Proceedings, Eleventh International Conference on Principles of Knowledge Representation and Reasoning (2008)

Scenario Update Applied to Causal Reasoning

Florence Dupin de Saint-Cyr
IRIT, Université Paul Sabatier, 31062 Toulouse, France

Abstract

In this paper, we propose to define the update of a scenario (sequence of observations at different time points) by a piece of information (value of a fluent or event occurrence) at a given time point. This operation computes the possible world evolutions (called trajectories) satisfying this piece of information that are the most in accordance with the initial scenario. It enables us to identify the consequences that a modification may involve on the evolution of the world.

Updating scenarios allows us to define formally the counterfactual aspect of causation: to check if an event is a cause in a given scenario amounts to update this scenario by the non-occurrence of this event.

Introduction

The question tackled in this article is the following one: being given a scenario describing the evolution of a system by means of a series of observations (events occurrences or facts) at various moments, what would have happened if something in the scenario had been different? In many situations, this question is fundamental, since it may help to assign responsibilities, it may clarify causal relations, and enable people to distinguish variables which have a determining impact on the future from those which finally have no influence.

As far as we know, this simple question does not have been formalized in artificial intelligence. It seems close to the field of diagnosis. But diagnosis aims at detecting the reasons of the failures of a system, whereas, here, one wishes to infer what would occur if the current scenario were modified. Diagnosis uses models of behaviors and uses the notion of distance between the observed behavior and the behavior predicted by the model. This distance is used to determine which component of the system can be at the origin of a failing behavior. Scenario update also uses the concept of distance between trajectories but the principle is different and the interesting distances are not the same ones. The update of scenario consists in selecting for each trajectory compatible with the initial scenario, the trajectories compatible with a new piece of information and that are the closest to it. The most interesting distances for the update of scenarios are asymmetrical compared to the moment in which the change takes place in the scenario.

One of the most important field of application of scenario update is causal reasoning. Indeed, in order to determine if an event causes a fact, one is brought to calculate mentally what would have occurred if this event had not taken place. If this modification has no influence on this precise fact then the event is not a possible cause of it. The update of scenario can thus clarify causal relations.

In this article, we use the framework of belief extrapolation (Dupin de Saint-Cyr & Lang 2002) since it enables us to carry out the first stage of the computation. Indeed, given a scenario, belief extrapolation computes the less surprising evolutions of the world (called trajectories) compatible with this scenario. In this framework, several order relations were proposed to compare trajectories, allowing to define various extrapolation operators. The only condition imposed on these relations is inertia, i.e., trajectories in which no change occurred should be preferred. Here, we propose to extend the framework of basic belief extrapolation, in which there are no events nor static or dynamic laws, with a framework in which events can occur with a given degree of surprise. Extrapolation within a wider framework makes it possible to calculate the preferred trajectories satisfying a complex scenario (which may contain event occurrences). The second stage consists in selecting the trajectories closest to each one of them which satisfy the new piece of information.

Belief Extrapolation

Temporal Formulas and Trajectories

We consider a propositional language $\mathscr{L}$ built on a finite set of variables (or fluents) $\mathscr{V} = \{v^1, ..., v^n\}$, the connectors $\wedge$ (and), $\vee$ (or), $\neg$ (not), $\rightarrow$ (implies) and $\leftrightarrow$ (equivalent) and the Boolean constants $\top$ (tautology) and $\bot$ (contradiction). $\mathscr{M} = 2^{\mathscr{V}}$ is the set of interpretations for $\mathscr{V}$. Formulas of $\mathscr{L}$ are denoted by lowercase Greek letters (φ, ψ etc.) and interpretations are denoted by m^i, m^j etc. If $\varphi \in \mathscr{L}$ then $Mod(\varphi)$ is the set of models of φ. $\models$ is the classical logic consequence. A literal is a variable or its negation; if $l = v^i$ (resp. $\neg v^i$) then $\neg l$ is equal to $\neg v^i$ (resp. v^i). The set of literals is $LIT = \{v^1, \neg v^1, ..., v^n, \neg v^n\}$.

Let N be an integer representing the number of considered time-points. A *time-stamped propositional variable* is a propositional variable indexed by one time point. If $v \in \mathscr{V}$

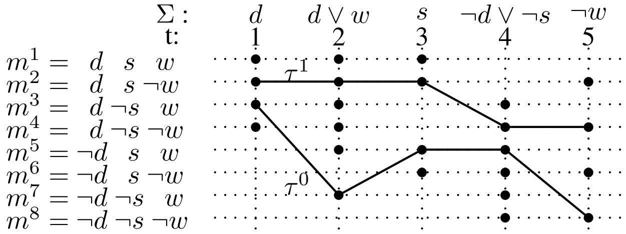
Σ :
d
$d \vee w$
s
$\neg d \vee \neg s$
$\neg w$
t:
1
2
3
4
5
$m^1 = \ d \ \ s \ \ w$
$m^2 = \ d \ \ s \ \neg w$
$m^3 = \ d \ \neg s \ \ w$
$m^4 = \ d \ \neg s \ \neg w$
$m^5 = \neg d \ \ s \ \ w$
$m^6 = \neg d \ \ s \ \neg w$
$m^7 = \neg d \ \neg s \ \ w$
$m^8 = \neg d \ \neg s \ \neg w$
τ^1
τ^0

Figure 1: Trajectories τ^0 and τ^1 of Example 1

and $t \in [\![1, N]\!]$ then the intuitive meaning of $v_{(t)}$ is that $v_{(t)}$ is true iff v is true at time-point t. $\mathscr{V}_{(N)} = \{v_{(t)} | v \in \mathscr{V}, t \in [\![1, N]\!]\}$ stands for the set of every time-stamped propositional variables (w.r.t. $\mathscr{V}$ and N). $\mathscr{L}_{(N)}$ is the language generated from $\mathscr{V}_{(N)}$ and the usual connectors. A formula of $\mathscr{L}_{(N)}$ is called *temporal formula*. The temporal formulas are denoted by uppercase Greek letter (Φ, Ψ etc.). An *instantaneous formula* is a formula whose variables are stamped by the same time-point. A *scenario* Σ is a particular temporal formula which can be written in the form of a conjunction of N instantaneous formulas $\varphi^1_{(1)} \wedge \ldots \wedge \varphi^N_{(N)}$. The formula of time-point i of the scenario Σ, is denoted by $\Sigma(i)$. $TRAJ_{(N)} = 2^{\mathscr{V}_{(N)}}$ is the set of every interpretations of $\mathscr{V}_{(N)}$, called *trajectories*. A trajectory τ can be represented by a sequence $\tau = \langle \tau(1), ..., \tau(N) \rangle$ of interpretations of $\mathscr{M}$ (i.e, $\forall i \in [\![1, N]\!]$, $\tau(i) \in \mathscr{M}$) . Finally, a trajectory τ is said *static* iff $\tau(1) = ... = \tau(N)$.

Example 1 *We consider a system which can be described by 3 variables: d, s and w. This system is an office whose door and window can be opened or closed, and whose occupant can be sitting or standing. The variables are defined by d is true if the door is opened, s is true if the occupant is sitting and w is true if the window is opened. Let us consider the scenario in which the door is opened at time point 1, the door or the window is opened at 2, the occupant is sitting at 3, the door is closed or the occupant is standing at 4, the window is closed at 5. This scenario is represented by the following formula $\Sigma = d_{(1)} \wedge (d_{(2)} \vee w_{(2)}) \wedge s_{(3)} \wedge (\neg d_{(4)} \vee \neg s_{(4)}) \wedge \neg w_{(5)}$. The set of interpretations, $\mathscr{M} = \{m^1, \ldots, m^8\}$, is described Figure 1. Many trajectories satisfy Σ (trajectories passing by the black spots of the Figure), for instance: $\tau^0 = \langle m^3, m^7, m^5, m^5, m^8 \rangle$ and $\tau^1 = \langle m^2, m^2, m^2, m^4, m^4 \rangle$.*

Standard Belief Extrapolation

Given a temporal formula Φ, *belief extrapolation* (Dupin de Saint-Cyr & Lang 2002) consists to supplement it by using persistence assumptions (this process is called *chronicle completion* in (Sandewall 1995)). The guiding idea of extrapolation is that, as long as nothing prevents it, the fluents value do not change. The goal is thus to find the best trajectories (i.e., the most static) satisfying the observations. This is why, it is necessary to have a *preference relation* on trajectories, this relation $\preceq$ should be reflexive and transitive in $TRAJ_{(N)}$ (it is not necessarily complete[1]). We use $\tau \prec \tau'$ (read "τ is strictly preferred to τ'") for ($\tau \preceq \tau'$ and not $\tau' \preceq \tau$), $\tau \sim \tau'$ for ($\tau \preceq \tau'$ and $\tau' \preceq \tau$). Let $X \subseteq TRAJ_{(N)}$, a trajectory $\tau \in X$ is *minimal* w.r.t. $\preceq$ in X iff it does not exist any trajectory τ' in X s.t. $\tau' \prec \tau$.

