(\#5), \mathtt{G}, 620)$$

Here $color(\#5)$ is an expression whose value is a feature, and G for green is a member of the value domain for the feature. Propositional features such as "alive" or "loaded" are treated as features whose value domain is $\{\mathtt{T}, \mathtt{F}\}$.

The temporal prefix may use an integer timepoint such as 620 or a timepoint constant symbol such as t_2. It may also refer to an open or closed interval rather than a single timepoint, for example

$$[620, t_2]\, color(\#5) \doteq \mathtt{G}$$

represents the fact that the object is green at each individual timepoint between times 620 and t_2 inclusive. Terms such as #5 refer to specific physical objects, and are analogous to integer timepoints in that they are the same across all interpretations. One may also use object constants whose values are specified in each interpretation, e.g.

$$[620, t_2]\, color(o_5) \doteq \mathtt{G}.$$

Actions are expressed using action symbols which may have arguments, e.g. as in

$$[530, 560]\, Paint(o_5, \mathtt{G}).$$

In such a case the same action does *not* apply over subintervals or superintervals of $[530, 560]$. Ordering relations between timepoints are expressed in the natural way, for example as in $t_2 > 620$.

The surface semantics is straightforward: for a given object domain $\mathcal{O}$, an interpretation I is a pair $\langle M, R\rangle$, where M assigns values to temporal and object constants independently of time, and R assigns a value to each combination of feature and timepoint. For a feature symbol with arguments such as *color* above, there is one feature for every choice of the argument as an object name (e.g. #5) in the given object domain. I will write $R(t)$ for $\lambda f.R(f,t)$.

The set of all features will be denoted $\mathcal{F}$. A *state* r is a mapping from features to feature values in the appropriate domain for each feature. The set of all states will be denoted $\mathcal{R}$.

The full details of the syntax and semantics are given in the book.

3 Underlying semantics: the game

The underlying semantics is intended to capture the intuition of situations where there is a world with inertia, so that features do not change value unless there is a positive reason why they must or may do so, and agents which may perform actions that override the inertia for some of the features. It is defined in two steps, first as a *game* between an ego and a world, and then as a formal characterization of the actions that the ego may invoke during the game.

For the purpose of the game, the ego is understood as the "knowledge level" [New82] of an autonomous agent or robot, and the world is understood as the combination of the physical robot vehicle and the robot's environment. The ego and the world can therefore be assumed to communicate in terms of discrete descriptions of the environment and invocations of actions which the ego requests and the world performs.

Formally, the ego and the world interact in terms of *finite developments*, which are tuples $\langle \mathcal{B}, M, R, \mathcal{A}, \mathcal{C}\rangle$. Here $\mathcal{B}$ is a set of timepoints (integers) whose largest member n is referred to as the *now*. M is like in interpretations. R is a history of the world from time 0 to time n, and formally a mapping from $[0, n]$ to $\mathcal{R}$. $\mathcal{A}$ is a set of actions which have been completed at time n, i.e. a set of triples $\langle s, A, t\rangle$ where $s < t \leq n$. Finally $\mathcal{C}$ is a set of actions which have been started but not completed at time n, represented as pairs $\langle s, A\rangle$ where $s \leq n$.

The interaction between the ego and the world is realized as a game where the two players alternate, and the moves consist of sucessively modifying and extending a partial development which serves as the "board" of the game. If $\mathcal{J}$ is the domain of all finite developments, then an ego $\mathbf{K}$ is formally speaking a mapping $\mathcal{J} \rightarrow \mathcal{J}$ specifying what move the ego does in each possible case, and a world $\mathbf{W}$ is formally a subset of $\mathcal{J} \times \mathcal{J}$ where $\forall j \exists j' [\mathbf{W}(j, j')]$, which specifies nondeterministically what move the world does in each case that may arise.

The moves of the ego must leave the first and third component of the development unchanged, which means in particular that the ego does not change the setting of "now". Typically the move of the ego is to add an element $\langle n, A\rangle$ to $\mathcal{C}$, meaning to invoke an action starting at the present now-time n. The moves of the world are typically to increase "now", extend R correspondingly, and to terminate actions by moving them from $\mathcal{C}$ to $\mathcal{A}$ while adding the present "now" as the ending time of the action.

4 Underlying semantics: the trajectory semantics

Although the world is in the most general case a relation over $\mathcal{J} \times \mathcal{J}$, one needs to impose some more structure on it in order to obtain any useful results. We define worlds in terms of a *trajectory semantics.* A *trajectory* for a set $F \subseteq \mathcal{F}$ of features is a sequence $\langle r'_1, \ldots r'_k\rangle$ where each r'_i is a partial state assigning values to the features in F.

