Histoires et évolution de croyances

Florence Dupin de Saint-Cyr
https://www.irit.fr/~Florence.Bannay/

Séminaire du pole Interaction du Lab-STICC
18 janvier 2024

Mon Histoire

Belief Change, Consistency and Argumentation

Garant :	J. Lang
Rapporteurs :	G. Kern-Isberner, P. Marquis, N. Maudet
Jury :	S. Benferhat, P. Besnard, C. Cayrol, A. Hunter, H. Prade

Gestion des croyances et de l'évolutif en logiques pondérées

Directeurs :	D. Dubois et H. Prade
Rapporteurs :	T. Schaub, C. Froidevaux
Jury :	A. Herzig, M. Cayrol

IRIT
Institut de Recherche en Informatique de Toulouse
CNRS - Toulouse INP - UT3 - UT Capitole - UT2

ADRIA
Argumentation Décision Raisonnement Incertitude Apprentissage

Mon équipe: Argumentation Décision Représentation Incertitude Apprentissage
40 chercheurs en IA Fondamentale

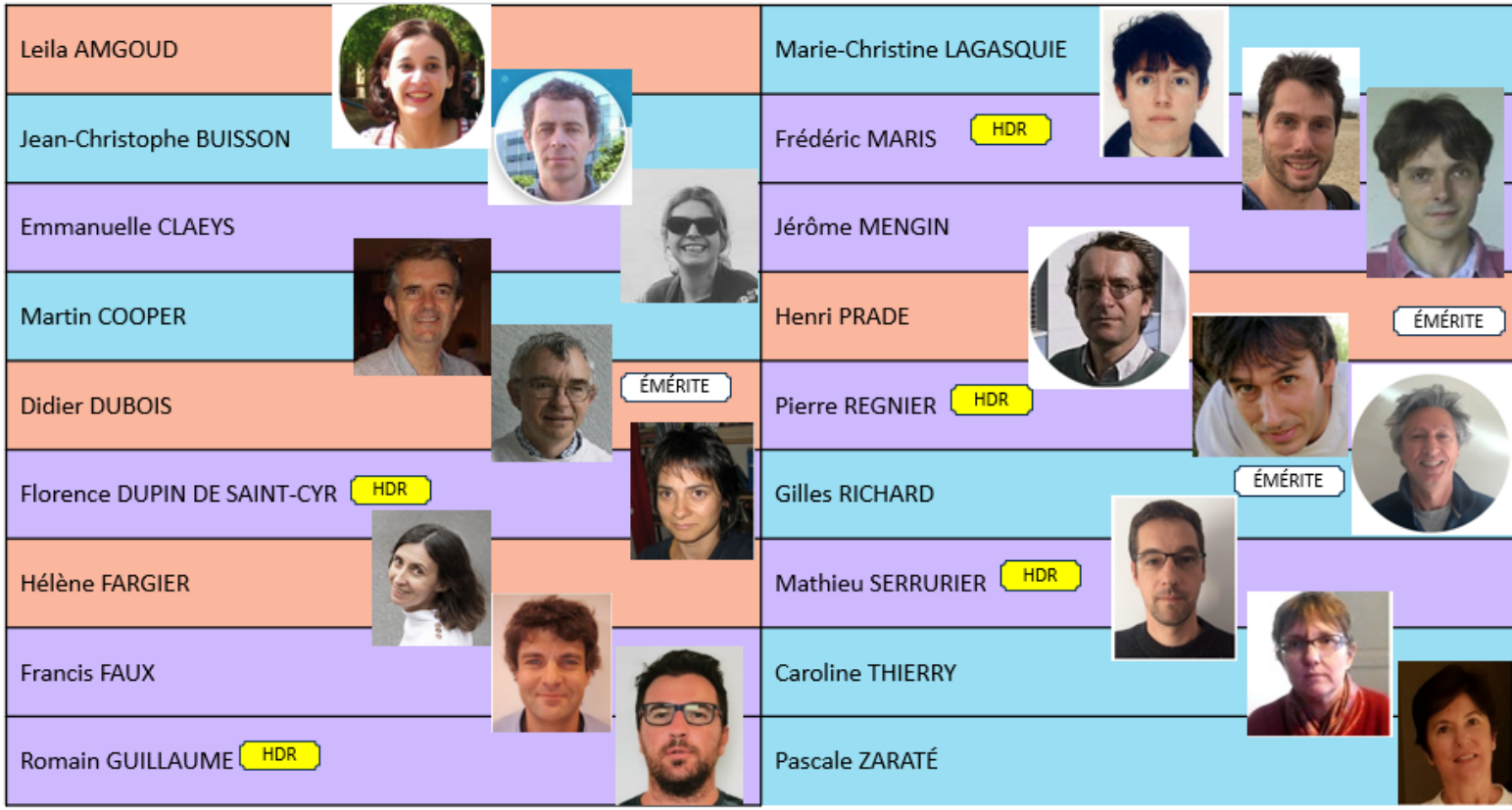
ADRIA

4 DR CNRS

6 PR

8 MC(5HDR)

+ 20 doctorants
+ 3 postdoc
+ 4 associés



13 UNIVERSITÉ TOULOUSE III PAUL SABATIER + 2 UNIVERSITÉ TOULOUSE Jean Jaurès + 1 UNIVERSITÉ TOULOUSE 1 CAPITOLE + 1 INP ENSEEIHT + 1 ISIS INGÉNIEURS

Domaine de recherche :

IA: rep. connaissances et raisonnement

Mon domaine de recherche : IA

Sous-domaine : Représentation connaissances & raisonnement

"Intelligence" = aptitude délibérée d'acquérir et de mettre en œuvre des connaissances et des compétences

- IA actuelle:
  - IA des données (apprentissage): données $\rightarrow$ prédiction (données)
  - IA des connaissances : connaissances $\rightarrow$ connaissances
  - ▶ XAI (IA explicable): intégration de connaissances sur IA des données
- ▶ Analogie avec systèmes 1 (intuition) et 2 (raisonnement) de *(TvKa74)*.

La représentation des connaissances

- Connaissance : "possession symbolique des choses" *(Godin18)*
  - VTL (représentations visuelles)
  - Connaissances explicitables (langue orale, langue écrite, graphique, codage)
  - Connaissances implicites : savoir-faire, savoir être, savoir reconnaître
