Time in a Causal Theory

Aïcha MOKHTARI
Institut d'Informatique
USTHB
BP 32 El Alia Alger Algérie
mokhtari@ist.ibp.dz

Daniel KAYSER
LIPN URA 1507 du CNRS
Institut Galilée Université Paris-Nord
93430 Villetaneuse France
Daniel.Kayser@ura1507.univ-paris13.fr

Abstract

This paper discusses the temporal aspect of a causal theory based on an "interventionist" conception of causality, i.e. a preference to select causes among a set of actions which an agent has the ability to perform or not to perform (free will). Casting causal reasoning in this framework leads to explore the problem of reasoning about actions, generally considered as a nonmonotonic temporal reasoning.

Most of the works on nonmonotonic temporal reasoning have used simple temporal ontologies, such as situation calculus, or temporal logic with discrete time. The theory presented in this paper also has a simple temporal ontology based on "time points" organized on branching "time lines", with the possibility of modelling continuous evolutions of the world for various futures (prediction) or pasts (diagnostic).

1. Introduction

A definition of the concept of cause, if at all possible, would involve deep philosophical questions ; we do not need to tackle them, however, in order to use a practical notion of cause. Intuitively, this notion is necessary in our everyday reasoning, both to anticipate what should happen if we decide to perform an action, and to diagnose what might have happened to yield a given state of affairs.

We propose to prune the collection of propositions which might be considered as causes by preferring to take as causes the result of the **free will** of an agent, i.e. his/her ability to opt for performing or not performing a given action. Casting causal reasoning in this framework leads to explore the problem of reasoning about actions, generally considered to be a nonmonotonic (i.e. defeasible) temporal reasoning.

Temporal reasoning is said to require a choice among two ontologies : point-based or interval-based. Although we consider that many non-trivial inferences can be achieved without ever making that choice (see e.g. [4]), we adopt in this paper a point-based point of view, for reasons which will appear in the next section. However, it is likely that an interpretation in terms of intervals of what we call below time points would not change drastically the core of our approach.

The next section motivates our choices, in order to satisfy the kind of reasoning we wish to capture. We then go on to provide a formalism intended to link an incomplete description with any normal course of the world: we give preliminary notations and definitions, then describe a theory of action, and compare it with related works. We conclude with a discussion about current and future work.

2. Actions, effects, and time points

Actions and effects will be the only temporal propositions considered in our framework. An effect can be, among other things, an event or a fact. But as argued in [13], a fine distinction is unnecessary. In order to introduce temporal aspects, some choices must be made.

The first one concerns the basic temporal element: point or interval ? Intervals can be related in various ways: "I1 is completely before I2"; "I1 abuts I2", "I1 overlaps I2", etc.[1]. All these relations are possible between a cause and an effect, but they are subsumed by the general principle, according to which effects never precede their causes; therefore we find it simpler to use only **time points.**

We do not mean to reduce causation merely to a temporal relationship: what we present below shows the contrary. Let us mention, in addition, that most approaches choose discrete sets, isomorphic to integers, to represent time. In order to better reflect our intuitions on continuity, we prefer to take reals, but we do not consider this choice as critical, especially as our examples need only to consider a finite number of time points.

Our second choice amounts to select either a linear or a branching model of time ; time is intuitively linear. However, as we deal with choice making, the very representation of a choice immediately suggests the use of **branching time**. Among the time points, some particular ones, called *choice points,* are intended to represent states of affairs where an agent can take the decision of performing or not an action: obviously, not all time points are choice points, since the conditions allowing for the action are not always satisfied.

The decision of the agent is represented as a *time line* splitting into two futures, or more accurately, as two distinct time lines having the same time points up to the choice point. This is consonant with most systems, which have a branching future [10].

But we consider as well a branching past [14], because we often need to examine two different courses of events leading to the same situation. We also allow time lines to meet again in the future (case of an action without long-range effects, for example). Fig.1 below will provide an illustration.

We now present more formally the general framework corresponding to our choices.

3. Temporal ontology

Definition 1: *A **time point** t is a "snapshot" state of the universe defined by a subset of true propositions at a certain date and by this date. T is the set of time points.*

The subset mentioned in the definition is not arbitrary: we sometimes refer to it as the set of "interesting propositions", i.e. in our framework, the result of causal relations. The possibility of having a branching past implies that the "interesting propositions" do not include "historical" statements, since otherwise different pasts could never lead to the same time point.