Each world is characterized as a pair $\langle \mathtt{Infl}, \mathtt{Trajs}\rangle$, where Infl (influenced features) and Trajs (trajectories) are functions such that if A is an action and $r \in \mathcal{R}$ is a state, then $\mathtt{Infl}(A, r) \subseteq \mathcal{F}$, and $\mathtt{Trajs}(A, r)$ is a non-empty set of trajectories for $\mathtt{Infl}(A, r)$.

Informally speaking, $\mathtt{Trajs}(A, r)$ is the set of possible ways that the action can be realized in the world. A world description $\langle \mathtt{Infl}, \mathtt{Trajs}\rangle$ specifies a world $\mathbf{W}$ with the following behavior in the game. If the ego has invoked the action A at time s by adding the pair $\langle s, A\rangle$ to the $\mathcal{C}$ component of the development, then the world chooses an arbitrary member $\langle r'_1, ...r'_k\rangle$ of $\mathtt{Trajs}(A, R(s))$. The world's move is to modify the current partial development by (1) adding $s+k$ as a member of $\mathcal{B}$, thereby making it the new "now" time; (2) extending the history component R by defining $R(s+i) = R(s) \oplus r'_i$, for $1 \leq i \leq k$. Here the symbol $\oplus$ represents "override" so $[u \oplus v](f)$ equals $v(f)$ when it is defined, otherwise $u(f)$; (3) adding the completed action $\langle s, A, s+k\rangle$ to the $\mathcal{A}$ component of the development; (4) resetting the $\mathcal{C}$ component for ongoing actions to the empty set.

After this it is the ego's turn to make its move, which it does in the new situation for the new "now" of $s+k$. The described procedure works for sequential actions; otherwise a more elaborate definition is required.

Complete developments are obtained in the game as the limiting case as time goes to infinity, and are characterized by $R(t)$ being defined for all non-negative t. They can be obtained in two ways: either the game goes on with an infinite number of moves each of which has a finite duration, or the agent ceases to make any moves and the world remains constant to infinity after the agent's last move.

5 Chronicles

A scenario will not be described as a single set of logic formulae, but as a tuple $\langle \mathcal{O}, \mathbf{A}, \text{SCD}, \text{OBS}\rangle$ (possibly with additional components), where $\mathcal{O}$ is a set of objects $\{\#1, \#2, ...\#k\}$ that are used in interpretations e.g. as arguments when constructing features. The other elements in the tuple are sets of formulae. **A** is a set of "laws" characterizing the effects of actions, and is in fact an exhaustive description of `Trajs` in logic. SCD is a schedule i.e. a set of formulae characterizing the actions, chosen among the following two simple kinds:

- $[s,t]\,E(...)$
- A disjunction of one or more $s < t$ or $s \leq t$.

where E is an action name, e.g. *Paint* in the example above, and the arguments may be objects or feature values. In this way it is possible to specify what actions there are, and what are the values or constraints on the timing of each action. Finally OBS is a set of observation statements, which can be chosen as any formulae not containing any action statements. A scenario of this kind will be called a *chronicle.*

I will write $\mathbf{A}(\text{SCD})$ for the result of replacing each action statement in SCD by the effects specified by the laws in **A**. For example the effect of loading the gun in the Yale shooting scenario could be expressed as the following formula in **A**:

$$[s,t]\,Load \Rightarrow [t]\,l \mathrel{\hat{=}} \mathrm{T}$$

saying that if the gun-loading action takes place over the interval from s to t, then the feature for the gun being loaded is true (T) at time t. If SCD has the member $[4,6]\,Load$, $\mathbf{A}(\text{SCD})$ would then in its place have the member $[6]\,l \mathrel{\hat{=}} \mathrm{T}$. From a formal point of view, action statements are considered to be a separate language, and action laws in **A** are rules for translating that language to the main language.

6 Intended models

The correctness of a logic was defined above in terms of equality between the sets of intended and selected models, and we can now finalize the definition of the intended models. If a chronicle $\langle \mathcal{O}, \mathbf{A}, \text{SCD}, \text{OBS}\rangle$ is given, then the set of ***intended models*** is defined using the set of infinite developments obtained as follows. Select an arbitrary world **W** which is exactly described by **A**, and select also an arbitrary ego and an arbitrary initial state. Generate all possible developments which can be obtained in games between them. Add an arbitrary M component (mapping from constant symbols to corresponding values) to each development. Then restrict the set of developments to those where all formulas in $\text{SCD} \cup \text{OBS}$ are satisfied, and where there is a one-to-one correspondence between actions in the development and action statements in SCD. Finally extract the M and R components from the remaining developments, obtaining a set of interpretations $\langle M, R\rangle$. This is the set of intended models for the given chronicle.