Definition 1 (Inertial Preference Relation)
The relation $\preceq$ on $TRAJ_{(N)}$ is inertial *iff*
- for any two static *trajectories, it holds that $\tau \sim \tau'$ and*
- for any static *trajectory τ and any* not static *trajectory τ', it holds that $\tau \prec \tau'$.*

The preference for inertia implies that if there exists a static trajectory satisfying a temporal formula Φ then every preferred trajectories satisfying Φ must be static as well. It means that if one can suppose consistently that no change occurred, completion consists in imposing it.

Definition 2 (Extrapolation Operator)
Any inertial *preference relation $\preceq$ on $TRAJ_{(N)}$ allows to induce an extrapolation operator which maps a temporal formula Φ to a formula $E_{\preceq}(\Phi)$ characterizing the preferred trajectories for $\preceq$ which satisfy Φ:*
$E_{\preceq} : \mathscr{L}_{(N)} \rightarrow \mathscr{L}_{(N)}$ such that:

$$Mod(E_{\preceq}(\Phi)) = \{\tau \mid \tau \in \min(\preceq, Mod(\Phi))\}$$

Definition 3 (Changes in a Trajectory)
The set of changes *$Ch(\tau)$ of the trajectory τ is defined by :*

$$Ch(\tau) = \left\{ \langle l, t \rangle \;\middle|\; \begin{array}{l} l \in LIT,\ t \in [\![2, N]\!], \\ \tau(t-1) \models \neg l \text{ and } \tau(t) \models l \end{array} \right\}$$

Example 1 (continued): *Note that τ^1 is more static than τ^0 w.r.t. the number of changes, denoted by $\tau^0 \preceq_{nc} \tau^1$, since $Ch(\tau^0) = \{\langle \neg d, 2 \rangle, \langle s, 3 \rangle, \langle \neg s, 5 \rangle, \langle \neg w, 5 \rangle\}$ and $Ch(\tau^1) = \{\langle \neg d, 4 \rangle\}$. It seems indeed more natural to explain the sequence of observation of the example by only one change: the door was closed at time-point 4 rather than to explain it by the fact that the door was closed at 2, the occupant sat down at 3, then raised at 5 while the window was closed. Besides, τ^1 is one of the three preferred trajectory for $\preceq_{nc}$ satisfying Σ. The two others are $\langle m^2, m^2, m^6, m^6, m^6 \rangle$ and $\langle m^2, m^2, m^2, m^6, m^6 \rangle$. Hence, $E_{\preceq_{nc}}(\Sigma) = d_{(1)} \wedge s_{(1)} \wedge \neg w_{(1)} \wedge d_{(2)} \wedge s_{(2)} \wedge \neg w_{(2)} \wedge s_{(3)} \wedge \neg w_{(3)} \wedge (d_{(4)} \oplus s_{(4)}) \wedge \neg w_{(4)} \wedge (\neg d_{(3)} \rightarrow \neg d_{(4)}) \wedge \neg w_{(5)} \wedge (d_{(4)} \leftrightarrow d_{(5)}) \wedge (s_{(4)} \leftrightarrow s_{(5)})$ where $\oplus$ stands for "exclusive or"[2].*

In (Dupin de Saint-Cyr & Lang 2002), several other preference relations on trajectories have been proposed. They all are based on the minimization of the changes. They take into account the number of literals that changed, or the costs of the changes (w.r.t. the penalties associated with the change of each literal), or the old or recent character of the changes.

[1] For example, the "change-set-inclusion" relation prefers a trajectory to another one if the changes which characterize the first one are included in those characterizing the second one. This preference relation is not complete, since, for instance, τ^0 and τ^1 of Example 1 can not be compared.

[2] $a \oplus b \equiv (a \wedge \neg b) \vee (\neg a \wedge b)$.

Event-Based Extrapolation

In this section, we develop an extension of standard belief extrapolation by introducing the concept of event. An event is an operation which induces a change in the normal course of the evolution. In the literature, events are often described by their effects and their preconditions. Here, we assume that the system evolution is described by static and dynamic laws. The encoding of these laws is out of the scope of this article, but we suppose that we two functions ke and km are available. ke quantifies the surprise degree associated with every set of events in a given situation. km measures the surprise degree attached to every transition. An infinite surprise degree means that the occurrence of the event (resp. the transition) is impossible. ke generalizes the function "Precond" (for preconditions) of STRIPS (Fikes & Nilsson 1971) which defines necessary conditions for the occurrence of one event. It is supposed to be defined for any set of events (events which took place simultaneously). This function may be computed from a representation of the dynamic and static laws and an event encoding. Here, the function km does not impose to have deterministic events as in the case of the function `Result` of the situations calculus (McCarthy & Hayes 1969). If no event occurs the function, applied to the current situation and an empty set of event, will describe the normal evolution of the system (it will give preference to the same situation if the system is inert or to a different situation if there exists temporary fluents[3]).

Definition 4 (Events) *Let Ev be a set of event symbols (denoted $\varepsilon^1, \dots, \varepsilon^P$)[4]. The two following functions are available:*
- the function ke measures the surprise degree associated with the simultaneous occurrence of a set of events in a situation: $ke : \mathscr{M} \times 2^{Ev} \rightarrow \mathbb{N} \cup \infty$
- the function km that allows to obtain a penalty distribution over the possible situations (representing the surprise degree associated with the different situations after the simultaneous occurrences of a set of events in a situation): $km : \mathscr{M} \times 2^{Ev} \times \mathscr{M} \rightarrow \mathbb{N} \cup \infty$.

The two available characteristic functions ke and km are supposed to be coherent with a given action theory. Note that this definition is under the strong assumption of a Markovian behavior of the system, i.e., the evolution of the system does not depend on its history but only on its current state. Taking into account not Markovian fluents would need a more heavy formalism and is left outside the scope of scenario update.

An "inert" system (in the classification of (Sandewall 1994)) is a system where no fluent is temporary. Thus, if no event occurs then the state of the world does not change and the occurrence of any event is surprising.

Definition 5 (Inert System) *The system is* inert *iff $\forall m, m' \in \mathscr{M}$, $\forall ev \subseteq Ev$,*

1. $km(m, \emptyset, m') = 0 \Leftrightarrow m = m'$ and

2. $ke(m, ev) = 0 \Leftrightarrow ev = \emptyset$

Now, it is possible to handle scenarios containing at the same time observations of facts and observations of event occurrences.

Definition 6 (Mixed Temporal Formula) *Let $\mathscr{V}' = \mathscr{V} \cup Ev$ and $\mathscr{V}'_{(N)} = \{v_{(t)} | (v \in \mathscr{V}$ and $t \in [\![1, N]\!])$ or $(v \in Ev,$ and $t \in [\![1, N-1]\!])\}$ its set of associated time-stamped variables. We denote by LIT' the set of literals built from $\mathscr{V}'$.* A mixed temporal formula *is built on the variables of $\mathscr{V}'_{(N)}$ with the usual connectors and constants. Let $\mathscr{L}'_{(N)}$ be the set of this formulas.*

The formula: $d_{(1)} \wedge \varepsilon^{cd}_{(1)} \wedge \neg d_{(2)}$ is an example of mixed temporal formula expressing that the door was open (d was true) at time point 1 and the event ε^{cd} has occurred (meaning for instance that "somebody has closed the door") at time point 1 and at time point 2 the door was closed.