- Représentation logique ou graphique de connaissances explicites
- Pour formaliser le raisonnement
  - produire un raisonnement idéal (rationnel)
  - simuler un raisonnement humain :
    - pour comprendre, prédire, guider, compléter un être humain

Histoire de mes contributions (1)

- Mon expertise de base : traitement de l'évolution des connaissances
- Contributions en argumentation
- Contributions en logique
- Langage visuel typé zoomable
- Contributions en apprentissage

Mes thèmes actuels et mes projets au CERV

- Modèles cognitifs pour le raisonnement, la communication et la décision
  - biais de raisonnement (fallacies, rationalité limitée) *(BDB19)(BiDu21)*
  - humour *(DuPr23)*
  - les préjugés *(DuFa23)*
  - évaluation d'arguments analogiques
- Avec des membres du CERV
  - jam création de textes poétiques ou drole à PFIA22 *(BEDDLCPAADGHM22)*
  - atelier clé en main "DriveToGæther" à PFIA23 *(DPVMMBE24)*
  - le mode candide
  - caractérisation d'une narration intéressante et logique
  - narration interactive vue comme un dialogue
  - l'élicitation de valeurs morales, découverte de biais, de préjugés en RV
  - détection/déconstruction de biais dans des histoires

Logique en bref

Logique en bref

Base de connaissances

$lime(a)$
$yellow(b)$
$q(nixon) \wedge r(nixon)$
$p(tweety)$
$b(tweety)$
$lime(X) \rightarrow yellow(X)$
$lime(X) \rightarrow green(X)$
$green(X) \rightarrow color(X)$
$yellow(X) \rightarrow color(X)$
$lime(X) \rightarrow cold_color(X)$
$green(X) \rightarrow cold_color(X)$
$yellow(X) \rightarrow warm_color(X)$
$cold_color(X) \rightarrow color(X)$
$warm_color(X) \rightarrow color(X)$
$warm_color(X) \wedge cold_color(X) \rightarrow \bot$
$r(X) \rightarrow \neg p(X)$
$q(X) \rightarrow p(X)$
$q(X) \rightarrow a(X)$
$a(X) \rightarrow b(X)$
$q(X) \rightarrow \neg b(X)$
$b(X) \rightarrow f(X)$
$p(X) \rightarrow b(X)$
$p(X) \rightarrow \neg f(X)$
$b(X) \rightarrow l(X)$
...