7 The ontological and epistemological taxonomy

I will characterize classes of systems using a set of letters, where each letter indicates the presence of some special property or "speciality". Basic inertia or the classical frame problem is denoted as **IA**, where **I** stands for inertia as such, and **A** represents "alternative results": the results of an action are conditional on the starting state, for example as when firing the gun in the Yale shooting problem. **C** represents that concurrent actions are allowed, **L** that actions may have delayed effects (resulting changes that occur after the period of the action itself), etc.

The trajectory semantics defined above corresponds to the ontological family **IAD**, where **D** represents dependencies between features. The simpler case **IA** is obtained by imposing the following restriction on the trajectory semantics. For each action there must be a *range* of influence for the action, consisting only of those objects which occur as the action's arguments. The result of the action may only depend on features where all arguments are in the action's range, and only those features can have their values changed by the action. (Therefore ramification is in the **IAD** family).

In addition there is a need to introduce ***sub-specialities*** which provide additional detail in the taxonomy and which will be written with a small letter. For example **Is** denotes the subfamily of **I** systems where all actions take a single time-step. **Id** denotes the subfamily where in every action and for every feature affected by the action, the feature makes a single change from its old to its new value. **Ad** denotes the subfamily of **A** where all actions are deterministic, and so on.

All of these subspecialities can be precisely defined in terms of the trajectory semantics. For example **Is** is characterized so that every member of $\mathtt{Trajs}(A, r)$ is a trajectory of length 1, for every A and r. **Ad** is characterized so that $\mathtt{Trajs}(A,r)$ is a singleton for every A and r, and so on. The full catalogue of specialities and subspecialities, with their exact definitions, are in the book.

There is also a need to characterize the epistemological assumptions, for example regarding complete information about actions. I will use $\mathcal{K}$ to denote the basic assumption of full knowledge about actions, as defined above. $\mathcal{Q}$ will represent the weaker case that is often used in planning, i.e. that for each action there is a known precondition, and in case the precondition is satisfied there is full knowledge about the possible effects of the action, otherwise not. $\mathcal{K}$s represents that in addition there is full knowledge about the initial state (time 0), and $\mathcal{K}$p that there are no observations about any timepoint after the initial one. The combination of the last two conditions is written as $\mathcal{K}$sp.

A combination of an epistemological and an ontological descriptor is formed as for example $\mathcal{K}$**−IsAd**. Such a combination characterizes a class of systems that one may wish to reason about (in this case single-step and deterministic actions) and a specific requirement on how a chronicle may describe a set of developments. Such combinations are used for characterizing the applicability of logics for action and change. If a logic is stated to be correct for chronicles in the $\mathcal{K}$p-**IsAd** family, it means that if one chooses an arbitrary chronicle which is formed according to the syntactic restrictions in $\mathcal{K}$p, then the set of models that the logic obtains for that chronicle is essentially equal to the set of infinite developments that are obtained by the following process:

- choose an arbitrary world, described by the trajectory semantics, which satisfies the **IsAd** restrictions and which is correctly described by the **A** component of the given chronicle;
- choose an arbitrary ego;
- obtain all the possible games between the chosen world and the chosen ego, from arbitrary initial states;
- among the developments in those games, select those which are correctly described by the given chronicle according to the requirements of $\mathcal{K}$p.

8 Assessments of some simple nonmonotonic logics for action and change

The following are the assessments of some currently proposed logics for action and change. For the full proofs and for some fine points regarding the conditions, please refer to the book. Throughout it is assumed that the schedule part of the chronicle is constructed so that the actions are necessarily sequential.

8.1 Original chronological minimization

The original chronological minimization (OCM) according to Kautz[Kau86] is correct for $\mathcal{K}$sp-**IsAd**. In other words, the initial state must be completely specified, there must not be any observations for times later than the initial one, and all actions must take a single timestep and be deterministic. It is easy to find counterexamples when any of these restrictions is violated.

