Causalité : Comparons les Approches

Salem Benferhat[1], Jean-François Bonnefon[2], Philippe Chassy[2], Rui Da Silva Neves[2], Didier Dubois[2], Florence Dupin de Saint-Cyr[2], Daniel Kayser[3], Farid Nouioua[3], Sara Nouioua-Boutouhami[3], Henri Prade[2] et Salma Smaoui[3]

[1]Université d'Artois, [3]Université Paris 13 et [2]Université de Toulouse

+

Philippe Besnard[2], Marie-Odile Cordier[4], Robert Demolombe[2], Yves Moinard[4], Camille Schwind[5]

[4]Université de Rennes [5] Universitté d'Aix-Marseille

+

Leila Amgoud[2], Denis Hilton[2], François Lévy[3]

Projet ANR NT05-3-44479

MICRAC

Modèles Informatiques et Cognitifs du Raisonnement Causal

- Un projet de 3,5 ans (Janvier 2006 - Juin 2009)
- 4 partenaires :
  - Centre de Recherche en Informatique de Lens (CRIL)
  - Institut de Recherche en Informatique de Toulouse (IRIT)
  - Laboratoire d'Informatique de Paris-Nord (LIPN)
  - Le laboratoire Cognition Langues Langage Ergonomie (CLLE)

- Intelligence Artificielle (IA) & Psychologie Cognitive

- http://www.irit.fr/MICRAC

Références-1 (Micrac)

Salem Benferhat, Jean-François Bonnefon, Philippe Chassy, Rui Da Silva Neves, Didier Dubois, Florence Dupin de Saint Cyr - Bannay, Daniel Kayser, Farid Nouioua, Sara Nouioua-Boutouhami, Henri Prade, Salma Smaoui. *A comparative study of six formal models of causal ascription.* Proc. *Inter. Conf. on Scalable Uncertainty Management (SUM 2008)*, *Naples, Oct.1-3, 2008*, (S. Greco, T. Lukasiewicz, eds.), Springer, LNAI 5291, 47-62.

J. F. Bonnefon, R. M. Da Silva Neves, D. Dubois, H. Prade. Predicting causality Ascriptions from background knowledge: Model and experimental validation. *Inter. J. of Approximate Reasoning, 48,* 752-765, 2008.

Florence Dupin de Saint Cyr - Bannay. *Scenario update applied to causal reasoning.* Proc. *Inter. Conf. on Principles of Knowledge Representation and Reasoning (KR 2008)*, *Sydney, Australia*, Sept.*16-19, 2008*, (G. Brewka, J. Lang, eds.), AAAI Press, 188-197, 2008.

D. Kayser, F. Nouioua. About norms and causes. International Journal on Artificial Intelligence Tools, 1–2, 7–23, 2005.

Jean-François Bonnefon, Didier Dubois, Henri Prade. *Transitive observation-based causation saliency, and the Markov condition.* Proc. *Inter. Conf. on Scalable Uncertainty Management (SUM 2008)*, *Naples, Oct.1-3, 2008*, (S. Greco, T. Lukasiewicz, eds.), Springer, LNAI 5291, 78-91, 2008.

Autres références: http://www.irit.fr/MICRAC/

Références-2

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

Philippe Besnard, Marie-Odile Cordier, Yves Moinard. *Deriving explanations from causal information.* Proc. *European Conference on Artificial Intelligence (ECAI 2008)*, *Patras, Greece*, July*21-25,08*, (M. Ghallab, C. D. Spyropoulos, N. Fakotakis, N. Avouris, eds.), IOS Press, 44-45, 2008.

Camilla Schwind. Causality in action theories. Electronic Transactions on Artificial Intelligence (ETAI), 3, 27-50, 1999. www.ep.liu.se/ej/etai/1999/007

Vers une comparaison des approches existantes en matière de jugements de causalité

- Causalité: de nombreux problèmes en IA:
  - Diagnostic de causes potentielles
    à partir d'effets observés
  - Induction de lois causales
    à partir de séries d'observations
  - Logiques de l'action
  - Simulation qualitative de systèmes dynamiques

- De nombreux modèles !

- La présentation portera pour une grande part sur
  la perception des relations causales
  et l'attribution ("ascription") causale

- Qu'est-ce que **l'ascription causale**?

  consiste à déterminer quels éléments
  dans une séquence de *faits rapportés*
  sont reliés de manière causale sur
  la base de *connaissances/croyances* sur la marche du monde

- Six modèles d'ascription causale considérés dans la suite
  - **Equations structurelles** (Halpern et Pearl, 2005)
  - Logique **non monotone**
    (Bonnefon, Da Silva Neves, Dubois et Prade, 2006)
  - Relations de préférence entre **trajectoires**
    (Dupin de Saint Cyr, 2008)
  - Approche basée sur des **normes** (Kayser et Nouioua, 2005)
  - Modèles graphiques et **intervention**
    (Pearl, 2000; Benferhat et Smaoui, 2007)
  - Théorie de la **cohérence explicative**
    (Thagard, 1989, Thagard et Verbeurgt, 1998)

+

- Modélisation en logique modale

  (R. Demolombe)

- Causalité et explication

  (P.Besnard, M. O. Cordier, Y. Moinard)

- Causalité et logique de l'action

  (C. Schwind)

- Causality is a complex notion !!
- No general agreement on a definition of causality
- Only main ideas underlying each model will be presented

pardon pour les transparents en anglais !

Différentes positions sur la causalité :

- une *commodité heuristique* qui doit être évacuée du discours scientifique (B. Russell, 1913):

La science recherche les *relations fonctionnelles*

- la causalité est une *caractéristique fondamentale* du monde, et doit être traitée comme une notion primitive

- la causalité *peut être réduite* à d’autres concepts (non causaux)

Réduire les relations causales à

- des *processus physiques*

… mais quid de la causalité … en économie ?