Plusieurs problèmes :

- incomplétude
- inconsistance
- évolution

Logique en bref

Base de connaissances

$lime(a)$
$yellow(b)$
$q(nixon) \wedge r(nixon)$
$p(tweety)$
$b(tweety)$
$lime(X) \rightarrow yellow(X)$
$lime(X) \rightarrow green(X)$
$green(X) \rightarrow color(X)$
$yellow(X) \rightarrow color(X)$
$lime(X) \rightarrow cold_color(X)$
$green(X) \rightarrow cold_color(X)$
$yellow(X) \rightarrow warm_color(X)$
$cold_color(X) \rightarrow color(X)$
$warm_color(X) \rightarrow color(X)$
$warm_color(X) \wedge cold_color(X) \rightarrow \bot$
$r(X) \rightarrow \neg p(X)$
$q(X) \rightarrow p(X)$
$q(X) \rightarrow a(X)$
$a(X) \rightarrow b(X)$
$q(X) \rightarrow \neg b(X)$
$b(X) \rightarrow f(X)$
$p(X) \rightarrow b(X)$
$p(X) \rightarrow \neg f(X)$
$b(X) \rightarrow l(X)$
...

Plusieurs problèmes :

- incomplétude :
  plusieurs interprétations possibles

  $lime(a), \overline{lime(b)}$, $\overline{f(a)}, yelow(b)$, ...

  $lime(a), lime(b)$, $\overline{f(a)}, yelow(b)$, ...

  ...

  $lime(a), \overline{lime(b)}$, $f(a), yelow(b)$, ...

- inconsistance

- évolution

Logique en bref

Base de connaissances

$lime(a)$
$yellow(b)$
$q(nixon) \wedge r(nixon)$
$p(tweety)$
$b(tweety)$
$lime(X) \rightarrow yellow(X)$
$lime(X) \rightarrow green(X)$
$green(X) \rightarrow color(X)$
$yellow(X) \rightarrow color(X)$
$lime(X) \rightarrow cold_color(X)$
$green(X) \rightarrow cold_color(X)$
$yellow(X) \rightarrow warm_color(X)$
$cold_color(X) \rightarrow color(X)$
$warm_color(X) \rightarrow color(X)$
$warm_color(X) \wedge cold_color(X) \rightarrow \bot$
$r(X) \rightarrow \neg p(X)$
$q(X) \rightarrow p(X)$
$q(X) \rightarrow a(X)$
$a(X) \rightarrow b(X)$
$q(X) \rightarrow \neg b(X)$
$b(X) \rightarrow f(X)$
$p(X) \rightarrow b(X)$
$p(X) \rightarrow \neg f(X)$
$b(X) \rightarrow l(X)$
...

Plusieurs problèmes :

- incomplétude
- inconsistance : contradiction
  pas d'interprétation
  $\emptyset$
- évolution