8.2 Prototypical chronological minimization

With a minor correction, OCM can be changed to prototypical chronological minimization (PCM) which is correct for $\mathcal{K}$p-**IsAn**. Here the initial state does not have to be completely specified, but still there must not be any observations for times later than the initial one. All actions must take a single timestep and satisfy the condition of "necessary change", which is weaker than the deterministic requirement. For example if a feature with three possible values red, yellow, green is influenced by an action, then the action is allowed to nondeterministically change the value from red to yellow or red to green, but it is not allowed to choose between switching from red to green or keeping it red.

The analysis of Lin and Shoham, which used another methodology (compare section 1), considered the $\mathcal{K}$p-**IsAd** family of reasoning problems, and they proved that PCM is correct for that family. The present result for PCM confirms and subsumes theirs.

8.3 Prototypical global minimization

The original proposal by McCarthy[McC84], which Hanks and McDermott formulated in general form and gave a counterexample for in their Yale shooting problem paper[HM87], can be characterized as prototypical global minimization of change (PGM). It is correct for consistent linear chronicles in the $\mathcal{K}$**−IsuAn** family. Here there are no restrictions on the timepoints that observations refer to, but actions must be single-step (**Is**), satisfy the necessary change condition (**An**), and satisfy uniform change (**Iu**). The last condition says that $\mathtt{Infl}(A,r)$ must be independent of r, i.e. the set of features that change as the result of the action must be independent of the starting state. This means that only delay actions (having no effect except the passage of time) and toggle-type actions are allowed. Even an action such as loading the gun (if the gun was unloaded it becomes loaded, if it was already loaded then nothing happens) does not satisfy uniform change.

The definition of linear chronicle is roughly speaking that the temporal order of all actions and observations must be the same in all classical models of the chronicle.

The $\mathcal{K}$**−IsuAn** family is very restricted. However it does not seem possible to strengthen the result: even in a slightly broader family there are counterexamples where PGM gives incorrect results. It is a surprising fact that PGM is restricted to single-step actions.

8.4 Formal definitions of PGM, OCM, and PCM

The formal definitions of the three entailment criteria that have been discussed so far are as follows. Let a chronicle $\langle \mathcal{O}, \mathbf{A}, \textsc{scd}, \textsc{obs}\rangle$ be given, and let W be the set of those classical models $\langle M, R\rangle$ for $\mathbf{A}(\textsc{scd}) \cup \textsc{obs}$ where $\mathcal{O}$ is the object domain. Then select a subset of W, which will be called the *selected* models, defined as the minimal ones according to the preference relations

$\ll_{pgm}$, $\ll_{ocm}$, and $\ll_{pcm}$ which are defined as follows. Let $I = \langle M, R\rangle$ and $I' = \langle M', R'\rangle$ be two members of W. The changeset of I is defined as

$$changeset(I) = \{\langle f, t\rangle \mid R(f, t-1) \neq R(f, t)\}$$

i.e. the set of pairs of a feature and a timepoint t where the feature changes value from time $t-1$ to time t. Then $I \ll_{pgm} I'$ iff $M = M'$ and $changeset(I) \subset changeset(I')$. The breakset of I at time t is defined as

$$breakset(I, t) = \{f \mid R(f, t-1) \neq R(f, t)\}$$

i.e. the set of features which change value from time $t-1$ to time t. Then $I \ll_{ocm} I'$ iff $M = M'$ and there is some timepoint $\mathbf{t}$ such that both

- for all $t < \mathbf{t}$, $breakset(I, t) = breakset(I', t)$
- $breakset(I, \mathbf{t}) \subset breakset(I', \mathbf{t})$.

Also $I \ll_{pcm} I'$ iff $M = M'$ and there is some timepoint $\mathbf{t}$ such that both

- for all $t < \mathbf{t}$, $R(t) = R'(t)$
- $breakset(I, \mathbf{t}) \subset breakset(I', \mathbf{t})$.

9 Assessments of logics based on occlusion and filtering

It is already known that the logics that were assessed in the previous section fail fairly easily, and the new results give a precise form to that insight. In order to obtain a logic that gives correct results for a larger class of systems while still retaining the preferential character, it is necessary to use the concepts of *filtering* and *occlusion*.

The idea with filtering [San89] is to separate the premises in the sets SCD and OBS. In the methods defined above, the set of selected models is chosen by preferential entailment as

$$Min(\ll, [\![\mathbf{A}(\text{SCD}) \cup \text{OBS}]\!])$$

where $[\![\Gamma]\!]$ denotes the set of classical models for the set Γ. In filtered preferential entailment it is instead chosen as

$$Min(\ll, [\![\mathbf{A}(\text{SCD})]\!]) \cap [\![\text{OBS}]\!]$$

so that the minimization is imposed "before" the observations. If the preference relation $\ll$ is chosen as $\ll_{pcm}$ then the resulting logic is correct for the $\mathcal{K}$–**IsAn** family of chronicles, which is already some improvement.