Definition 7 (Cost of Mixed Trajectory) *A* mixed situation *s is an interpretation of $\mathscr{V}'$. Let $f(s)$ denote the facts that hold in the situation s (interpretations of $\mathscr{V}$) and $e(s)$ the events that occur in s (interpretations of Ev). A* mixed trajectory *corresponds to a truth value assignment to the variables of $\mathscr{V}'_{(N)}$. Every mixed trajectory τ can be represented by a sequence $\tau = \langle \tau(1), \dots \tau(N) \rangle$ of interpretations of $\mathscr{V}'$. Let $TRAJ'_{(N)}$ denote the set of mixed trajectories.*
A mixed trajectory is called static *if*
$$\begin{cases} \forall t \in [\![1, (N-1)]\!], & e(\tau(t)) = \emptyset \text{ and} \\ \forall t \in [\![1, N]\!], & f(\tau(1)) = \dots = f(\tau(N)). \end{cases}$$
The cost *$k(\tau)$ of a trajectory τ is:*

$$\sum_{t=1}^{N-1} ke(f(\tau(t)), e(\tau(t)) + km(f(\tau(t)), e(\tau(t)), f(\tau(t+1)))$$

The cost of a trajectory thus corresponds to the sum for each time point of the surprise degree associated to the occurrence of the events at this time point added to the surprise degree to reach the following situation being given the occurrence of these events at the previous time point. We consider the cost of the events which occurred between time point 1 and N-1 without taking account of those which occurred at the last moment since they have no impact. Let us note that, here, the surprise degree associated with the occurrence of an event or the surprise degree associated with obtaining a given situation are quantified in terms of penalties. Besides, the use of penalties to characterize surprise degrees has been proposed initially by Sandewall in (Sandewall 1994). This choice is justified by the intrinsically additive and compensatory character of surprises, but the use of other measures like probabilities or possibilities is completely possible. One can refer to (Dupin de Saint-Cyr, Lang, & Schiex 1994) for a study of the links between these various measures.

In an inert system, a static trajectory is not surprising, this is expressed by the following property.

Proposition 1 *If the system is inert then*

$$\forall \tau \in TRAJ'_{(N)}, \tau \text{ is static} \Leftrightarrow k(\tau) = 0.$$

[3] Temporary fluents (also called "dynamic" in (Sandewall 1995)) are not persistent fluents, they correspond to variables whose values are not naturally constant during time.

[4] More precisely, the symbol ε^i does not denote the event itself but its occurrence.

Proof: Using Definition 7, τ is static $\Leftrightarrow$
$$\begin{cases} \forall t \in [\![1,(N-1)]\!], & e(\tau(t)) = \emptyset \text{ and} \\ \forall t \in [\![1,N]\!], & f(\tau(1)) = ... = f(\tau(N)). \end{cases}$$

- given a static trajectory τ, let $m = f(\tau(t))$, the cost of τ is $k(\tau) = \sum_{t=1}^{N-1} ke(m,\emptyset) + km(m,\emptyset,m)$. Using Definition 5, the system being inert, we get $k(\tau) = 0$.
- Conversely, given a trajectory τ such that $k(\tau) = 0$, we get $\sum_{t=1}^{N-1} ke(f(\tau(t)), e(\tau(t)) + km(f(\tau(t)), e(\tau(t)), f(\tau(t+1))) = 0$. Now, since the penalties are positive. We obtain that each term of the sum should be equal to zero, hence every event occurring in τ is not surprising: $\forall t \in [\![1,N-1]\!], ke(f(\tau(t)), e(\tau(t))) = 0$, since the system is inert it means that $\forall t \in [\![1,N-1]\!], e(\tau(t))) = \emptyset$. It means that no event occurs in τ. Using Definition 5 with $\forall t \in [\![1,N-1]\!], km(f(\tau(t)), \emptyset, f(\tau(t+1))) = 0$, we get that $\forall t \in [\![1,N-1]\!], f(\tau(t)) = f(\tau(t+1))$. Hence τ is static.

□

Belief extrapolation extended to mixed formulas is defined by means of a preference relation on trajectories which minimizes their cost, i.e., which minimizes the surprise degree associated with the events of these trajectories.

Definition 8 *Given a mixed temporal formula $\Phi \in \mathscr{L}'_{(N)}$, the extended extrapolation of Φ is defined by:*
$EE : \mathscr{L}'_{(N)} \to \mathscr{L}'_{(N)}$ *such that:*

$$Mod(EE(\Phi)) = \{\tau \mid \tau \in \min(k, Mod(\Phi))\}$$

Example 2 *Let $\mathscr{V}' = \{d, r, \varepsilon^{cd}, \varepsilon^{b}, \varepsilon^{rd}\}$ where d is a variable representing the fact that the door is opened, the variable r means that the door is red, ε^{cd} is an event meaning that "somebody is closing the door", ε^{b} means "the wind is blowing", ε^{rd} means "someone is painting the door in red". Let us consider the scenario $\Sigma = d_{(1)} \wedge \neg r_{(1)} \wedge \neg d_{(3)}$. Computing the extrapolation of this scenario can be done as follows: the facts that are true at time point 1 are completely known since d and $\neg r$ are true, let m^2 such a situation (see Figure 2). We suppose that the surprise degree associated to the events are known, with for instance $0 < ke(m^2, \{\varepsilon^{cd}\}) = ke(m^2, \{\varepsilon^{b}\}) < ke(m^2, \{\varepsilon^{cd}, \varepsilon^{b}\}) < ke(m^2, \{\varepsilon^{rd}\}) < ke(m^2, \{\varepsilon^{cd}, \varepsilon^{rd}\}) = ke(m^2, \{\varepsilon^{w}, \varepsilon^{rd}\}) = ke(m^2, \{\varepsilon^{cd}, \varepsilon^{b}, \varepsilon^{rd}\}) = \infty$. These constraints mean that in the initial state m^2, the fact that someone closes the door is as little surprising as a strong gale but less surprising as the simultaneous occurrence of these two events, and even less surprising than the fact that someone paints the door (regarded as rare). In addition, it is impossible to paint the door during a strong gale or while it is being closed.*

Let us suppose now that we have a description of the events allowing to deduce that when someone closes the door in the state m the states of the system in which the door is closed are less surprising, etc. Then, among the $2^3 \times 2^5 \times 2^2$ possible trajectories satisfying this scenario, four trajectories are preferred: those where only one event occurs (either the agent closes the door, or the wind blows at time point 1 or 2). In these four trajectories r remains false in 2 and 3. So, $EE(\Sigma) \models (((\varepsilon^{cd}_{(1)} \oplus \varepsilon^{b}_{(1)}) \wedge \neg d_{(2)}) \vee (d_{(2)} \wedge (\varepsilon^{cd}_{(2)} \oplus \varepsilon^{b}_{(2)}))) \wedge \neg r_{(2)} \wedge \neg r_{(3)}$.

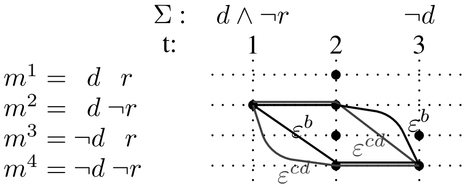

Figure 2: Mixed Trajectories for Example 2

Proposition 2 *If the system is inert then an extended extrapolation operator EE satisfies the extended postulates of belief extrapolation $E1, E2' \dots E6$*[5]
E is a belief extrapolation operator $\Rightarrow$ E satisfies :
E1 : $E(\Phi) \models \Phi$
E2' : if $PERS' \wedge \Phi$ is consistent then $E(\Phi) \equiv PERS' \wedge \Phi$
E3 : if Φ is consistent then $E(\Phi)$ is consistent
E4 : if $\Phi \equiv \Phi'$ then $E(\Phi) \equiv E(\Phi')$
E5 : $E(\Phi) \wedge \Phi' \models E(\Phi \wedge \Phi')$
E6 : if $E(\Phi) \wedge \Phi'$ is consistent then $E(\Phi \wedge \Phi') \models E(\Phi) \wedge \Phi'$

with $PERS' = \bigwedge_{t \in [\![1,N-1]\!]} \left(\bigwedge_{v \in \mathscr{V}} v_{(t)} \leftrightarrow v_{(t+1)} \wedge \bigwedge_{\varepsilon \in Ev} \neg \varepsilon_{(t)} \right)$.