- des *relations probabilistes* entre variables

- des *conditionnelles contrefactuelles*

A cause B =

B ne se serait pas produit si A n’avait pas eu lieu

- la *capacité d’agents à atteindre des buts* en agissant sur ce qui les produit

Le modèle **probabiliste** le plus ancien utilise
comme définition de **'A cause B'**
la relation quantitative standard

$$\mathbf{Prob(B \mid A) > Prob(B)}$$

ou de manière équivalente

$$Prob(B \mid A) > Prob(B \mid \neg A)$$

(I. J. Good)

Mais attention aux *corrélations fallacieuses* !
(« spurious correlations »)

Peut s'écrire de manière équivalente de façon *symétrique*
sous la forme d'une *corrélation* positive

$$\mathbf{Prob(AB) > Prob(A)\ Prob(B)}$$

Fumer donne les dents jaunes

Fumer donne le cancer du poumon

$$\text{Prob(cancer} \mid \text{dents jaunes)} > \text{Prob(cancer)}$$

Avoir les dents jaunes cause(rait) le cancer du poumon !

pour s’en sortir, recours à la notion d’**Interventions**!

(J. Pearl)

Illustrative Example

- We were at "…", I was surprised by the person who braked in front of me, not having the option of changing lane and the road being wet, I could not stop completely in time.

- All models will use the same common core of variables and pieces of knowledge

Driver A ***follows*** *Driver B*

- **Variables** :
  - Acc (occurrence of an accident)
  - Wet (road being wet)
  - Brak (driver B brakes in front of driver A)
  - Reac (driver A brakes in reaction to driver B's braking)

  with variants ReacS and ReacL:

  driver A brakes shortly after B brakes, or with a longer delay)
  - Ncl (A does not have the option of changing lane)
  - Sur (A is surprised)

- Models incorporate this core of **knowledge**:

  – Accidents are *abnormal*

  – Being surprised is *abnormal*

  – ReacL and Wet *promote* Acc

  – Brak and Ncl and Sur *promote* ReacL

  – Brak and Ncl and neg Sur *promote* ReacS

different modelings of ‘abnormal’ and ‘promote’

Structural equations model
(Halpern and Pearl)

- Exogenous (U) and endogenous variables (V)
- Exogenous variables are assumed to be known and out of control
- Only endogenous variables can be causes, or be caused
- Causal model $M = (U, V, F)$
- F is a function that assigns a value to each variable given each value of its parents.
- Each assignment of the exogenous variables $U = u$ determines a unique value x of each subset X of endogenous variables (i.e. $X \subseteq V$)

- Definition
  - The event X=x is said to be an actual cause of an event φ if and only if:
    - AC1: X(u)=x and φ(u) is true (when U takes the value u).
    - AC2: There exists a partition (Z,W) of V with $X \subseteq Z$ and some settings (x', w') of (X,W) such that if Z(u)=z* (z* is the value assigned to Z when U=u), both of the following conditions hold:
      - AC2a: $\varphi \; X \leftarrow x' , W \leftarrow w' \; ^{(u)}$ is false, namely, if X is set to x' and W is set to w' then φ becomes false.
      - AC2b: $\varphi X \leftarrow x, W \leftarrow [w'], Z' \leftarrow [z^*]^{(u)}$ is true for all $W' \subseteq W$ and for all $Z' \subseteq Z$. Namely, if X is set to x, W' is set to [w'], ([w'] is an instantiation of W consistent with w'), and Z' is set to [z*] then φ remains true.
    - AC3: The subset X is minimal.

Allows the identification of *“actual causes”*

- A causal relation between two events A and B represents a process that generates possibilities and presumes *a counterfactual relation* of dependence between the two events.

- Halpern and Pearl’ model involves comparisons between several causal models modified by *interventions* that change the value of some variables in order to verify that the above conditions are satisfied

- Causal models can be represented by a DAG
- The car accident example:

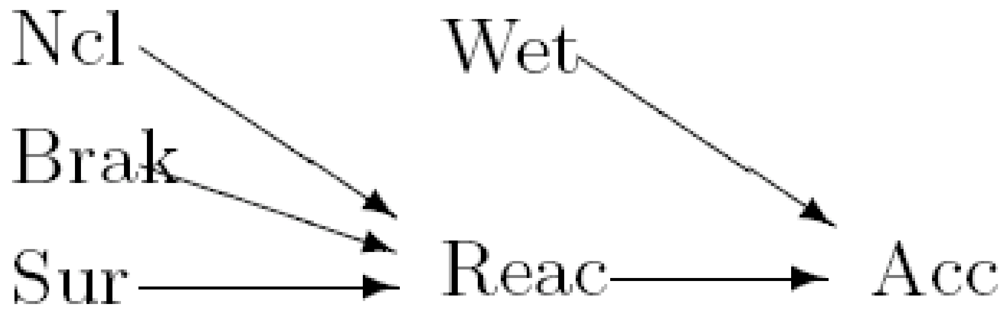


Fig. 1. A causal network

- Structural equations encode background knowledge and the actual context.

  Acc = 1 if wet=1 and Reac=ReacL

  0 otherwise

  Sur=1 and Brak=1 and Ncl=1 and Reac=ReacL and Wet=1 and Acc=1.

- What is the cause of the accident?
  - Consider each possible cause and test if conditions AC1 to AC3 hold.
  - If they do, the possible cause is cause in fact, and do the same test with another possible cause.

- In the example, we conclude for example that Ncl=1 is a cause of Acc=1.

But maintaining the same context we obtain that each event is a cause of Acc=1.

- Limitations:
  – Reasoning with structural equations means that all required information must be available. But, this is not always the case, which may limit the scope of application.

  – The apparent **lack of selective power** of this model may also be considered a weakness, as an event is very easily designated as a cause of another.

  – In order to select preferred causes, it may be interesting to assign `weights' on the basis of levels of normality assigned to each cause according to its implication in making the event happening.

Nonmonotonic Logic Approaches

- *Abnormal facts* are privileged
  when providing causal explanations

- *Material implication is insufficient* for representing causation

- => Nonmonotonic logic-based approaches
  for causal ascriptions.

Nonmonotonic Consequence Approach

- Assume an agent learns of the sequence $\neg B_t, A_t, B_{t+k}$. Call K_t (the context) the conjunction of all other facts known by the agent at time t.

- Let **|~** denote a nonmonotonic consequence relation (in the sense of System P of Kraus et al., 1990).

- If the agent believes K |~ ¬B and K ∧ A |~ B, the agent will perceive <u>A to **cause** B</u> in context K, denoted A ***c*** B.

- If the agent believes that K |~ ¬B, and K ∧ A |~/~ ¬B rather than K ∧ A |~ B, then <u>A is perceived as **facilitating** rather than causing B</u>, denoted A ***f*** B

Nonmonotonic Consequence Approach

- The formalization of the common core of knowledge is :

  (4) $\mid\sim \neg Acc$; (5) $\mid\sim \neg Sur$; (6) $ReacL \wedge Wet \mid\sim Acc$;

  (7) $Brak \wedge Ncl \wedge Sur \mid\sim ReacL$;

  (8) $Brak \wedge Ncl \wedge \neg Sur \mid\sim ReacS$.