Logique en bref

Base de connaissances

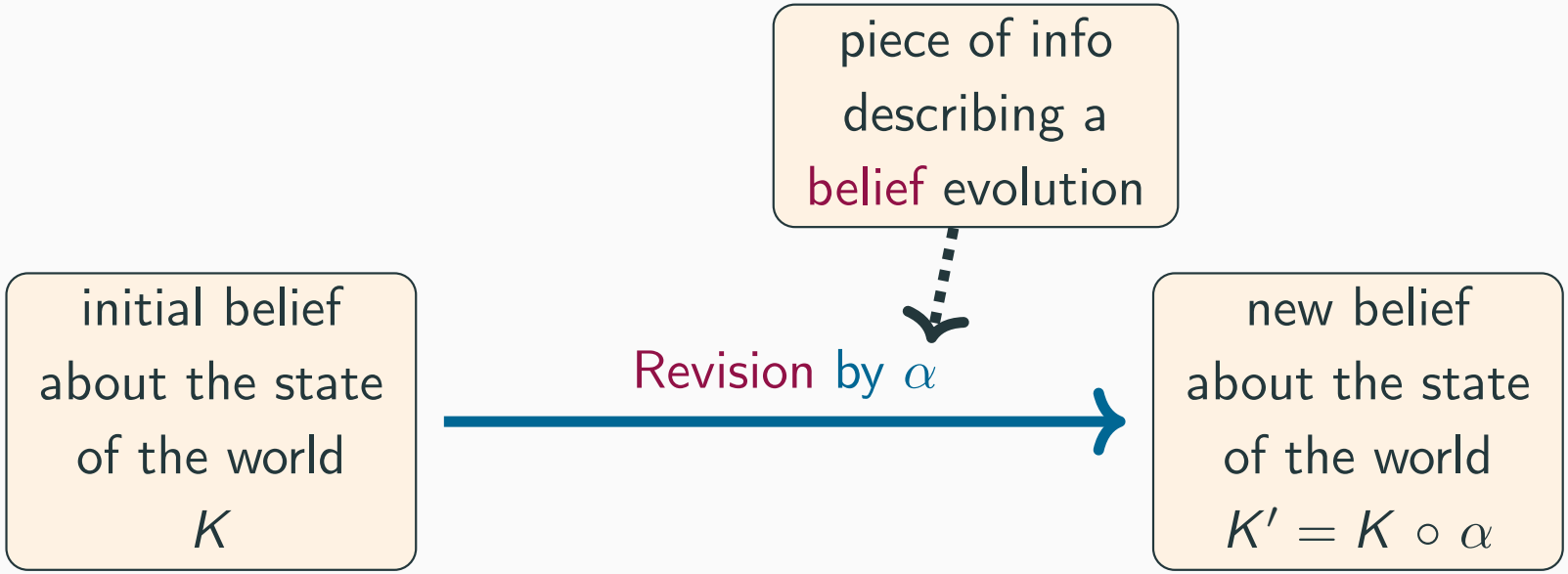
$lime(a)$
$yellow(b)$
$q(nixon) \wedge r(nixon)$
$p(tweety)$
$b(tweety)$
$lime(X) \rightarrow yellow(X)$
$lime(X) \rightarrow green(X)$
$green(X) \rightarrow color(X)$
$yellow(X) \rightarrow color(X)$
$lime(X) \rightarrow cold_color(X)$
$green(X) \rightarrow cold_color(X)$
$yellow(X) \rightarrow warm_color(X)$
$cold_color(X) \rightarrow color(X)$
$warm_color(X) \rightarrow color(X)$
$warm_color(X) \wedge cold_color(X) \rightarrow \bot$
$r(X) \rightarrow \neg p(X)$
$q(X) \rightarrow p(X)$
$q(X) \rightarrow a(X)$
$a(X) \rightarrow b(X)$
$q(X) \rightarrow \neg b(X)$
$b(X) \rightarrow f(X)$
$p(X) \rightarrow b(X)$
$p(X) \rightarrow \neg f(X)$
$b(X) \rightarrow l(X)$
...

Plusieurs problèmes :

- incomplétude
- inconsistance
- évolution :
  on apprend une information α
  - révision
  - mise à jour

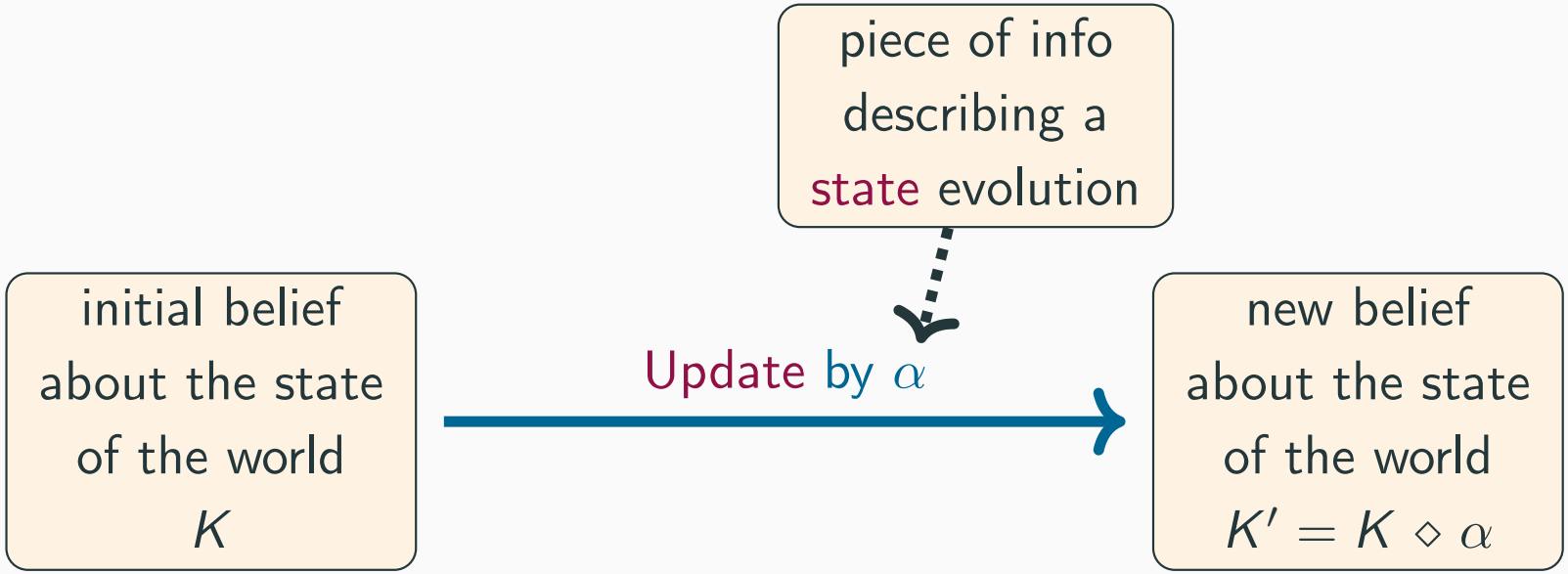
Évolution des croyances : au commencement il y avait la révision et la mise à jour