Having time points defined both by the subset of true propositions and by the date allows to distinguish several occurrences of the same state of affairs. If we need to represent cyclic phenomena, where this distinction is useless, we may add an equivalence relation on time points: $t1 \equiv t2$ if they differ only by their date, and then reason on the quotient set. The alternative, i.e. define time points only by the subset of true propositions, does not allow to restore the notion of date when needed.

To extract the date of a time point, we define a function *date* : $T \mapsto \Re$ which maps every time point on a real number representing its date. To simplify, we write *date(t)* as d_t.

Definition 2: *We call **time line** l (somehow similarly to McDermott's "chronicle" [10]) a set of time points in bijection with the set of dates, meant to represent a possible evolution of the universe.*

A time line hence conveys the complete evolution of the truth value of the "interesting propositions". The time points of a time line are supposed to comply with the general principle: "there is no effect without a cause". Their propositions are then the result of cause-effect relations governed by **causal rules**. The set of causal rules is gathered in a rule base called *BR*.

The time points of a time line are totally ordered by a precedence relation written "$\leq$", where $t1 \leq t2$ means that time point t2 does not precede t1, hence whenever $t1 \leq t2$ we have $d_{t1} \leq d_{t2}$, but the converse does not hold (see Fig.1).

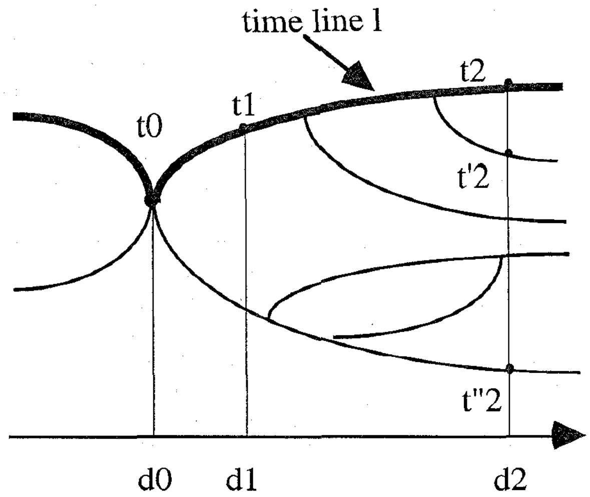

Figure 1: the structure of branching time in the past and in the future.
The thick line represents a time line, l, including among others time points t0, t1, t2. The other curved lines represent other time lines. t0 < t1 < t2 holds, as do t0 < t"2 and t1 < t'2, but there is no relation between t1 and t"2, although $d_{t1} < d_{t''2}$ is true.

4. The language

The proposed langage $\mathcal{L}$ is defined at two levels:

• the first level is meant to represent static information. It is a plain propositional language in which: P is a set of propositions we are interested in, A, subset of P, is a set of actions, and E, subset of P, is a set of effects, with $A \cap E = \emptyset$ and $A \cup E = P$.

• the second level expresses dynamic information. It contains predicates with time variables. If p is a formula of the first level, l a time line, $t \in l$ a time point, a formula of the second level has the form:

- $v(p,l,d_t)$ with the intended meaning that formula p is true in l at the date of time point t, and:

- $nocc(p,l,d_t,\Delta)$ with the intended meaning that p is never true in line l from the date of time point t on, during the delay Δ. In other words, $nocc(p,l,d_t,\Delta)$ is a short-hand for :

$$(\forall t')\ ((t' \in l \wedge 0 \leq d_{t'} - d_t < \Delta) \supset \neg v(p,l,d_{t'}))$$

We are going to extend gradually this language, but first let us define its semantics. The associated model theory is a generalization of Kripke [8] possible world semantics. In this model, an interpretation is defined as function I mapping the cartesian product $L \times \Re$ into a subset of propositions, i.e. $I : L \times \Re \mapsto 2^P$, where:

- L is the set of time lines,
- $\Re$ is the set of real numbers,
- $t \in l$ is true in the model if t has exactly as its true propositions the set $I(l,d_t)$,
- $v(p,l,d_t)$ is true in the model iff $p \in I(l,d_t)$, i.e. proposition p is true at the time point determined by time line l and date d_t.