- From (4) and (6), we derive $ReacL \wedge Wet$ ***c*** Acc.

The cause of the accident is the conjunction of braking late and the road being wet.

The derivation is based on definitions of ‘cause’ and ‘facilitate’, and on rules of system P, plus rational monotony.

Nonmonotonic Consequence Approach

- In the definitions of ***c*** and ***f***, |~ is a preferential entailment, and a rational closure entailment, respectively.
- Causes and facilitations are **abnormal** in context:
 If A ***f*** B or A ***c*** B then K|~ ¬A.
- Causality is **transitive** *only in particular cases*:
If A is the normal way of getting B in context K, i.e., K ∧ B |~ A,
and if A ***c*** B and B ***c*** C, then A ***c*** C.
- The distinction between causation and facilitation,
as well as the restricted transitivity property,
have been validated by behavioral experiments.
- *Temporality* plays an important role.

Other notions:

- ‘**prevent** to take place’

- ‘**necessary condition**’ (or enabling condition) can be defined in order to deal with *normal* events without which nothing would have happened

- ‘**justification**’:

$$K \mid\!\sim\!/\!\sim B, \quad K \mid\!\sim\!/\!\sim \neg B, \quad \text{and} \quad K \wedge A \mid\!\sim B$$

Nonmonotonic Consequence Approach

- This approach relies on the beliefs about the `normal' states and courses of the world.
- Such beliefs are agent-dependent, which explains that different individuals may have different readings of events
- Exceptional events are favored as potential causes, which help discriminating causes;
- Finally, this approach does not embed the notion of intervention and thus cannot readily distinguish spurious correlation from causation.

Mise à jour et causalité

Principe : “A_t cause B_{t+N}” dans le scénario Σ si

- A_t et B_{t+N} ont eu lieu dans Σ
- si A_t n’avait pas eu lieu dans Σ alors B_{t+N}
- n’aurait pas eu lieu (contre-factualité)

$\Longrightarrow$ calculer si A_t et B_{t+N} ont eu lieu dans Σ

(extrapolation de Σ)

$\Longrightarrow$ calculer ce qui se serait passé

si A_t ne s’était pas produit (mise à jour de Σ par $\neg A_t$)

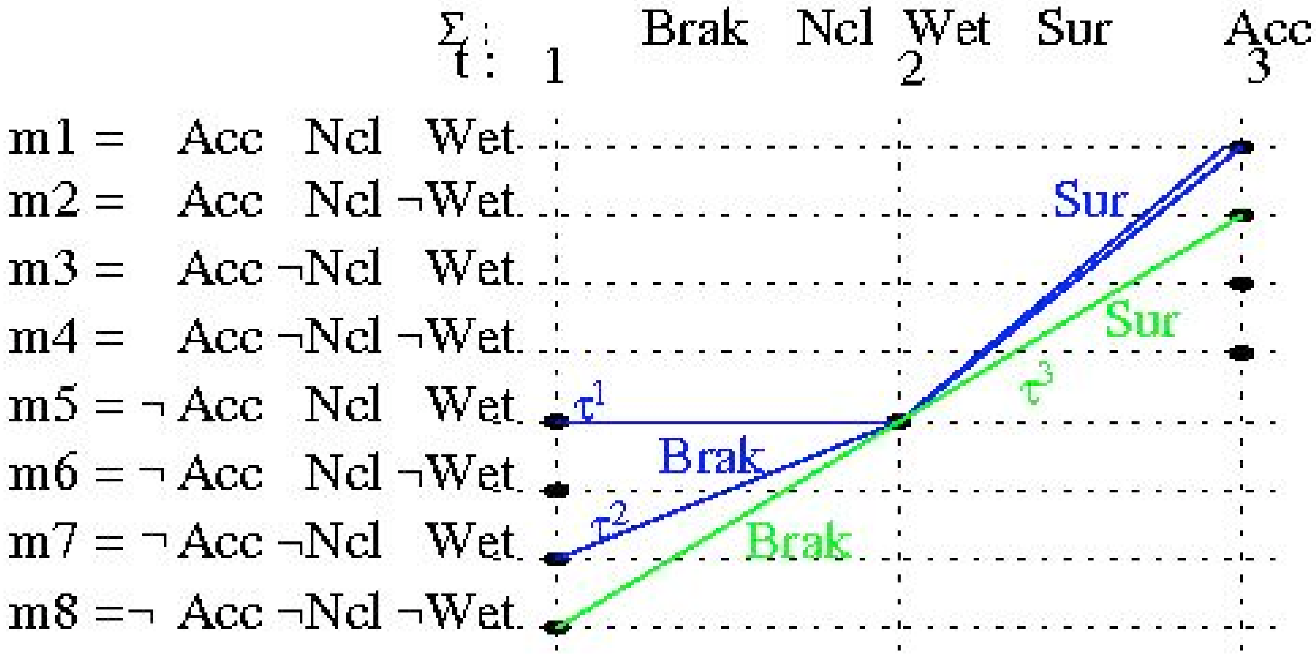
Exemple

- "Nous étions à …, j'ai été surprise (Sur) par la personne qui a freiné (Brak) devant moi, n'ayant pas la possibilité de changer de voie (Ncl) et la route étant mouillée (Wet), je n'ai pu m'arrêter complètement à temps (Acc)"

- $\Sigma = Brak_1 \wedge Sur_2 \wedge Ncl_2 \wedge Wet_2 \wedge Acc_3$
- La surprise est-elle la cause de l'accident ?
  - Calculer si Sur_2 et Acc_3 sont vraies dans Σ
  - (calcul des trajectoires les plus "normales" qui satisfont Σ : extrapolation)
  - Calculer ce qui se serait passé dans Σ si $\neg Sur_2$

Extrapolation du scénario Σ : $E(\Sigma)$

Calcul des trajectoires *les moins surprenantes* pour Σ :