Proof: If the system is inert, then we can prove, using the same principles as in the proof for the representation theorem of classical extrapolation (Dupin de Saint-Cyr & Lang 2008), that for any extrapolation operator EE there exists a revision operator $\star$ on $\mathscr{L}'_{(N)}$ such that for every $\Phi \in \mathscr{L}'_{(N)}$, $EE(\Phi) \equiv PERS' \star \Phi$. Conversely, for any revision operator $\star$, there exists a unique extrapolation operator EE such that for every $\Phi \in \mathscr{L}'_{(N)}$, $EE(\Phi) \equiv PERS' \star \Phi$. This result is based on the fact that k respects inertia (Proposition 1). The postulates $E1$, $E2'$, $E3$, ... $E6$ correspond to Katsuno and Mendelzon postulates for belief revision (Katsuno & Mendelzon 1991) in which the formula to be revised has been changed into $PERS'$. □

As evoked in the proof, in the case of an inert system traditional extrapolation amounts revising the knowledge that the values of the fluents persist by the temporal formula to extrapolate. A similar result holds for extended extrapolation applied to any system (not necessarily inert). This time, it is the formulas describing the natural evolution of the system (static and dynamic laws) that must be revised by the formula to extrapolate.

Updating Mixed Temporal Formulas

The Question of "What would have occurred if..."

The question tackled in this paragraph is: being given a factual story and a knowledge of the dynamics of the system, what would occur if someone imposes a change in this story? Within our formal framework, the story is a sequence of observations (events or facts), i.e., a mixed scenario or more generally, a mixed temporal formula. The knowledge

[5] Here $E2'$ replace the postulate $E2$ of standard belief extrapolation since $PERS'$ involve event occurrence where as the initial postulate did not, more precisely $E2$ refers to the formula $PERS = \bigwedge_{v \in \mathscr{V}, t \in [\![1,N-1]\!]} (v_{(t)} \leftrightarrow v_{(t+1)})$.

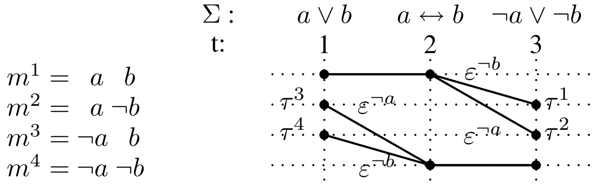

Figure 3: Spy Story

of the dynamics of the system corresponds to laws of evolution of the world summarized by the two cost functions ke and km.

Let us examine a toy spy-story example:

Example 3 *Two agents are in contact with me, they share their life between Toulouse and London. I received a post-card from London but I could not read the signature, it was one of them but I do not know who exactly. Hence* a *or* b *was in London at time point 1:* $a_{(1)} \vee b_{(1)}$ *where* $a_{(t)}$ *(resp.* $b_{(t)}$*) denotes the fact that the first (resp. second) agent is in London at time point* t*. I learnt that the day after they met each other (since they have exchanged secret documents), hence either they were both in London that day or either they both were elsewhere:* $a_{(2)} \leftrightarrow b_{(2)}$*. I know that one of them was seen in Toulouse two days after but I do not know who it was:* $\neg a_{(3)} \vee \neg b_{(3)}$*. I know that they prefer not to travel together in order to avoid suspicions. Hence, I can extrapolate four possibilities (considering only the less surprising ones, see Figure 3): either they were both in London, they met there and then one of them left (it makes two possibilities* τ^1 *and* τ^2 *according to the identity of the agent who left), either only one of them was in London, he left London and they met in Toulouse (it also gives two possibilities* τ^3 *and* τ^4*).* $Traj(EE(\Sigma)) = \{\tau^1, \tau^2, \tau^3, \tau^4\}$*. If we encode the event of living London to Toulouse by* $\varepsilon^{\neg a}$ *for the first agent (and* $\varepsilon^{\neg b}$*) for the second agent, we obtain* $EE(\Sigma) \models (\varepsilon^{\neg a}_{(1)} \oplus \varepsilon^{\neg b}_{(1)}) \oplus (\varepsilon^{\neg a}_{(2)} \oplus \varepsilon^{\neg b}_{(2)})$

Now, there are two interesting questions:

1. *what can I conclude if I learn that Gatwick airport in London was closed (because of a strike of the controllers) at time point 1, preventing any flight between London and Toulouse between time point 1 and 2:*$\neg\varepsilon^{\neg a}_{(1)} \wedge \neg\varepsilon^{\neg b}_{(1)}$.
2. *what would have happened if the police had set up a security check in Gatwick airport? (and would have been able to arrest any suspect agent, hence forbidding them to fly):* $\neg\varepsilon^{\neg a}_{(1)} \wedge \neg\varepsilon^{\neg b}_{(1)}$

While the two questions are expressed by the same formula they do not imply the same reasoning. The first question implies that I should take into account a new piece of information about this story, I should complete my knowledge, it means that among all possible trajectories, I am able to be more selective and pick the ones that are in agreement with the new piece of information, or if no trajectory is in agreement, I should correct minimally my beliefs in order to obtain a possible trajectory. In our case, I will conclude that they were both in London and they left after their meeting.

The second question implies that I should make an hypothetical reasoning about what the story becomes if something change, hence I should consider every possibility and see how it may evolve. Hence the four initial trajectories should be reviewed, for τ^1 *and* τ^2 *the hypothesis does not change anything, if the two agents where in London and met there then they were not at the airport when the police check point could have been set up. While for* τ^3 *and* τ^4 *corresponding to the trajectories in which an agent should leave London at time point 1, the consequences of such a supposition could be more dramatic, since the agent could have been arrested in London and then he could not have exchange information with its colleague and so on...*

In order to analyze the two different kinds of reasoning used in this example, let us recall that *belief revision* consists in integrating a new piece of information into old beliefs while trying to preserve initial information as much as possible. In practice, that amounts calculating the set of interpretations satisfying new information closest to the initial beliefs. *Belief update* is based on another technique (one can refer to (Katsuno & Mendelzon 1991) for a discussion on the difference between revision and update) because it relates to the arrival of a piece of information which characterizes an evolution of the world. In that case, what was believed before is not questioned but must be modified to take into account the fact that the world evolved in accordance with this new piece of information. In practice, that amounts calculating for each possible initial world, what is the most natural evolution which leads to a world where the new piece of information is true.

In this toy example, we can see that the first reasoning is a belief revision while in the second case the reasoning is an update. A main claim of this article is that the question of "what would have happened if ..." is an *update operation*. In our problem, we want to know what *would have occurred* if something - let us call it φ - had been true at time t in a given story. That does not means that this temporal information $\varphi_{(t)}$ was *actually true*. The difference is due to the fact that:

- in the case we learn that $\varphi_{(t)}$ was true, the history must be corrected in order that φ holds at time point t. One must thus find the trajectories satisfying the new piece of temporal information $\varphi_{(t)}$ closest to the initial scenario: it is a revision.
- in the other case, one wishes to calculate for each trajectory initially compatible with the history, what it would become if one forced φ to hold at time point t: it is thus typically a belief update operation.

Note that the update of this example has been chosen with a formula containing only occurrences of events but we could have chosen to update by a fact, for instance we could have wonder what would have occurred if the first agent had been in Toulouse at time 2: $\neg a_{(2)}$. It would mean that if we had initially supposed that he was in London at time 1 (according to trajectories τ^1, τ^2 and τ^3) we should either suppose that he has flown or that he was not there at time 1. If we examine each initial possibility: for trajectories τ^3 and τ^4 the story could have been the same. Meanwhile for trajectories τ^1 and τ^2 the new hypothesis has a stronger influence (a solution is proposed below).