It follows that $nocc(p,l,d_t,\Delta)$ is true in the model iff $(\forall t')\ ((t' \in l \wedge d_t \leq d_{t'} < d_t + \Delta) \supset p \notin I(l,d_{t'}))$.

The dependence of the effect on the cause may vary according to the context. We shall therefore introduce a subset of **preferred time lines** and augment the language with an operator denoted "$\Rightarrow$" meaning **normally implies** (in a more comprehensive presentation, see [11], we also have an operator "$\rightarrow$" meaning *implies in all cases*).

The intuitive idea behind these two new notions is as follows: when an agent chooses to perform an action, he or she does not anticipate every possible outcome of his or her choice: several circumstances, unknown to, or neglected by, the agent at the moment of the choice, may alter the predictable effect of the action. Informally, the "preferred" time lines are the futures that the agent normally "had in mind" when he or she opted to perform the action. The effects which are present in all "preferred time lines" following an action are said to be "normally implied" by the action. The idea of "normal implication" is inspired by the work of

Delgrande [3]. We defer a more precise explanation until we introduce some more notions.

To make sense of the notion of "normality" requires to reason with uncertain information: in the absence of specific information, we are entitled to believe that things behave normally. This brings us to a problem similar to the well-known "frame problem", which is inherent to any theory of change. We must therefore take into account:

- the preconditions of an action a, i.e. represent what is reasonable to assume whenever performing a is considered;
- the hindrances of an action a, i.e. represent the effects of other actions that can inhibit the effects of a;
- the persistence of states, corresponding to the fact that some propositions continue to hold true for some duration, unless an external event entails their falsity.

All these aspects generally require the use of a nonmonotonic reasoning. That is the reason why we devote next section to this issue.

5. Nonmonotonicity

In [11], we show how to include implicit premises in a normal inference. We suppose the existence of a function, called *norm*, to define the normal conditions under which an action is executed. Technically, $norm : A \mapsto 2^P$ is such that for any action a, $norm(a)$ contains those propositions (preconditions) which are true unless otherwise specified when an agent considers to perform a. Extending the domain of *norm* to our "first-level" language, i.e. defining compositionally *norm (a* op *a')* where op is a boolean operator is not a trivial task, if we want to remain compatible with our intuitions. We will not treat this problem here.

As we saw the importance of defining "hindrances", we suppose similarly the existence of a function *inhibit* that, for any couple $<e,a>$ where e is a normal effect of a, determines the events which are liable to prevent e from following a. Technically, $inhibit : E \times A \to 2^E$ is such that $inhibit(e,a)$ is the subset E' of E where $e' \in E'$ iff whenever e' occurs during the delay after a where e should turn true, e may actually remain false. Notice that action a' causing e' may happen before, with, or after a. Similarly, extending the domain of *inhibit* to couples of formulas instead of couples of atoms is a very thorny issue.

We turn now to what we mean by "normal case". This notion is often attached to a preference ordering, but we notice that the definition of a so-called "correct order" is rather difficult; therefore, we find it more convenient to use the notion of preferred time line, as we introduced it in §4. Its definition requires the preliminary notion of coincidence, viz.

Definition 3: *Two time lines l_1 and l_2* ***coincide*** *up to time point t, property written $coincide(l_1,l_2,t)$ iff for every time point t' preceding t,*

$$t' \in l_1 \equiv t' \in l_2.$$

In other words, a model satisfies *coincide* (l_1,l_2,t) iff $(\forall t')(t' \leq t \supset I(l_1,d_{t'})=I(l_2,d_{t'}))$

Definition 4: *The set of* ***preferred time lines*** *for line l at time point t, noted $Lp(l,d_t)$ is a subset of L obtained by a function $Lp : L \times \Re \mapsto 2^L$ such that : $(\forall l,l',t)\ (l' \in Lp(l,d_t) \supset coincide(l,l',t))$*

We are now in position to define formally normal implication:

Definition 5: *Action a* ***normally implies*** *effect e within the delay Δ, noted $a \Rightarrow e\ [\Delta]$, iff $(\forall l,t)$ $(C1 \wedge C2)$ where C1 and C2 stand for the following conditions:*

