

## La logique des incohérences : un modèle formel pour l'analyse de l'erreur humaine

Thèse de Valentin Fouillard, encadrée par Nicolas Sabouret (LISN), Safouan Taha et Frédéric Boulanger (LMF)

Les humains prennent parfois des décisions en apparence illogiques, avec des conséquences graves :

- 49% des causes d'accidents d'aviation
- 46% des causes d'accidents de la route

Analyse des accidents pour éviter qu'ils se reproduisent :

- amélioration de la sécurité des systèmes
- amélioration de la formation des agents humains

Utilisation des méthodes formelles pour :

- assurer la pertinence de l'explication des causes de l'accident
- explorer des possibilités ignorées des enquêteurs humains
- définir formellement la plausibilité d'un scénario d'accident



Vol 447 Rio-Paris du 1er juin 2009

### Problèmes :

- Modéliser les comportements illogiques en utilisant la logique
- Évaluer la plausibilité d'un comportement illogique dans le contexte d'un accident

**Langage logique de modélisation**

- Proposition indicée temporellement

alarme<sub>t</sub> : il y a une alarme au pas de temps t

---

**Point de vue**

Les observations : Obs<sub>t</sub>  
alarme<sub>t</sub> ∈ Obs<sub>t</sub>

Les règles de raisonnements :  $\mathcal{R}$   
 $R_1 \equiv \text{alarme}_t \rightarrow \text{danger}_t \in \mathcal{R}$

Les désirs :  $\mathcal{D}$   
 $\neg \text{danger}_t \in \mathcal{D}$

---

**Les actions effectuées :  $\mathcal{T}$**

tirer<sub>t</sub> ∈  $\mathcal{T}$ , pousser<sub>t</sub> ∈  $\mathcal{T}$

---

**États de croyances**

état précédent :  $B_{t-1}$ , état courant :  $B_t$

---

**Exemple**

Point de vue :

- une alarme qui lui indique une panne d'essence
- une réserve d'essence suffisante pour continuer

Action :

- atterrissage en urgence

Ignorance :

- Se concentre sur l'alarme et ignore la réserve

Dekker (2006) : l'agent agit rationnellement. Son comportement paraît irrationnel parce que ses croyances ne sont pas celles que nous pensons.

Dans notre modèle, un agent a accès à des observations, raisonne selon des règles et a des désirs. Pour maintenir la cohérence de son état de croyance, il peut ignorer certains de ces éléments

À partir d'un état de croyance à la date t-1, la prise en compte des règles de raisonnement, des désirs et des observations au temps t peut mener à un ensemble incohérent. Les prérequis de l'action effectuée à l'instant t peuvent également contribuer à cette incohérence. Il faut donc retirer de cet ensemble de faits un ensemble de faits ignorés pour construire l'état de croyance à la date t.

**Minimal Correction Set (MCS)**

Pour un système  $\Phi = \{\phi_1, \phi_2 \dots \phi_n\}$ ,  
 $M \subseteq \Phi$  est un MCS de  $\Phi$  ssi :

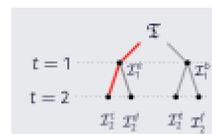
- $\Phi \setminus M$  est cohérent
- $\forall \phi_i \in M, (\Phi \setminus M) \cup \{\phi_i\}$  est incohérent

**Définition de l'opérateur**

$\mathfrak{M}(\Phi, \text{shielded}) \Leftrightarrow$  Révision AGM

- Utilisation de l'algorithme de (Liffton & Sakallah, 2008)
- La fonction  $\mathfrak{M}$  retourne les MCSes de  $\Phi$
- $\text{shielded} \subset \Phi$  et ne peut pas être corrigé

La détermination des ignorances se fait par le calcul d'un MCS. On choisit un ensemble de correction minimal afin d'éviter des hypothèses gratuites qui ne seraient pas nécessaires au rétablissement de la cohérence. *shielded* est un ensemble de faits qui ne peuvent pas être remis en cause. Le calcul des MCS peut se faire à l'aide d'un solveur tel que Z3.

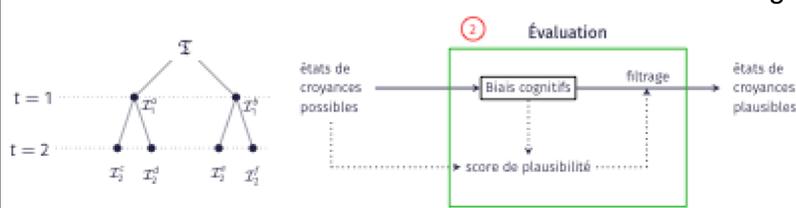


Plusieurs MCS sont possibles à chaque instant, on obtient donc un arbre des états de croyance possibles au fil du scénario de l'accident.

Pour restreindre le foisonnement de l'arbre des états de croyance (900 000 états pour le scénario du Rio-Paris) et limiter le nombre de diagnostics remontés aux experts, nous utilisons un modèle de biais cognitifs pour évaluer la plausibilité des diagnostics.

L'hypothèse est ici que les diagnostics dans lesquels on reconnaît des biais cognitifs, qui sont des manières de raisonner habituelles, sont plus plausibles.

Les biais cognitifs sont définis formellement comme des motifs de comportements psychologiques primitifs. La recherche des biais se fait en recherchant ces motifs le long des branches de l'arbre des états de croyance.



<https://hal.science/tel-03999313v1>