Update operators and Preorderings

To define in practice the update of mixed temporal formulas, we need to define a preference relation between trajectories w.r.t. any initial given trajectory. This comes from the representation theorem of (Katsuno & Mendelzon 1991) linking any update operator to the existence of a faithful assignment which maps each interpretation m to a pre-order $\leq_m$. The assignment is called "faithful" if the pre-order associated with m always strictly prefers m to any other interpretation. Note that it has been shown in (Dubois, Dupin de Saint-Cyr, & Prade 1995) that a larger class of update operator maybe defined without imposing "faithfulness" and allowing for the existence of unreachable states. A representation theorem links this broader class of update characterized by a minimal and complete set of postulates to the existence of an underlying pre-order. These critics about Katsuno and Mendelzon postulates have also been done by other authors see e.g. (Boutilier 1994).

Many preorderings can be proposed to compare two trajectories with respect to a same third. Following the schema of classical extrapolation, we propose first to define a change set associated to two trajectories by the set of literals that differ in the two trajectories associated with their change time point:

Definition 9 (Changes Between two Trajectories) *The* set of changes $Ch(\tau, \tau')$ *between trajectory* τ *and* τ' *is:*

$$Ch(\tau, \tau') = \left\{ \langle l, t \rangle \,\middle|\, \begin{array}{c} l \in LIT', t \in [\![1, N]\!] \\ \tau(t) \models l, \tau'(t) \models \neg l \end{array} \right\}$$

We also define $Ch(\tau, \tau')_{(t)} = \{l \mid \langle l, t \rangle \in Ch(\tau, \tau')\}$

Then we can adapt every change-based preorderings defined for classical extrapolation to obtain preorderings w.r.t. a given trajectory:

number of change:

$$\tau' \leq^{nc}_{\tau} \tau'' \text{ iff } \mid Ch(\tau, \tau') \mid \quad \leq \quad \mid Ch(\tau, \tau'') \mid$$

change set inclusion:

$$\tau' \leq^{csi}_{\tau} \tau'' \text{ iff } Ch(\tau, \tau') \subseteq Ch(\tau, \tau'')$$

penalty of changing events:

$$\tau' \leq^{k}_{\tau} \tau'' \text{ iff } kev(\tau, \tau') \leq kev(\tau, \tau'')$$

where $kev(\tau^1, \tau^2)$ is the sum of the surprise degrees associated to events occurrence in τ^1 when these events do not occur in τ^2 added to the symmetric sum for τ^2 w.r.t. τ^1:

$$kev(\tau^1, \tau^2) = \left\{ \begin{array}{c} \sum_{\varepsilon \in Ev, \langle \varepsilon, t \rangle \in Ch(\tau^1, \tau^2)} ke(\tau^1(t), \varepsilon) \\ + \\ \sum_{\varepsilon \in Ev, \langle \neg\varepsilon, t \rangle \in Ch(\tau^1, \tau^2)} ke(\tau^2(t), \varepsilon) \end{array} \right.$$

chronological minimization: $\tau' \leq^{csi}_{\tau} \tau''$ iff

$$\left\{ \begin{array}{l} N = 1 \\ \text{or } Ch(\tau, \tau', 1) \subset Ch(\tau, \tau'', 1) \\ \text{or} \left\{ \begin{array}{l} Ch(\tau, \tau', 1) = Ch(\tau, \tau'', 1) \text{ and} \\ \langle \tau'(2), \ldots \tau'(N) \rangle \leq^{chr}_{\langle \tau(2), \ldots \tau(N) \rangle} \langle \tau''(2), \ldots \tau''(N) \rangle \end{array} \right. \end{array} \right.$$

In this preference relation, priority is given to old similarities as in the chronological minimization of (Shoham 1988).

All these proposals are valid in order to compare trajectories with respect to a reference trajectory, not that the preference relation $\leq^{csi}$ is included in the three others (it is less discriminating). Two of them are complete ($\leq^{nc}$ and $\leq^{k}$) while the two others are partial. In the following, we define a fifth relation which seems to us better adapted to the hypothetical reasoning. In principle, we propose that the trajectories that are identical to the initial trajectory w.r.t. events occurrences should be preferred. However, it seems important to check that not only events but also facts should be identical since if there are non-deterministic events we suppose that they should give the same result, and this, before the change time point. This last supposition seems less natural after this point. Indeed in the spy story example, we can notice that the facts that were believed to hold after the change time point are no longer necessarily true. However, it seems natural to have trajectories that are closest as possible concerning occurrence of events until the end point. In summary, we choose to minimize chronologically the event distances on the entire trajectory and minimize chronologically the fact distances only until the change time point. Before defining the preordering, we need to introduce chronological event-distance preference relation:

Definition 10 (Chronological Event-distance Preference) *Given two mixed trajectories* τ *and* τ'*, let* $kev(\tau, \tau')_{(t)}$ *be the surprise degree associated to distinct events in* τ *and* τ' *at time point* t*:*

$$kev(\tau, \tau')_{(t)} = \left\{ \begin{array}{c} \sum_{\varepsilon \in Ev, \varepsilon \in Ch(\tau, \tau')_{(t)}} ke(\tau(t), \varepsilon) \\ + \\ \sum_{\varepsilon \in Ev, \neg\varepsilon \in Ch(\tau, \tau')_{(t)}} ke(\tau'(t), \varepsilon) \end{array} \right.$$

Now, given three mixed trajectories τ*,* τ' *and* τ''*,* τ' *is closer to* τ *for event-distance chronological preference than* τ'' *denoted* $\tau' \leq^{chre}_{\tau} \tau''$ *iff*

$$\left\{ \begin{array}{l} N = 1 \\ \textit{or } kev(\tau, \tau')_{(1)} < kev(\tau, \tau'')_{(1)} \\ \textit{or} \left\{ \begin{array}{l} kev(\tau, \tau')_{(1)} = kev(\tau, \tau'')_{(1)} \textit{ and} \\ \langle \tau'(2) \ldots \tau'(N) \rangle \leq^{chre}_{\langle \tau(2) \ldots \tau(N) \rangle} \langle \tau''(2) \ldots tau''(N) \rangle \end{array} \right. \end{array} \right.$$

Example 3 (continued): *If we compare the trajectories* τ^1*,* τ^2*,* τ^3*,* τ^4 *and the trajectory* $\tau^5 = \langle (m^2, \emptyset), (m^2, \{\varepsilon^{\neg a}\}), (m^4, \emptyset) \rangle$ *with respect to their closeness to* τ^2*, we get* $\tau^2 =^{chre}_{\tau^2} \tau^5 <^{chre}_{\tau^2} \tau^1 <^{chre}_{\tau^2} \tau^3$*. Considering* τ^4 *we know that* τ^4 *will be less preferred than* τ^1*, the relation between* τ^4 *and* τ^3 *depends on the surprise degree associated to the events* $\varepsilon^{\neg a}$ *and* $\varepsilon^{\neg b}$*.*

Now, we need to define fact-distance chronological minimization before a given time-point.

Definition 11 *Given three mixed trajectories* τ*,* τ' *and* τ'' *and a time point* t*,* τ' *is closer to* τ *before* t *for fact-distance chronological minimization than* τ'' *denoted* $\tau' \leq^{chrf}_{\tau,t} \tau''$ *iff*

$$\left\{ \begin{array}{l} N = 1 \textit{ or } \ t = 0 \\ \textit{or } |Chf(\tau, \tau')_{(1)}| < |Chf(\tau, \tau'')_{(1)}| \\ \textit{or} \left\{ \begin{array}{l} |Chf(\tau, \tau')_{(1)}| = |Chf(\tau, \tau'')_{(1)}| \textit{ and} \\ \langle \tau'(2) \ldots \tau'(N) \rangle \leq^{chrf}_{\langle \tau(2) \ldots \tau(N) \rangle, t-1} \langle \tau''(2) \ldots tau''(N) \rangle \end{array} \right. \end{array} \right.$$

where $Chf(\tau, \tau')_{(t)}$ *is the set of literal that are not representing event occurrences and that differs from* τ *to* τ' *at time* t*:* $Chf(\tau, \tau')_{(t)} = \{l \in Ch(\tau, \tau')_{(t)} \cap LIT\}$