The idea with occlusion [San89] is to not minimize change directly, but instead to have a separate property of occlusion. Interpretations are extended from $\langle M, R\rangle$ to $\langle M, R, X\rangle$ where X is a relation over features times timepoints. An action law saying that a feature f changes its value from x to any of $\{x_1, x_2, ... x_k\}$ over an interval of time $[s, t]$, is now expressed as saying that f is occluded throughout $(s, t]$, and in addition there is some information about its value within or at the end of the interval. The minimization criteria are changed so that they minimize *unoccluded* change rather than any change. Also of course occlusion itself has to be minimized. Observations are not allowed to use the occlusion operator.

The exact syntax for referring to the occlusion predicate X in action laws is not important here. It can be found in the book, together with a more extensive motivation of the intuitions for occlusion and filtering.

9.1 Prototypical chronological minimization with filtering

If filtering is used with the $\ll_{pcm}$ preference ordering, but without the use of occlusion, then the resulting entailment criterion is correct for all $\mathcal{K}$–**IsAn** temporal reasoning problems. This subsumes the three methods described above, but is still restrictive.

9.2 Chronological minimization of occlusion and change

A preference relation that chronologically minimizes occlusion and change together, is defined as follows and will be referred to as CMOC. Let $\langle \mathcal{O}, \mathbf{A}, \text{SCD}, \text{OBS}\rangle$ be a chronicle in $\mathcal{K}$–**IA**, and let $I = \langle M, R, X\rangle$ and $I' = \langle M', R', X'\rangle$ be two models for $\mathbf{A}(\text{SCD})$, both having $\mathcal{O}$ as their object domains. Define a modified breakset function as

$$brs(I, t) = \{f \mid R(f, t-1) \neq R(f, t) \wedge \neg X(f, t)\}$$

i.e. as the set of features having unoccluded changes at time t. Then the preference relation $\ll_{cmoc}$ is defined so that $I \ll_{cmoc} I'$ iff $M = M'$ and there is some timepoint $\mathbf{t}$ such that both of the following hold:

- For all $t < \mathbf{t}$, $R(t) = R'(t)$ and $X(t) = X'(t)$.
- Either of the following hold:
  - $X(\mathbf{t}) \subset X'(\mathbf{t})$
  - $X(\mathbf{t}) = X'(\mathbf{t}) \wedge brs(I, \mathbf{t}) \subset brs(I', \mathbf{t})$.

This entailment criterion was tested against a fairly extensive set of test examples, which contains or subsumes most problems that have been used by other authors in the field, and it passed all the tests. However when attempting to prove the correctness of this criterion for $\mathcal{K}$–**IA** it was discovered that it is in fact not completely general, and the correct assessment is that CMOC is correct for $\mathcal{K}$–**IAe** chronicles. The **Ae** subspeciality says that actions are *equidurational*, i.e. the duration of an action or the set of possible durations of an action is not allowed to differ depending on the starting state of the action. This is in fact a quite strong restriction: certainly if you allow actions to have conditional effects depending on the starting state and you allow them to have a duration different from 1, you would expect that the duration of the action may depend on the starting state. However once you know of this restriction, it is also easy to construct counterexamples outside $\mathcal{K}$–**IAe** where CMOC gives the wrong results.

9.3 Chronological assignment and chronological minimization of occlusion and change

Fortunately it is possible to correct the problem and have a method which obtains the correct results throughout $\mathcal{K}$–**IA**. The modified method, CAMOC, is characterized by an additional syntactic restriction and a modification of the preference relation.

The syntactic restriction is as follows. All action statements must be written on the form $[s_i, t_i]A_i$ where s_i and t_i are constant symbols, and the same constant symbol is only used in one single action statement. The schedule SCD may only contain such action statements, temporal ordering statements of the form $s_i < t_i$ or $t_i \leq s_j$ which serve to order the actions, and disjunctions between the latter kind of statements. All other information about the order and distance between timepoints must be placed in the set OBS of observations.

The preference relation is modified as follows. Given a chronicle and two models I and I' like for CMOC, we first define $M_{0:t}$ as a partial function which is obtained from M by restricting it to object constants and to those temporal constants whose value is $\leq t$. (Remember that M is a mapping from constant symbols to corresponding values). This operation is a restriction on a function to a restricted value range, not to a restricted argument range.