C1 $\{[v(a,l,d_t) \wedge (\forall p)\ (p \in norm(a) \supset v(p,l,d_t))] \supset (\forall l')\ (l' \in Lp(l,d_t) \supset [(\exists t')\ (t' \in l' \wedge d_t \leq d_{t'} \leq d_t+\Delta \wedge v(e,l',d_{t'})) \vee (\exists e',t'')\ (e' \in inhibit(e,a) \wedge t'' \in l' \wedge v(e',l',d_{t''}) \wedge d_t \leq d_{t''} \leq d_t+\Delta)])\}$

C2 $\{v(\neg a,l,d_t) \supset (\exists l')\ (l' \in Lp(l,d_t) \wedge nocc(e,l',d_t,\Delta))\}$

This rather intricate definition calls for some explanations:

C1 tells that whenever a occurs under normal conditions, in all preferred futures, either there exists a subsequent occurrence of event e within the delay Δ, or there is an occurrence of event e' known to inhibit the effect of a; e' must then occur after t within the prescribed delay (notice that, even in this case, e may become true);

C2 checks that if a is not executed, there exists at least one preferred future in which e will not occur within the specified delay Δ. This condition reflects the implicit counterfactuality always present in causation: we are not ready to say that a normally implies e if we think that, even without performing a, e will nevertheless occur in every likely future.

We now turn to the last problem related with the "frame problem", namely persistence.

Example 1: *Suppose that the following facts and rule are given. They represent the well-known "Yale Shooting Problem" (Y.S.P.) according to our notation:*

- $v(Fred\ is\ alive,l,d_{t0})$
- $v(gun\ is\ loaded,l,d_{t1})$

- $v(shoot\ at\ Fred, l, d_{t2})$
- $shoot\ at\ Fred \Rightarrow \neg alive\ Fred\ [\Delta]$

with $d_{t0} \le d_{t1} \le d_{t2}$ *and* Δ*: a few seconds.*

Two problems have to be considered in relation with the persistence of a given event *e*:

1. temporal nonmonotonicity, i.e. the possibility that an external event prevents *e* from remaining true, and
2. the estimation of the duration of the persistence of a fact, i.e. how long, after *e* has begun to be true, is it likely that it is still true ?

Suppose that we heve the answer to point 2., and let ∂ be the "normal duration" of fact *e*. We can define persistence as follows:

Definition 6 (persistence): *We note* $persist(e,\partial)$ *the fact that, without any external influence, event e is believed to remain true for a duration* ∂*. We have:*
$persist(e,\partial) \supset (\forall t_0,l)\ ((v(e,l,d_{t0}) \wedge (\exists t')\ (t' \in l \wedge nocc(e,l,t',d_{t0}\text{-}d_{t'}))) \supset (\forall l',t)\ ((t \in l' \wedge C1 \wedge C2 \wedge C3) \supset v(e,l',d_t)))$
where C1, C2, and C3 abbreviate the following conditions:
C1 $l' \in Lp(l,d_{t0})$
C2 $d_{t0} \le d_t \le d_{t0} + \partial$
C3 $(\forall a',\Delta)\ ((a' \Rightarrow \neg e[\Delta]) \in BR \supset nocc(a',l',d_{t0}\text{-}\Delta, d_t\text{-}d_{t0}+\Delta))$

t_0 is the time point where fact *e* becomes true in time line *l*; C1 expresses that persistence is predicted at least in the preferred futures; C2, that persistence lasts at least for ∂ (without prohibiting it beyond this duration), provided that no action *a'*, known in *BR* to make *e* false, occurs during the relevant lapse (C3). (We cannot use the function *inhibit* here, since we want the persistence of an event to be defined without reference to its cause, while *inhibit* defines a set of events related to both a cause and its effect).

In the same way as implications are gathered in a rule base *BR*, the known persistences are collected in a "table of persistences" *TP*.

We now have all the prerequisites necessary to investigate which set of actions can reasonably be held as causally responsible for a given event *e*.

6. Explanation

The reader should remember that the notion of action is essential in our theory, since we decided to privilege, when asked to find the causes of a state of affairs, voluntary causes — which we take to be actions executed by agents in virtue of their free will — over "natural laws".