Belief Change Theory



change of the belief about the state of the world

Belief Change Theory

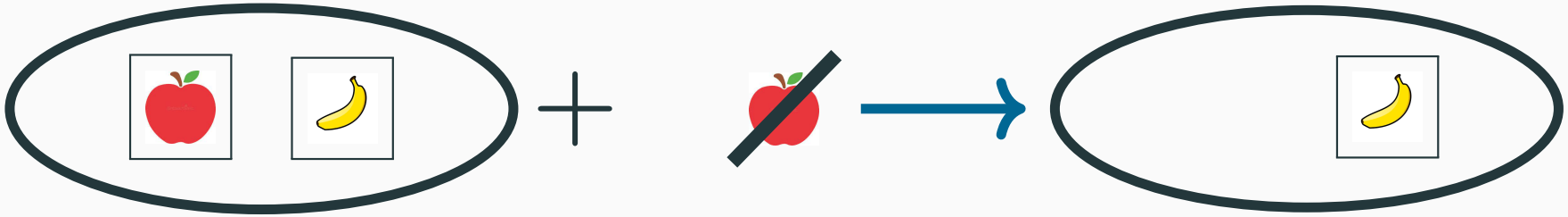


change of the state of the world

Belief Change Theory: Example

- There is either an apple or a banana in the box but not both.
- I learn that there was no apple in the box.

initial belief about the world $a \oplus b =$ $(a \wedge \neg b) \vee (\neg a \wedge b)$ $\xrightarrow{\text{Revision by } \neg a}$ new belief about the world $(a \oplus b) \circ \neg a$

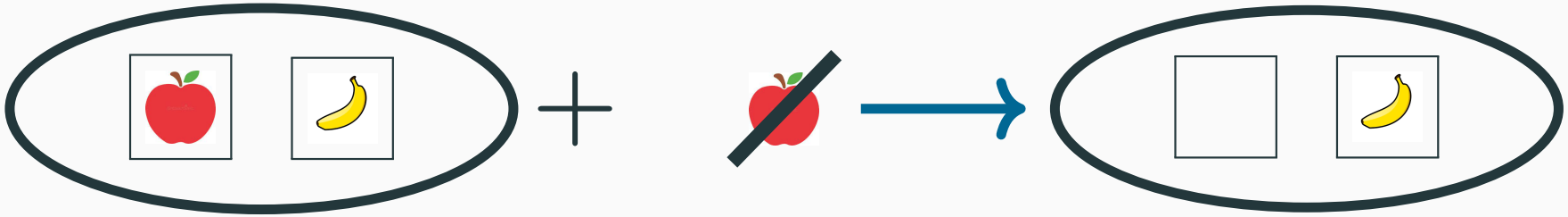


$(\neg a \wedge b)$

Belief Change Theory: Example

- There is either an apple or a banana in the box but not both.
- I learn that the apple stealer has visited the box

initial state of the world $a \oplus b =$ $(a \wedge \neg b) \vee (\neg a \wedge b)$ $\xrightarrow{\text{Update by } \neg a}$ new state of the world $(a \oplus b) \diamond \neg a$