Then the preference relation $\ll_{camoc}$ is defined as follows: $I \ll_{camoc} I'$ iff there is some timepoint $\mathbf{t}$ such that all of the following hold:

- $M_{0:\mathbf{t}} = M'_{0:\mathbf{t}}$.
- for all $t < \mathbf{t}$, $R(t) = R'(t)$ and $X(t) = X'(t)$.
- Either of the following hold:
  - $X(\mathbf{t}) \subset X'(\mathbf{t})$
  - $X(\mathbf{t}) = X'(\mathbf{t}) \wedge brs(I, \mathbf{t}) \subset brs(I', \mathbf{t})$.

The set of selected models is defined like before as

$$Min(\ll_{camoc}, [\![\mathbf{A}(\text{SCD})]\!]) \cap [\![\text{OBS}]\!].$$

It has been proved that this entailment criterion is correct for all $\mathcal{K}$–**IA** chronicles.

10 Syntactic approaches

Schubert[Sch90] and Reiter[Rei91], following a proposal by Haas, have recently analyzed a "monotonic" approach to reasoning about action and change, called *explanation closure*. This approach works by incrementing the axiom set with additional axioms. It therefore represents a syntactic method for reducing the intended model set, as compared to the semantic method in the logics discussed here. The explanation closure approach is monotonic once the additions to the original axiom set have been performed, but as seen from the original axioms it is still a nonmonotonic method.

Schubert's and Reiter's analyses are restricted to situation-calculus formalizations, and it remains to obtain a more general analysis for all cases offered by the use of explicit time and the trajectory semantics.

11 Summary

The results of this paper are on two levels. A new, systematic methodology was defined in the first half of the paper. The second part of the paper presented a number of hard results which have been obtained using this methodology. In particular we have shown that although several widely accepted temporal reasoning methods are not correct for unrestricted use, for certain classes of reasoning problems they are provably correct. In several cases we have also found that the methods were more restricted than what was previously known or intuitively obvious.

The systematic methodology has made it possible to identify and validate the CAMOC entailment criterion which has been proven correct for all $\mathcal{K}$–**IA** problems i.e. all problems with simple inertia. Work is in progress for assessing additional, existing logics using the same methodology, and it seems that the other classical aspects of the frame problem (such as ramification) can be analyzed by extension of the results described here.

References

[GN87] Michael R. Genesereth and Nils J. Nilsson. *Logical Foundations of Artificial Intelligence.* Morgan-Kaufmann Publishing Co., 1987.

[HM87] Steve Hanks and Drew McDermott. Nonmonotonic logics and temporal projection. *Artificial Intelligence*, 33(3):379–412, 1987.

[Kau86] Henry Kautz. The logic of persistence. In *Proc. AAAI 1986*, pages 401–405, 1986.

[Lif91] Vladimir Lifschitz. Toward a metatheory of action. In *International Conf. on Knowledge Representation and Reasoning*, pages 376–386, 1991.

[LS91] Fangzhen Lin and Yoav Shoham. Provably correct theories of action (preliminary report). In *National (U.S.) Conference on Artificial Intelligence*, pages 349–354, 1991.

[McC84] John McCarthy. Applications of circumscription to formalizing common sense knowledge. In *Proc. of the Nonmonotonic reasoning workshop*, pages 295–324, October 1984.

[New82] Allen Newell. The knowledge level. *Artificial Intelligence*, 18(1):87–127, 1982.

[Rei91] Ray Reiter. The frame problem in the situation calculus: a simple solution (sometimes) and a completeness result for goal regression. In Vladimir Lifschitz, editor, *Artificial Intelligence and Mathematical Theory of Computation*, pages 359–380. Academic Press, 1991.

[San89] Erik Sandewall. Filter preferential entailment for the logic of action in almost continuous worlds. In *Proc. International Joint Conf. on Artificial Intelligence, Detroit, USA*, 1989.

[San92] Erik Sandewall. Features and fluents. Review version of 1992. Technical Report LiTH-IDA-R-92-30, Linköping University, Department of Computer and Information Science, 1992.

[Sch90] Lenhart Schubert. Monotonic solution of the frame problem in situation calculus. In Henry E. Kyburg, Ronald P. Loui, and Greg N. Carlson, editors, *Knowledge Representation and Defeasible Reasoning*, pages 23–67. Kluwer, 1990.