We recall that a *choice point* is a time point from which stem (at least) two different time lines, one in which the action is performed, the other in which it is not. The free will of the agent is exactly his/her ability to choose which of these two lines will correspond to reality.

Definition 7: *t is called a* ***choice point*** *relative to action a, among lines* l_1 *and* l_2 *(noted* $pchoix(a,l_1,l_2,t)$*) iff* $(\forall t')\ ((t' < t \supset coincide(l_1,l_2,t')) \wedge v(a,l_1,d_t) \wedge v(\neg a,l_2,d_t))$

The set of causal rules *BR* and the definitions provided so far allow us to define the set of voluntary causes of an effect *e* observed on time line *l'* at time point *t'*:

Definition 8: *the voluntary causes are defined as a partial function:*

$$cause_v : E \times L \times T \rightarrow 2^A$$

defined only if $v(e,l',d_{t'})$ holds. Then, $cause_v\ (e,l',t')$ is the subset *A'* of *A* such that $a \in A'$ iff *a* satisfies conditions C1-C4 (the scope of *t* and *l* extends from C2 to C4, and of Δ includes C1 and C2):
C1 $(\exists\Delta)\ (a \Rightarrow e[\Delta] \in BR)$
C2 $(\exists t,l,l'')\ (pchoix(a,l,l'',t) \wedge d_t \le d_{t'} \le d_t + Relevant\text{-}Delay)$, where:
if $(\exists\ \partial)\ (persist(e,\ \partial) \in TP)$ **then** $Relevant\text{-}Delay = \Delta + \partial$ **else** $Relevant\text{-}Delay = \Delta$,
C3 $l' \in Lp(l,d_t)$
C4 $v(\neg e,l,d_t)$.

C1 selects the set of causal rules of *BR* containing the effect *e* in their right part (we examine in [11] the possibility of exploring what we call *closure(BR)* instead of *BR*, in order to take into account the actions which are known to cause an effect *e'*, of which *e* is a tautological consequence);

C2 means that the agent had the choice (at a time which is relevant for the observation of *e*) between doing and not doing action *a*, and that he or she chose to do *a* ...

C3 ... in a time line *l* for which time line *l'* (where event *e* has actually occurred) is among the preferred futures at the time where the choice has been made;

C4 specifies the relevance of the action to the observed event: *e* must not already be true at the moment of executing *a* in *l*.

Suppose that we have a description of the evolution of a world by means of a set *DL* of statements using the predicates *v* and *nocc*. The above definition can be used to solve the explanation problem, if we also have at our disposal general information such as *BR* and *TP*, and provided that some assumptions concerning the completeness of

DL are accepted. [11] gives further results, and extends Definition 8 to the case where an operator $\rightarrow$ for "implies in all cases" is added.

7. Related works

Recent publications [12,15] contain thorough discussions of other approaches, namely chronological minimization [6,7,13] or causal minimisation [9]. They show why such approaches fail to handle adequately prediction, explanation, or ramification problems. The aim of this section is not to restart these discussions.

However, we would like to show briefly how we tackle the central problem illustrated by the already mentionned Y.S.P. To *DL* and *BR* given in section 5, we add *TP* containing the facts that *alive* and *loaded* persist indefinitely; we add also the fact that *inhibit(¬alive, shoot)* contains facts like *deviation-of-bullet* and so on.

The conflict of persistence between *alive* and *loaded* does not arise in our approach. As a matter of fact, the assumption of completeness on *DL* enables us to derive that no inhibiting fact prevents *¬alive* to occur, once *shoot* has been done; therefore, we **predict** *¬alive*. If we also have a rule like *shoot* $\Rightarrow$ *¬loaded* in *BR*, definition 6 cannot be used to predict the persistence of *loaded*. Knowing what belongs to *inhibit(¬loaded, shoot)* — or assuming this set to be empty, in the absence of any information on the subject —, we predict *¬loaded* as well.