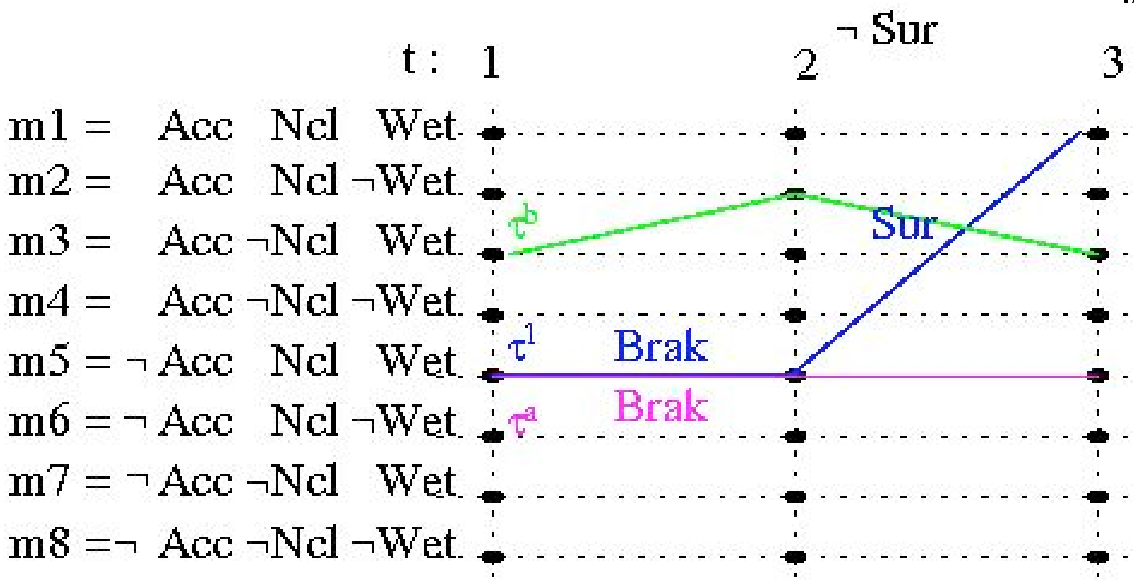
étant données des lois d'évolution et le scénario Σ, τ^1 et τ^2 sont des trajectoires moins surprenantes que τ^3 (où la route est sèche puis mouillée puis sèche)

extrapolation de Σ => ensemble des trajectoires les moins surprenantes pour Σ

Mise à jour

Que ce serait-il passé si le conducteur n'avait pas été surpris en 2?

- *mise à jour* de Σ par $\neg Sur_2$: revient à calculer pour chaque trajectoire de $E(\Sigma)$ les trajectoires “les plus proches” qui satisfont $\neg Sur_2$
- famille de relations de préférence δ_τ [Katsuno Mendelzon90]

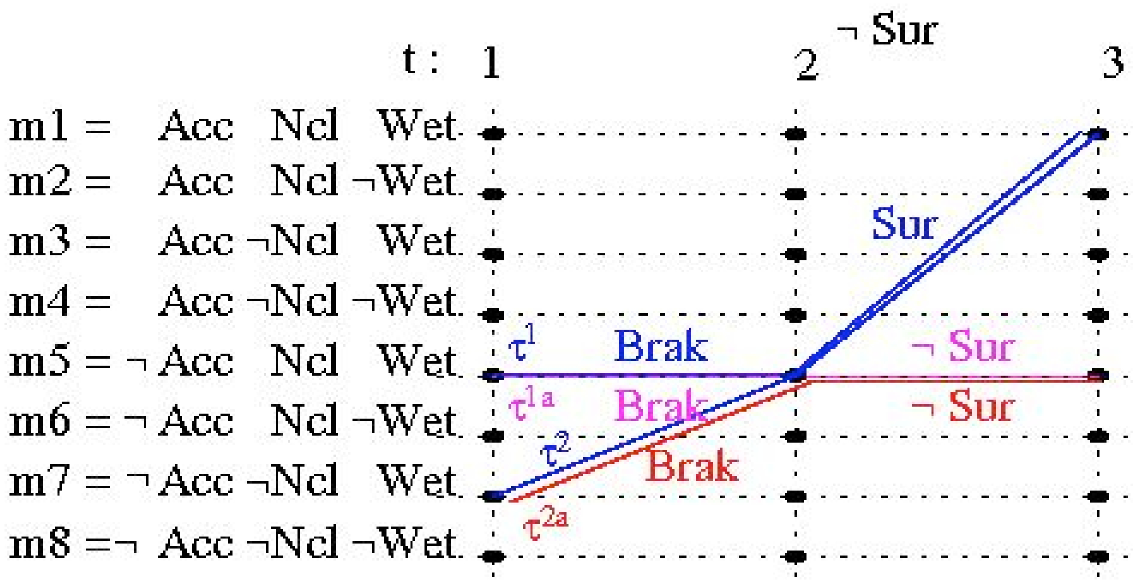


On a proposé $\tau^a \, \delta_\tau \, \tau^b$ t.q. événements de τ^a et τ sont moins différents que entre τ^b et τ (chronologiquement) jusqu'à l'instant du changement

en cas d'égalité, on compare les faits qui diffèrent jusqu'à l'instant du Changement en cas d'égalité, on s'intéresse à la différence entre τ^a et τ après le changement mais seulement au niveau des événements

- ici $\tau^a \, \delta_{\tau 1} \, \tau^b$ (car τ^a identique à τ jusqu'au changement (contrairement à τ^b))

Conclusion



- Mise à jour de Σ par $\neg Sur_2$ donne $\{\tau^{1a}, \tau^{2a}\}$ qui satisfont $\neg Acc_3$
- Donc **la surprise est bien une Cause de l'accident** (car en supposant qu'il n'y a pas de surprise on obtient que normalement il n'y aurait pas d'accident)

- On répète le raisonnement pour toutes les causes possibles ($Brak_1$, Wet_1, ..)

- La **contrefactualité** est liée à la mise à jour (et l'extrapolation) de scénario.

- La mise à jour et l'extrapolation se basent sur des relations de préférence entre trajectoires, qui codent la « **normalité** » (par rapport aux lois d'évolution normale).

- Pour limiter les causes possibles, on pourrait s'intéresser à l'**intention** de la demande de cause

- (ex : si recherche de responsabilité => cause agentive).