Example 3 (continued): *Considering τ^1, τ^2 and $\tau^6 = \langle (m^3, \{\varepsilon^a\}), (m^1, \emptyset), (m^1, \emptyset)\rangle$, we get $\tau^2 =^{chrf}_{\tau^2,2} \tau^1 <^{chrf}_{\tau^2,2} \tau^6$. Since, τ^1 differs from τ^2 after time point 2, while τ^6 differs at the first time point. Note that τ^6 is equal to a trajectory starting with m^2 while it will be preferred to any trajectory starting with m^4.*

Now, we are in position to define a faithful assignment:

Definition 12 (Chronological Closeness Assignment) *For any time point $p \in [\![1, N]\!]$, given a trajectory $\tau \in TRAJ'_{(N)}$, let $\preceq^p$ be a family of preference relations defined by:*

$$\forall \tau', \tau'' \in TRAJ'_{(N)}, \tau' \preceq^p_\tau \tau'' \text{ iff}$$

$$\begin{cases} \tau' <^{chre}_\tau \tau'' \\ \text{or } \tau' =^{chre}_\tau \tau'' \text{ and } \tau' \leq^{chrf}_\tau \tau'' \end{cases}$$

Proposition 3 *For any time point $p \in [\![1, N]\!]$, the chronological closeness assignment is* faithful*, i.e., for any trajectories τ and τ', we have $\tau \preceq^p_\tau \tau'$.*

Proof: We have $\forall t \in [\![1, N]\!]$, $kev(\tau, \tau)_{(t)} = 0$ hence for any τ', $\tau \leq^{chre}_\tau \tau'$, if $\tau =^{chre}_\tau \tau'$ then for any time point t, $\emptyset = fd(\tau, \tau)_{(t)} \subseteq fd(\tau, \tau')_{(t)}$. Hence, $\tau \preceq^p_\tau \tau'$. □

This property implies that any trajectory is closer to itself than to other trajectories. Note that this is not a strict preference, i.e., the relation suggested is not strictly faithful in the sense of Katsuno and Mendelzon, some trajectories may have the same sequence of events and differ only in the sequence of situations after time point p but be as preferred as the reference trajectory.

Definition 13 (Updating mixed temporal formula) *Given a mixed temporal formula Φ, the update of Φ by punctual information φ at time point t by the update operator $\diamond_t$, based on $\preceq^t$, is defined by:* $Mod(\Phi \diamond_t \varphi_{(t)}) =$

$$\bigcup_{\tau \in EE(\Phi)} \{\tau' \in TRAJ'_{(N)}, \tau' \in \min_{\preceq^t_\tau}\{\tau \models \varphi_{(t)}, k(\tau) \neq \infty\}\}$$

Intuitively, to update a mixed temporal formula Φ by an instantaneous mixed formula $\varphi_{(t)}$ consists in calculating for each preferred trajectory (i.e., less surprising) τ satisfying Φ, the possible trajectories satisfying $\varphi_{(t)}$ that are closest to τ. The result corresponds to the union of the trajectories obtained for each initial trajectory.

Example 3 (continued): *If we want to know what would have occurred if the first agent had been in Toulouse at time point 2 then it is necessary to calculate $\Sigma \diamond_2 \neg a_{(2)}$. According to Definition 13, it requires to compute the possible trajectories in which $\neg a_{(2)}$ holds; then select in this set the trajectories closest to each of the 4 initial trajectories drawn on Figure 3.*

In this example, we suppose that the definitions of the surprise degrees associated with the transitions of the system allow to obtain 32 possible trajectories satisfying $\neg a_{(2)}$: 16 trajectories passing by $\neg a_{(2)} b_{(2)}$ and 16 by $\neg a_{(2)} \neg b_{(2)}$ see Figure 4. Note that τ^3 and τ^4 belong to this set. τ^3 is strictly closer to itself than the 31 other trajectories. It is the same for τ^4. In τ_1 there was only one event occurring at time 2, namely $\varepsilon^{\neg b}$. We can found one trajectory having the same curses of events, namely: $\tau^5 = \langle (m^3, \emptyset), (m^3, \{\varepsilon^{\neg b}\}), (m^4, \emptyset)\rangle$. This trajectory is the closest to τ^1 w.r.t. chronological closeness among all the 32 possible trajectories. Concerning τ^2, there is no trajectory identical for event occurrences, but two trajectories have only one difference with it: namely the two static trajectories $\tau^6 = \langle (m^3, \emptyset), (m^3, \emptyset), (m^3, \emptyset)\rangle$ and $\tau^7 = \langle (m^4, \emptyset), (m^4, \emptyset), (m^4, \emptyset)\rangle$. But according to Definition 12, τ^6 is closer since its situation at time point 1 less differs from τ^2 than τ^7. Finally, we obtain exactly 4 trajectories satisfying $\Sigma \diamond_{(2)} \neg a_{(2)}$: τ^3, τ^4,τ^5 and τ^6. In each trajectory the first agent stays in Toulouse at time 3. In this hypothetical reasoning, one cannot conclude on the fact that the two agents could meet or not.

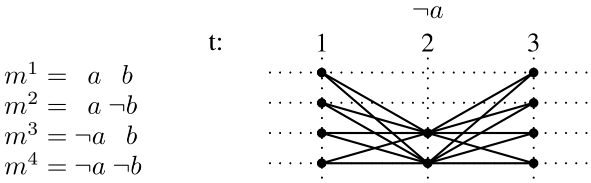

Figure 4: Trajectories satisfying $\neg a_{(2)}$

Let us note that we gave an example of update by a fact, but it is completely possible to carry out updates with occurrences of event.

Proposition 4 *The operator $\diamond_t$ based on $\preceq^t$ satisfies postulates U1, U3bis, U4, U5, U8, U9 and U10:*
U1 *$(K \diamond \varphi)$ implies φ.*
U3bis *$K \diamond \varphi$ consistent implies $K \diamond (\varphi \vee \psi)$ consistent.*
U4 *If $K1 \leftrightarrow K2$ and $\varphi^1 \leftrightarrow \varphi^2$ then $(K1 \diamond \varphi^1) \leftrightarrow (K2 \diamond \varphi^2)$.*
U5 *$(K \diamond \varphi) \wedge \psi$ implies $(K \diamond (\varphi \wedge \psi))$.*
U8 *$(K1 \vee K2) \diamond \varphi \leftrightarrow (K1 \diamond \varphi) \vee (K2 \diamond \varphi)$.*
U9 *If K is deductively closed and $(K \diamond \varphi^1) \wedge \varphi^2$ is consistent then $(K \diamond (\varphi^1 \wedge \varphi^2))$ implies $((K \diamond \varphi^1) \wedge \varphi^2))$.*
U10 *If φ is impossible then $K \diamond \varphi$ is inconsistent.*

Proof: Using the representation theorem proposed by (Dubois, Dupin de Saint-Cyr, & Prade 1995), $\forall t \in [\![1, N]\!]$, chronological closeness assignment is a mapping associating to each interpretation τ of $TRAJ_{(N)}$ a complete preordering $\preceq^t_\tau$ such that $Mod(\Phi \diamond_t \varphi_{(t)}) = \bigcup_{\tau \in EE(\Phi)}\{\tau' \in TRAJ'_{(N)}, \tau' \in \min_{\preceq^t_\tau}\{\tau \models \varphi_{(t)}, \tau \text{ is not impossible }\}\}$ where the impossible trajectories are those that have an infinite surprise degree. □

Last proposition shows that $\diamond_t$ is an update operator in the sense of (Dubois, Dupin de Saint-Cyr, & Prade 1995)[6]. Note that we have chosen to propose a more general update operator than in Katsuno and Mendelzon framework since we want to allow for impossible models (called trajectories in our context) and also we want to allow not strict faithfulness. This is why the postulates U2 and U3 of Katsuno Mendelzon have not been considered here. U3 imposes that updates are always possible: **U3** If K and φ are consistent

[6] In the proposal of (Dubois, Dupin de Saint-Cyr, & Prade 1995), the postulate U1 is not necessary since it can be deduce from U3bis, U5 and U10. The set U3bis, U4, U5, U8, U9 and U10 is minimal and complete with respect to the representation theorem.

then $(K \diamond \varphi)$ is consistent. U2 has been often discussed in the literature, it is well known to impose inert update when it is possible: **U2** If K implies φ then $(K \diamond \varphi)$ is equivalent to K. We do not want that U2 necessarily hold because we want to allow not strictly faithful assignments. Strict faithfulness, in this context, is not always desirable (see Example 4) except in the particular case of a deterministic system. Let us note however that it is completely possible to particularize the preference relation suggested in order to make it faithful.