We now consider backward reasoning, adding to *DL* a statement such as $v(Fred\ is\ alive, l, d_{t3})$ and $d_{t2}+\Delta<d_{t3}$; if we have to **explain** this anomalous state, our approach will consider two possible tracks to follow: the persistence of *loaded* has stopped (the gun has somehow become unloaded between d_{t1} and d_{t2}) or some inhibiting effect (e.g. *deviation-of-bullet*) has occurred between the action *shoot* and its normal effect *¬alive*, that is between d_{t2} and d_{t3}. However, we cannot prefer one track over the other, nor can we guess exactly when, in the intervals defined, the anomalous fact took place. In contrast to this intuition, the chronological systems tend to prefer the sequence of world states where the gun becomes unloaded just before shooting, because this sequence postpones the change as long as it is consistent to do so.

More recent approaches [2,12,15] do not present these anomalies, but they should be augmented with the possibility of inhibiting effects after the action; otherwise, they are not able to propose the second kind of explanations.

Finally, the solution advocated for in this paper also runs into some difficulties. An important one concerns concurrent actions (two or more actions simultaneously providing the same result) and the fact that we require the actions to be instantaneous.

8. Conclusion

In this paper, we have developped:

- a simple temporal ontology,
- the role of an agent in the evolution of the world.

This approach seems to provide an intuitively correct analysis of the main problems encountered in the A.I. literature: the explanation problem, the prediction problem, and the ramification problem (see [11] for examples).

Our approach should easily extend to the case where the first-level language is first-order. We do not anticipate too many difficulties to take into account the duration of actions: instead of a three-place predicate $v(a,t,d_t)$, we might abbreviate the formula: $(\forall t)\ (t1<t<t2 \supset v(a,l,d_t))$ into $v(a,l,d_{t1},d_{t2})$, a four-place predicate. This should also help us to handle the case of concurrent actions.

As another direction of further research, we are taking advantage of a corpus of car-crash reports, which is currently studied in our Laboratory [5]. The goal is to determine what should be put in *BR* and *TP* in order to find intuitively correct answers to questions concerning the causes of the accident. As it is often the case in Artificial Intelligence, real-size problems reveal issues which are not even visible in toy problems, such as those which illustrate the present paper.

REFERENCES

[1] James ALLEN: Towards a general theory of action and time. *Artificial Intelligence* vol.23 pp.123-154, 1984

[2] A.B.BAKER: A simple solution to the Yale Shooting Problem. *Intern.Conf. on Knowledge Representation and Reasoning* pp.11-20, 1989

[3] James P.DELGRANDE: A first-order conditional logic for prototypical properties. *Artificial Intelligence* vol.33 pp.105-130, 1987

[4] Françoise GAYRAL, Philippe GRANDEMANGE: Événements : ponctualité et durativité. *8th AFCET-RFIA Congress* pp.905-910, Lyon (F), Nov. 1991

[5] Françoise GAYRAL, Philippe GRANDEMANGE, Daniel KAYSER, François LÉVY: Interprétation des constats d'accidents : représenter le réel et le potentiel *Approches sémantiques* t.a.l. vol.35 n°1 pp.65-81, 1994

[6] B.A.HAUGH: Simple causal minimization for temporal persistence and projection. *AAAI* pp.218-223, 1987

[7] Henry A.KAUTZ: The logic of persistence. *5^{th} National Conference on Artificial Intelligence* pp.401-405, 1986
[8] Saul A.KRIPKE: Semantical considerations on modal logic. *Acta philosophica fennica* vol.16 pp.83-94, 1963
[9] Vladimir LIFSHITZ: Computing Circumscription. *9^{th} IJCAI* pp.121-127, Los Angeles, 1985
[10] Drew V.McDERMOTT: A Temporal Logic for Reasoning about Processes and Plans. *Cognitive Science* vol.6 pp.101-155, 1982
[11] Aïcha MOKHTARI: Action-based causal reasoning. *Applied Intelligence* to appear
[12] Erik SANDEWALL: The range of applicability of nonmonotonic logics for the inertia problem. *13^{th} IJCAI* pp.738-743, Chambéry, 1993
[13] Yoav SHOHAM: Reasoning about change: time and causation from the standpoint of Artificial Intelligence. *M.I.T.Press* 1988
[14] Yoav SHOHAM: Time for Action : On the Relation Between Time, Knowledge and Action. *11^{th} IJCAI* pp.954-959 & 1173, Detroit, 1989
[15] Lynn A.STEIN, Leora MORGENSTERN: Motivated action theory: a formal theory of causal reasoning. *Artificial Intelligence* vol.71 pp.1-42, 1994