Normed based approach

Idea that **norms** are crucial for people to find causes of events

- Searching for the cause of an abnormal event E occurring at time t amounts to finding an agent who should, according to some norm, adopt behavior b at a time t' < t, and actually adopted another behavior b', such that E appears as a normal consequence of b'

- At t', the agent had the possibility to have the normal behavior b; otherwise, b' is only a derived anomaly (then the search must be pursued to find a primary anomaly)

Normed based approach

- The fact that property P holds for agent A at time t is written:
  – holds(P,A,t).
- Two modalities are introduced to express norm violations:
  – should(P,A,t)
  – able(P,A,t)

  standing for: at time t, A should (resp. has the ability to) achieve P.

- For the running example of this paper, we only need a few of these literals:
- (1) $Wet \Rightarrow should(reduced\_speed, A, t)$.
- (2) $holds(Acc,A,t) \Rightarrow should(avoid\_obs,A,t-1)$
- (3) $should(avoid\_obs,A,t) \wedge \neg able(ch\_lane,A,t) \Rightarrow should(stop,A,t)$.

Expressed in this language, the cause of an abnormal event (the `primary anomaly' P_ano) obtains as

- (4) $should(F,A,t) \wedge able(F,A,t) \wedge \neg holds(F,A, t+1) \Rightarrow P\_ano(F,A,t+1)$.

Normed based approach

- In traffic accident examples, the norm-based approach views norms as normative duties.
- To generalize this approach to domains where norms are only what is normal (as opposed to mandatory), it is necessary to organize these norms in a hierarchy, and to conjecture that the most specific violated norm will be perceived as the cause of an abnormal event.
- Requires to gather a *reasonably complete set of norms* for the domain

Graphical Models and Intervention

- Intervention is a critical route to causation

- Ascribing causality becomes easier when experimenting, then observing the effects of the manipulation on the system

- Graphical causal models help make explicit the assumptions needed by allowing inference from *interventions* as well as *observations*

Graphical Models and Intervention

- A causal Bayesian network is a Bayesian network where directed arcs of the graph are interpreted as elementary causal relations between variables.

- When there is an influence relation between two variables, intervention allows to determine the causality relation between these variables. In this case, arcs between variables should follow the direction of the causal process.

Graphical Models and Intervention

- Pearl (2000) proposed an approach for handling interventions using causal graphs based on a **`do'** operator.

- Graphical models are compatible both with a probabilistic and a possibilistic modeling of uncertainty.
The *possibilistic* setting, more qualitative, is used here.
It allows us to more easily relate graphical models to nonmonotonic approaches.

Graphical Models and Intervention

- An intervention forcing a variable A_i to take the value a_i is denoted $do(A_i=a_i)$ or $do(a_i)$. This intervention consists in making A_i true independently from all its other direct causes (i.e. parents).

- Graphically, this modification is represented by deletion of links.

- The resulting graph is said to be mutilated.

Graphical Models and Intervention

- Another approach consists in adding a new variable DO_{Ai} as a parent node of A_i. DO_{Ai} takes value $do_{Ai\text{-}noact}$ when no intervention is observed, and value do_{ai} when an intervention occurs, forcing A_i to take value a_i (a_i belonging to the domain of A_i).

- The resulting graph is called *augmented*.

- Better for computing the effect of interventions

Graphical Models and Intervention

- For binary variables, possibilistic graphical models can encode causality relations as defined by nonmonotonic logic approaches.

  E |~ F is interpreted by $\Pi(E \wedge F) > \Pi(E \wedge \neg F)$.

Prior local possibility distributions : $\Pi(Sur = 0) = 1 > \Pi(Sur = 1) = \alpha$

$$\Pi(Acc = 1 \mid do(Reac = ReacL), Wet = 1) = 1 >$$
$$\Pi(Acc = 0 \mid do(Reac = ReacL), Wet = 1)$$

- Whereas only reported events can be causes, unreported but strongly plausible events can be causes in the possibilistic frameworks.
- Graphical models provide a computational tool for causality ascriptions in presence of interventions.

Theory of Explanatory Coherence

- Thagard's theory of explanatory coherence (1989) views causal ascriptions as attempts to maximize explanatory coherence between propositions.
- In the accident example, maximizing coherence would lead to accept the most plausible hypotheses that explain the accident and reject the alternative hypotheses
- If one proposition explains another, then there is a positive constraint between them. Negative constraints result from events that prevent or are inconsistent with other events.

Theory of explanatory coherence

- Maximizing coherence is implemented under the form of connectionist algorithms such as ECHO.

- ECHO creates a network of units with explanatory and inhibitory non directional links and then makes inference by spreading activation through the network until all activations have reached stable values.

Theory of explanatory coherence

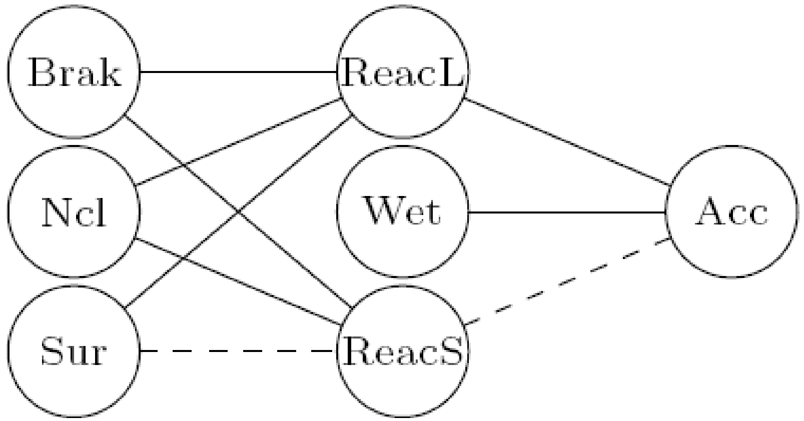


Fig. 2. Accident Example Network in ECHO

Each node represents a variable. The three nodes on the left and the Wet node correspond to variables with priority; in this case, initial conditions at the beginning of the accident process.
Dotted lines represent inhibitory links.

Theory of explanatory coherence

- ECHO establishes an ordering between accepted causes.   Final activation = causal power
- Previous experimental studies suggests that human distinction between facilitation and causation is based on the strength of the relation between events
- Inference in ECHO is not monotonic, not transitive, and can be forward or backward
- Here the only central notion is coherence, but abnormality, temporality and intervention can be added
- ECHO can be translated in probabilistic networks
- Has been used in diverse psychological domains

Argumentation

Agents may argue about where causation takes place
in a sequence of events
One may use a weaker notion of causality
Sequence of arguments, and counterarguments
… IAF'07

But agents may also use argumentation in a self-serving way:
in the case of a traffic accident, they may attempt to **present events in a favorable way** to produce a 'biased description,'
that remains respectful of the essential facts,
but triggers inferences to conclusions
that are in favor of the arguer.