Example 4 *Let us consider a scenario Σ in which a die has been rolled on a green carpet at time point 1 and its face is a six at time point 2. Σ does imply that the "carpet is green" but if we update Σ by the fact (that we already know) that "the carpet is green" denoted by cg, we do not necessarily obtain the same result: $\Sigma \diamond_2 cg_{(2)}$ is not equivalent to Σ, since, if the encoding of ke and km represent the transition of a fair die (hence non deterministic), it gives 6 possible trajectories.*

Note that Proposition 4 holds for update operators based on the complete pre-orderings $\leq^{nc}$ and $\leq^{k}$, meanwhile for the partial pre-orderings, the postulates $U6$ and $U7$ should be used instead of $U9$.

Causality and Scenario Update

The problem of causality was explored under many different points of view (one can refer to (Scheines 2004; Demolombe 2000) for an outline of the various approaches of causality). One point of view is to try to establish generic causal relations (it is called "causal generalization"), for example, it allows to affirm that "smoking causes cancer". One can be interested to discover the cause of a given result in a concrete example (it is called "event causation"), for example "the fact that Titanic hit an iceberg is the cause of its loss". One can also be interested in axiomatizing causality, i.e., characterizing the properties of the relation of causality: transitivity, asymmetry, decomposability, etc. Besides, axiomatizing can be studied within the framework of generic or concrete causation. A last aspect is the investigation into the "perceived cause", this problem consists in seeking the cause that would select an intelligent agent among all the possible causes explaining a scenario.

The numerous works in the field of causality lead to obtain several definitions of causality. One of the first definition was given by (Lewis 1973), which introduces "counter-factuality". This definition is based on the existence of possible worlds and on a similarity distance between worlds. A causes B in a counter-factual way when one can affirm that if A had not occurred in a world as close as possible of the current world then B would not have occurred. Other definitions were given in terms of probabilities by (Cartwright 2002), however, it is admitted that causality is related to probabilities but can not be only defined by probabilistic dependencies (mainly because of the need of asymmetry). Another type of definition uses manipulability theory, in particular (Von Wright 1976) who formalizes the idea that causality is related to the concept of intervention. A causes B if while forcing A to be true, one forces B to be true and by doing nothing to change the value of A then it does not change the value of B. In their structural model approach, (Halpern & Pearl 2001) have also proposed a counter-factual definition of causality encoded by causal graphs in which endogenous variables are distinguished from exogenous variables. A first-order modeling of this approach have been provided by (Finzi & Lukasiewicz 2003). A last definition based on a non-monotonic inference relation was given by (Bonnefon *et al.* 2006): given a series of observations $\neg B_{(t)}$, $A_{(t)}$, $B_{(t+k)}$, a context C and a relation of non-monotonic consequence $\mid\!\sim$, if the agent believes that $C \mid\!\sim B$, and that $C \wedge A \mid\!\sim \neg B$ then the agent will perceive A as being the cause of B in the context C. This approach requires to define a "context" which seems similar to what Halpern and Pearl call "exogenous variables" but this implies to already know what variables or formulas can not be considered as causes.

Here, we are interested in the research of a concrete cause ("event causation"). Our definition uses the fact that a cause must be counter-factual (if it had not been there, the conclusion would not have been obtained) and the assumption of manipulability. Thus, we will regard as possible causes only the values of the facts at the initial time point (allowed to be modified) and the events. The values of the facts at various moments are not considered as possible causes. The counter-factual nature of the cause will be used in the following way: A causes B if A and B are true in the initial mixed temporal formula, and if B is false after the update of this initial formula by $\neg A$. In short, we propose the following definition:

Definition 14 (cause)
Given a mixed temporal formula $\Phi \in \mathscr{L}'_{(N)}$,

$$A_{(t)} \textit{ causes } B_{(N)} \textit{ iff } \begin{cases} t = 1 \textit{ or } A \in \mathscr{L}_{Ev} \\ EE(\Phi) \models A_{(t)} \wedge B_{(N)} \\ \Phi \diamond_t \neg A_{(t)} \models \neg B_{(N)} \end{cases}$$

where $\mathscr{L}_{Ev}$ is the set of formulas built on Ev.

The following example come from a database of traffic accident reports submitted by drivers to insurance companies.

Example 5 (Car accident) *"The car B slipped and was found perpendicular to the road. I was coming behind it and, because of the glaze, I could not stop. My vehicle struck the vehicle B and my vehicle was projected in the ditch."*

This story can be modelled by the following scenario, $\Sigma = \varepsilon^{Bslips}_{(1)} \wedge Bperp_{(2)} \wedge AfB_{(2)} \wedge glaze_{(2)} \wedge \neg\varepsilon^{Astops}_{(2)} \wedge acc_{(3)}$. We considered the following facts $Bperp$, AfB, $glaze$, acc, $dABok$ meaning respectively B is perpendicular to the road, A is following B, there is glaze, an accident has occurred, the distance between A and B is adapted to the context; and the following events ε^{Bslips}, ε^{acc}, ε^{Astops} meaning respectively B is slipping, an accident is occurring, A is stopping. We assume that $glaze$, AfB, $\neg acc$, $\neg dABok$, $Bperp$ and $\neg Bperp$ are persistent by default (but they may change if some event occurs). For instance, if the event ε^{Bslips} occurs then it is normal that the fluent $Bperp$ change its value. When there is glaze, the event ε^{Astop} can occur only if $dABok$ is true. We suppose also that there is a rule that imposes to stop if possible when there is an obstacle. With this encoding, we obtain *one preferred trajectory corresponding to the scenario:* $\langle (\neg Bperp\ AfB\ glaze\ \neg acc\ \neg dABok, \{\varepsilon^{Bslips}\}), (Bperp\ AfB\ glaze\ \neg acc\ \neg dABok, \{\varepsilon^{acc}\}), (Bperp\ \neg AfB\ glaze\ acc\ \neg dABok, \emptyset) \rangle$. *It is a trajectory in which the distance was not adapted to glaze, since if it had been adapted the driver could have stopped (i.e.,* ε^{Astop} *would have occurred instead of* ε^{acc}*).*

Now for instance, if we want to know if $\varepsilon_{(1)}^{Bslips}$ *was a cause of the accident, we first check that* $EE(\Sigma) \models acc_{(3)} \wedge \varepsilon_{(1)}^{Bslips}$ *holds. Then we compute the set of possible trajectories satisfying* $\neg\varepsilon_{(1)}^{Bslips}$ *that are the closest from the initial trajectory. Since in all these trajectories the distance is not adapted (*$\neg dABok_{(1)}$*) it means that some preferred trajectories will lead also to an accident (assuming that when the distance is not adapted for glaze, it is normal that accidents occur). Hence, we will conclude that* $EE(\Sigma) \diamond_1 \neg\varepsilon_{(1)}^{Bslips} \not\models \neg acc_{(3)}$*, thus that* $\varepsilon_{(1)}^{Bslips}$ *is not a cause. A similar computation will lead to conclude that* $\neg dABok_{(1)}$ *is a cause of the accident.*

Related Work

Several authors also use chronicles to study causality, for instance, one can refer to (Belnap, Perloff, & Xu 2001) or (Mokhtari & Kayser 1998). These approaches use an interventionist modeling of causality. Indeed, in (Mokhtari & Kayser 1998), the authors define the concept of voluntary cause which implies a deliberated choice of the agent among its possible actions. In the approach of Belnap *et al.*, the authors are interested in the representation of the fact that the agent "could have act differently". Only actions of agent are regarded as "true" causes. This definition is not ours because, in this paper, we are not interested in the problem of perceived causality but in the event causation problem (looking for the particular causes of a fact in a given scenario). We are however in agreement with the fact that causes should correspond to actions or events (to reflect that causality is related to manipulability as preached by Von Wright). The works of (Mokhtari & Kayser 1998) and of (Belnap, Perloff, & Xu 2001) use modal logic for the definition of the possible evolutions of the worlds. In our work, we use similar concepts since we also define a preference relation on trajectories, but we use a less complex formalism based simply on propositional logic.