Synthesis

	Main Cause(s)	Can select conjunctive causes	Can select multiple causes.
StructEqM	All	Yes	Yes
NMonLoAp	ReacL ∧ Wet	Yes	Yes
TrajBPrefR	Sur	Yes	Yes
NormBasA	Not having reduced its speed in time.	Yes	Yes
GraMo&Int	ReacL & Wet	Yes	Yes
ExplCoh	Wet & Sur (& ReacL)	Yes	Yes

Synthesis

- Poor model's agreement

- Wet and Reacl the more given conjunctive cause

- Specificity of the Norm Based Model who provides an external (a norm) cause for Acc.

- Lake of specificity of Halpern and Pearl Model. Every considered event is cause in fact.

- Graphical Models and TEC provides degrees of causal power.

General discussion

- All models (excepted TEC and Halpern and Pearl model) make explicit the contrast between normal and abnormal states of affairs.

- One model (norm-based) privileges as causes events that are under the control of agents (agentivity).

- Explicit intervention-like manipulations, where a variable can be forced to take some value.

- Counterfactuality is central in Halpern and Pearl model, Graphical models and Trajectory based model.

General discussion

- Only TEC and the structural equation model do not make explicit the temporal relation between the factors they deal with
(e.g. braking occurs before stopping)

- Although the different models start with the same core of variables and pieces of knowledge, they rely on representation frameworks of different expressive power,
and may exploit additional pieces of knowledge
not assumed to be available in other models.

- Computational tractability is not discriminative criterion

  All the formalisms underlying the approaches we have reviewed have already been implemented.

Modélisation de la Causalité en Logique Mo

Robert Demolombe[1]

[1] IRIT
Groupe LILaC

Juin 2008

Logiques de la causalité

La causalité est par essence liée à une **action** qui influence l'évolution du monde,
pas à **un** état particulier du monde

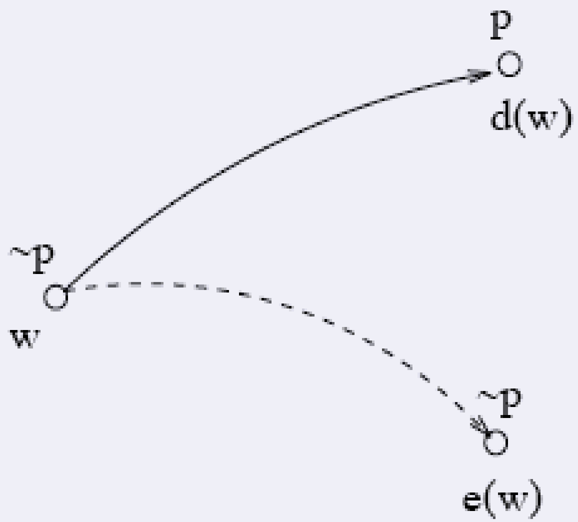
Approche de von Wright

Georg Henrik von Wright (Norm and Action, 1963)
Analyse des différentes attitudes possibles d'un agent par rapport
l'évolution d'une proposition p

- w : le monde dans lequel on est **avant** que l'agent ait agi
- $d(w)$: le monde dans lequel on est **après** que l'agent ait ag
(pour "done")
- $e(w)$: le monde dans lequel on serait **si l'agent n'avait pa**
agi (pour "empty action")

Dans chacun de ces mondes on peut avoir p ou $\neg p$
8 attitudes possibles de l'agent

Exemple p : la porte est fermée



L'agent fait en sorte que p ("*to bring it about that*")

Intuitivement : "l'agent a fermé la porte"

Modalité $Br_i p$

Conditions de satisfaisabilité

$M, w \models Br_i p$ ssi $M, w \models \neg p$ et $M, d(w) \models p$ et $M, e(w) \models \neg p$

DEUX conditions

Condition **suffisante** $M, d(w) \models p$: il suffit que l'agent ait fait ce qu'il a fait pour que l'on ait p

Condition **nécessaire** $M, e(w) \models \neg p$: s'il n'avait pas fait ce qu'il a fait, **toutes choses égales par ailleurs**, on n'aurait pas eu p

Conditions de satisfaisabilité

$M, w \models Br_i p$ ssi $M, w \models \neg p$ et $M, d(w) \models p$ et $M, e(w) \models \neg p$

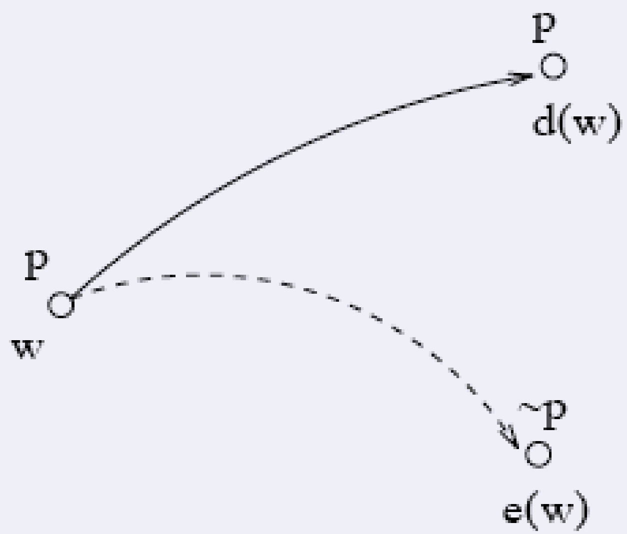
DEUX conditions

Condition **suffisante** $M, d(w) \models p$: il suffit que l'agent ait fait ce qu'il a fait pour que l'on ait p

Condition **nécessaire** $M, e(w) \models \neg p$: s'il n'avait pas fait ce qu'il a fait, **toutes choses égales par ailleurs**, on n'aurait pas eu p

$e(w)$ est un monde **contrefactuel**

L'agent maintient ("sustain") p



Si l'agent n'avait pas agi, le vent aurait ouvert la porte

$M, w \models Ss_i p$ ssi $M, w \models p$ et $M, d(w) \models p$ et $M, e(w) \models \neg p$

Variations sur la caractérisation des mondes contrefactuels

Relation de similitude entre mondes :