In a non logical framework, the trajectory comparison is close to the work of (Dousson & Ghallab 1994) whose aim is to recognize typical behavior (called temporal scenarios or chronicles) during the monitoring of an evolving system. It would be possible to use similar techniques to compute sequences of events as well as possible explanations for a given trajectory. However, it would require to have an inventory of all the possible sequences of events (which appears rather heavy even if there exist possible factorizations of the models).

In addition, the definition of the evolution of the world by cost functions is close to the concepts used in Markovian Decision Processes (MDP) (Putterman 1994) which provide methods to select the most useful actions in a given situation. The main difference carries on the significance of the weight associated with an event given a situation: here, it corresponds to a degree of surprise, whereas in the MDP, it is a utility (called reward). Moreover, the formalism of the MDP only handles controllable actions, properly speaking, there is no external events.

In this article, we do not mention the problem of the computation of the cost functions associated with the occurrence of each event and associated with the transitions between states of the system. We suppose that this computation comes from more or less defeasible generic laws, this kind of computation is similar to the computation of possibility distribution associated with possibilistic knowledge bases (see for example (Dubois, Lang, & Prade 1994)). One can also refer to (Dupin de Saint-Cyr, Lang, & Schiex 1994) for this type of approach within a penalty framework. These two types of computations start from a knowledge base containing weighted propositional formulas and lead to obtain possibility (or penalty) distributions. Extensions of this type of approach exist that allow to handle at the same time uncertain and defeasible formulas. This is very useful for representing dynamic systems in order to be able to avoid the traditional problems of actions representation (like the "frame" and "qualification" problems which require the use of default reasoning) and to be able to differentiate the fluent persistences (on a scale going from switching fluents to completely inert fluents).

Conclusion

In this article, our contribution is triple, first we extend the concept of extrapolation operator in order to handle events. This operator is based on a preference relation on trajectories which minimizes at the same time the occurrences of surprising events and the surprising transitions. Let us note that the formalism suggested is able to manage the simultaneous events. Secondly, we introduced an update operator on temporal formulas. This operator is defined by means of a distance relation between trajectories and the extended extrapolation operator. The objective of this definition is to be able to provide a logical answer to the question of "what would have occurred if one had modified such thing in a given history". This simple question has many interests in particular for the determination of causal relations and responsibilities. Our third contribution is to propose a definition of event causation based on the update of scenario.

Let us note that within the framework of causal reasoning, we restricted ourselves with the computation of the possible causes in a given story. The problem of *perceived causality* is a forthcoming stage of our work. We argue that the cause perceived by an agent depends on what the agent thinks that one will make of his answer. Most of the time, the cause investigation is done in order to determine responsibilities, this explains the tendency of the human agents to give priority to intentional causes. But it seems to us that it is not always the case, hence there is no need to privilege interventionism systematically. Thus, a promising direction will be to take into account the intention of the request for a cause.

In this article, we make the assumption that the reported observations are *reliable*, it would be interesting to consider the possible existence of bad perceptions of the world. The discovery of an error of observation in a scenario would require to carry out a revision of it. Thus, it would imply to manage, at the same time, modifications of scenario in order to take into account corrections in the initial knowledge and modifications of scenario in order to make hypothetical reasoning. It would require to utilize both revision and update operators, these two operators being applied to temporal formulas.

The distance between trajectories which we use in this work is rather simple, an interesting prospect would be to use the DNA sequences alignment techniques used in biodata processing in order to calculate events sequence alignments. Thus, as in the algorithm of (Needleman & Wunsch 1970), we could associate a cost with the addition, the withdrawal, the substitution or the shift of an event in a sequence. Then we could define the distance between a trajectory and another by the cost of the best alignment between these two trajectories.

Acknowledgments

This research was supported by grant NT05-3-44479 from Agence Nationale de la Recherche. The author thanks Jerome Lang, since this work benefited from recent developments on classical extrapolation to which Jerome contributed much.

References

Belnap, N.; Perloff, M.; and Xu, M. 2001. *Facing the Future: Agents and Choices in Our Indeterminist World*. Oxford University Press.

Bonnefon, J.; Da Silva Neves, R.; Dubois, D.; and Prade, H. 2006. Background default knowledge and causality ascriptions. In *European Conference on Artificial Intelligence (ECAI)*, 11–15. IOS Press.

Boutilier, C. 1994. An event-based abductive model of update. In *Proc. of the Tenth Canadian Conf. on Artficial Intelligence*.

Cartwright, N. 2002. Against modularity, the causal Markov condition, and any link between the two. *British Journal for Philosophy of Science* 53:411–453.

Demolombe, R. 2000. Action et causalité : essais de formalisation en logique. In Prade, H.; Jeansoulin, R.; and Garbay, C., eds., *Le Temps, l'Espace et l'Evolutif en Sciences du Traitement de l'Information*. Cépadues Editions.

Dousson, C., and Ghallab, M. 1994. Suivi et reconnaissance de chroniques. *Revue d'intelligence artificielle* 8(1):29–61.

Dubois, D.; Dupin de Saint-Cyr, F.; and Prade, H. 1995. Update postulates without inertia. In C. Froidevaux, J. K., ed., *Lecture Notes in Artificial Intelligence 946 (Proc. of ECSQARU-95)*, 162–170.

Dubois, D.; Lang, J.; and Prade, H. 1994. Possibilistic logic. In Gabbay, D.; Hogger, C.; and Robinson, J., eds., *Handbook of logic in Artificial Intelligence and logic programming*, volume 3. Clarendon Press - Oxford. 439–513.

Dupin de Saint-Cyr, F., and Lang, J. 2002. Belief extrapolation (or how to reason about observations and unpredicted change). In *Proc. of the 8^{th} KR*.

Dupin de Saint-Cyr, F., and Lang, J. 2008. Belief extrapolation. Technical report, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), Toulouse, France.

Dupin de Saint-Cyr, F.; Lang, J.; and Schiex, T. 1994. Penalty logic and its link with Dempster-Shafer theory. In *Proc. of the 10^{th} Conf. on Uncertainty in Artificial Intelligence*, 204–211. Morgan Kaufmann.

Fikes, R., and Nilsson, N. 1971. Strips : A new approach to the application of theorem proving to problem solving. *Artificial Intelligence* 2:189–208.

Finzi, A., and Lukasiewicz, T. 2003. Structure-based causes and explanations in the independent choice logic. In *Proc. of the 19^{th} Conf. on Uncertainty in Artificial Intelligence*, 225–232. Morgan Kaufmann.

Halpern, J. Y., and Pearl, J. 2001. Causes and explanations: A structural model approach. Part I: Causes. In Breese, J., and Koller, D., eds., *Proc. of the17th Conf. on Uncertainty in Artificial Intelligence (UAI'2001)*, 194–202. Morgan Kaufmann.

Katsuno, H., and Mendelzon, A. 1991. On the difference between updating a knowledge base and revising it. In *Proc. of the 2^{nd} KR*, 387–394.

Lewis, D. 1973. *Counterfactuals*. Harvard University Press.

McCarthy, J., and Hayes, P. 1969. Some philosophical problems from the standpoint of artificial intelligence. In *Machine Intelligence*, volume 4. 463–502.

Mokhtari, A., and Kayser, D. 1998. Time in a causal theory. *Annals of Mathematics and Artificial Intelligence* 22(1-2):117–138.

Needleman, S., and Wunsch, C. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* 48(3):443–53.

Putterman, M. 1994. *Markov Decision Processes. Discrete stochastic dynamic programming*. New York: Wiley-Interscience.

Sandewall, E. 1994. The range of applicability of some non-monotonic logics for strict inertia. *Journal of logic computation* 4(5):581–615.

Sandewall, E. 1995. *Features and Fluents*. Oxford University Press.

Scheines, R. 2004. Causation. *New Dictionary of the History of Ideas*.

Shoham, Y. 1988. *Reasoning about Change - Time and Causation from the Standpoint of Artificial Intelligence*. MIT Press.

Von Wright, G. 1976. Causality and determinism. *The Journal of Philosophy* 73(8):213–218.