David Lewis (Counterfactuals, 1973)

Actions non déterministes :

Stig Kanger (Law and Logic, 1972)

Ingamr Pörn (Action Theory and Social Science, 1977) (et aussi Andrew J.I. Jones)

Causalité et capacité de changer le monde :

Nuel Belnap, John Horty (The deliberative STIT, JPL, 1995)

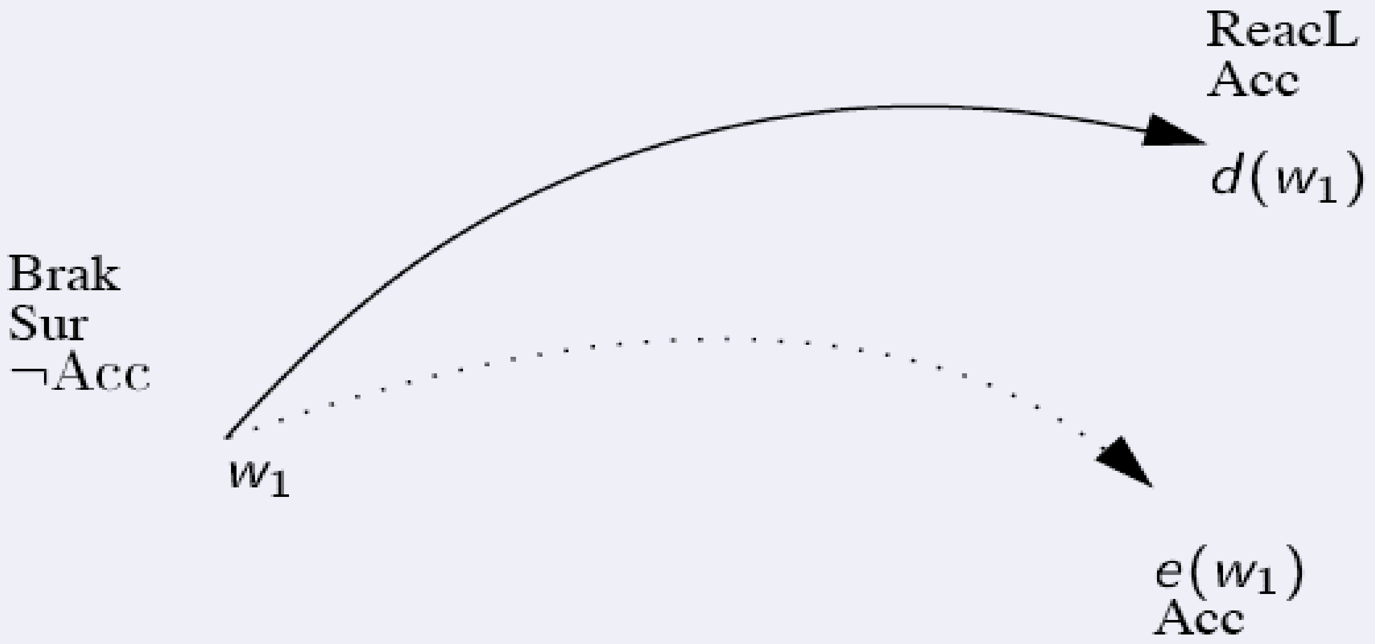
Relation de similitude définie pour chaque contexte

Actions simultanées

Risto Hilpinen (On Action and Agency, 1997)

Exemple de l'accident de voiture

Question : est-ce que A est la cause de l'accident Acc?
(selon von Wright)
Quels sont les mondes w_1, $d(w_1)$ et $e(w_1)$?



w_1 : B a fait Brak, A est surpris
$d(w_1)$: A a fait ReacL
$e(w_1)$: A n'a "rien" fait, ni ReacL, ni ReacS

On a $M, w_1 \models \neg Br_A Acc$
Il est faux que A est la cause de Acc, car si A n'avait "rien" fait, l'accident aurait quand même eu lieu
On a $M, w_1 \models \neg Ss_A \neg Acc$
A n'a pas maintenu l'absence d'accident

Explications générées à partir d'information causale et taxinomique

Philippe Besnard[1] Marie-Odile Cordier[2] Yves Moinard[3]

[1]CNRS, IRIT, Univ. Paul Sabatier Toulouse, France
[2]Univ. Rennes1 [3]INRIA Bretagne-Atlantique [23]IRISA, Rennes, France

Journées IAF, Paris, 21–23 oct. 2008

- Comment **représenter des énoncés causaux** et qu'**en déduire.**

Attention:
Les relations de causalité sont ici **données**

- **Causalité**: une relation entre états.
- La cause doit être ce qui produit l'effet.
  "Une explosion à l'air libre cause du bruit."
- La cause produit toujours l'effet (**exclut** *"fumer cause le cancer"*). Mais graduations faciles à introduire.
- **Une cause n'est pas une simple corrélation** *"La marée basse* ne cause pas *la marée haute environ six heures plus tard"*.

Le formalisme causal

Données

- **C**: Des *formules causales* construites sur des atomes classiques et des atomes *causaux* comme $On(alarm)$ *cause* $Heard(bell)$.
- **O**: Des *atomes ontologiques*: $soft\_bell \rightarrow_{IS-A} bell$, $Heard \rightarrow_{IS-A} Perceived$.
- **W**: Des *formules classiques* (incompatibilités, co-occurrences,...): $Heard(soft\_bell) \rightarrow \neg Heard(loud\_bell)$.

Résultats

- **Atomes d'explications**: α *explique* β *car_poss* Φ:
  α *est une explication de* β **car** Φ *est* **poss***ible*.

$\{\alpha, \beta\} \cup \Phi =$\{atomes classiques concrets comme $On(alarm)$\}.

Le formalisme causal: inférer les explications

Le cas général: *"ontologie en cassis"* $\mathit{cause} \searrow^{\mathit{IS-A}} \nearrow$

Si α **cause** β, $\gamma \rightarrow_{\textsf{IS-A}} \beta$, $\gamma \rightarrow_{\textsf{IS-A}} \delta$,
alors α **explique** δ **car_poss** $\{\alpha, \gamma\}$.

Exemple: $C = \{On(alarm)\ \ cause\ \ Heard(bell)\}$,
$O = \{loud\_bell \rightarrow_{IS-A} bell,\ loud\_bell \rightarrow_{IS-A} loud\_noise\}$.

Comme $W \not\models \neg(On(alarm) \wedge Heard(loud\_bell))$, on déduit les trois atomes d'explication $On(alarm)\ explique \cdots$

1. $\cdots Heard(loud\_noise)\ car\_poss\ \{On(alarm), Heard(loud\_bell)\}$

   et a fortiori

2. $\cdots Heard(loud\_bell)\ car\_poss\ \{On(alarm), Heard(loud\_bell)\})$
3. $\cdots Heard(bell)\ car\_poss\ \{On(alarm)\}$.

Transitivité de l'explication

L'explication est transitive (ou presque)

Schéma de transitivité de l'explication

Si $\left(\begin{array}{l} W \not\models \neg(\bigwedge_{\varphi \in \Phi \cup \Psi} \varphi), \\ \alpha \textit{ explique } \beta \textit{ car\_poss } \Phi \\ \beta \textit{ explique } \gamma \textit{ car\_poss } \Psi \end{array}\right)$ alors $\begin{array}{c} \alpha \textit{ explique } \gamma \\ \textit{car\_poss } (\Phi \cup \Psi) \end{array}$

L'accident

- Ce qu'exprime la narration, en termes de ce formalisme

$$\begin{array}{ll} & Brak\_Sur\_Wet\_Ncl \;\; cause \;\; Acc \\ \text{où} & Brak\_Sur\_Wet\_Ncl \leftrightarrow Brak \wedge Sur \wedge Wet \wedge Ncl \end{array}$$

- Ce qu'exprime la formalisation et le vocabulaire suggérés:

$$\begin{array}{llll} Brak\_Sur \;\; cause \;\; ReacL & \text{où} & Brak\_Sur \leftrightarrow Brak \wedge Sur; & \text{et} \\ Brak\_Non\_Sur \;\; cause \;\; ReacS & \text{où} & Brak\_Non\_Sur \leftrightarrow Brak \wedge \neg Sur. & \end{array}$$

- On en déduit les ▸ atomes d'explication:

$$\begin{array}{ll} & Brak\_Sur\_Wet\_Ncl \; explique \; Acc \;\; car\_poss \; \{Brak\_Sur\_Wet\_Ncl\} \\ \text{et} & Brak\_Sur \;\; explique \;\; ReacL \; car\_poss \; \{Brak\_Sur\}, \cdots \end{array}$$

L'accident: une variante

Utile d'examiner aussi des variantes ("elaboration tolerance").

Variante On n'a pas connaissance d'une route mouillée:

> *I was surprised by the person who braked in front of me, not having the option of changing lane, [crash!].*

Qu'attend-on? S'il s'agit de trouver une explication fondée sur une des causes "les plus accidentogènes": **Deux possibilités:**

Attribuer des degrés aux atomes causaux

$$\alpha \ \textit{cause} \ \beta \quad [\textit{force}]$$

Les atomes d'explication seraient "gradués" de la même façon.

Incorporer ce "degré" à l'atome *Acc*.

$Acc(force)$ où *force* est un nombre (*Acc* pour accident)

Ne garder que ce qui explique $Acc(f_{max})$ pour le plus grand f.

Conclusion sur ce formalisme et les accidents

Il existe un programme ASP (answer set programming) de ce formalisme.
Tous les ASP (Smodels, DLV,clingo) admettent degrés/poids
$\rightsquigarrow$ ajout ci-dessus faciles.

Données

$$\begin{array}{rcl} Brak & cause & Acc(5) \quad , \cdots , \\ Brak\_Sur\_Wet\_Ncl & cause & Acc(20) \end{array}$$

$voiture \rightarrow_{IS-A}$ $vehicule$... (atomes ontologiques au besoin)
et formules pour définitions, contraintes,...

Résultats

▸ Atomes d'explications dont la conclusion est le *Acc* avec le plus fort ▸ degré possible.

Conditional Logics for representing Causation

Camilla Schwind

Laboratoire d'Informatique Fondamentale
Marseille

October 1, 2008

Causal Implication

Properties of Causality Relations

- No reflexivity. *Acc* cannot cause *Acc*
- Conjunction
- *Brak* and *Wet* causes *Acc*
- *Brak* does not cause *Acc* and *Wet* does not cause *Acc*
- No contraposition.
- Reasoning by case (or)
- Cumulativity
- ...

Transitivity?

- yes: *Rain* causes *wet* and *Brak* $\wedge$ *Ncl* $\wedge$ *Sur* causes *ReacL* and *ReacL* $\wedge$ *wet* causes *Acc*. Hence *Rain* causes *Acc*
- no: *Storm* causes *Rain* and *Clima change* causes *Storm* and *Ozon* causes *Clima change* and ... *Burn* causes *Clima change* . . . BUT NOT *Burn* causes *Acc*
- Notion of "distance" of events
- Notion of relevance

Formalization

- Conditional logic
- Counterfactuals
- Weak (causal) implication $>$
- *ATM* set of atoms of the language,
- $A > B$ or $A > (B > C)$: imbricated causal chains are possible
- $(A > B) > C$: imbricated hypotheses are possible, but do perhaps not have a meaning.

Types of causal implication and related notions

- "Actual causality" B occurs and has been caused by A
  causal ascription
- "Potential" Causation: Normally A causes B. A occurs,but B does not need to occur
- "Prevention": A causes B and C causes $\neg A$. C occurs and B will not occur. C prevents B.

Example

1. $ReacL \wedge Wet$ causes Acc
2. 6 *weeks sunshine* causes $\neg Wet$
3. 6 *weeks sunshine* prevents Acc

Example

- $\mathit{ReacL} \wedge \mathit{Wet} > \mathit{Acc}$
- $\mathit{Brak} \wedge \mathit{Ncl} \wedge \mathit{Sur} > \mathit{ReacL}$
- $\mathit{Brak} \wedge \mathit{Ncl} \wedge \neg \mathit{Sur} > \mathit{ReacS}$
- $\mathit{ReacL} \rightarrow \mathit{Brak}$
- $\mathit{ReacS} \rightarrow \neg \mathit{ReacL}$
- $\mathit{Reac} \leftrightarrow \mathit{ReacS} \vee \mathit{ReacL}$

Prevention

- $Sun > \neg Wet$
- $\neg Wet \rightarrow \neg(ReacL \wedge Wet)$
- $ReacL \wedge Wet > Acc$
- Sun prevents Acc