$(\neg a \wedge \neg b) \vee (\neg a \wedge b)$

Théorème de représentation pour la révision

(R1): $K \circ \alpha \models \alpha$

(R2): If $K \wedge \alpha$ consistent, then $K \circ \alpha \equiv K \wedge \alpha$

(R3): If α consistent, then $K \circ \alpha$ consistent

(R4): If $K_1 \equiv K_2$ and $\alpha_1 \equiv \alpha_2$
then $K_1 \circ \alpha_1 \equiv K_2 \circ \alpha_2$

(R5): $(K \circ \alpha) \wedge \psi \models K \circ (\alpha \wedge \psi)$

(R6): If $(K \circ \alpha) \wedge \psi$ is consistent
then $K \circ (\alpha \wedge \psi) \models (K \circ \alpha) \wedge \psi$

Théorème de représentation Révision *(KaMe91a)*

$\circ$ satisfies the postulates (R1) - (R6) iff there is a faithful assignment s.t.:

$$Mod(K \circ \alpha) = \min(Mod(\alpha), \preceq_K)$$

Faithful assignment = mapping from K to a total pre-order $\preceq_K$ "proximity to K" s.t.:

1) If $\omega \models K$ and $\omega' \models K$ then $\omega =_K \omega'$.
2) If $\omega \models K$ and $\omega' \not\models K$ then $\omega \prec_K \omega'$.
3) If $K \equiv K'$ then $\preceq_K = \preceq_{K'}$.

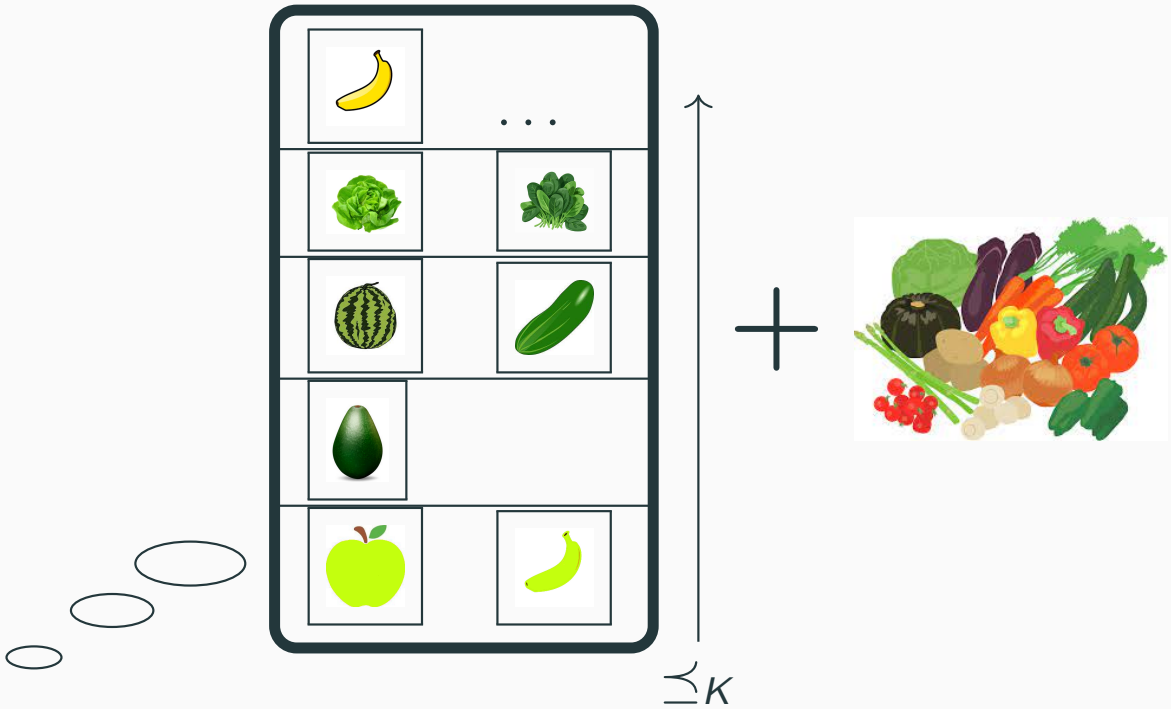
Belief Change Theory: Fruits pas mur

- Il y a une seule chose dans la boite, c'est un fruit rond ou long, il est vert.
- J'apprends que c'est un légume.

K: initial belief about the world
$fruit, rond \oplus long, vert$

$\xrightarrow{\text{Revision by } vegetable}$

new belief about the world



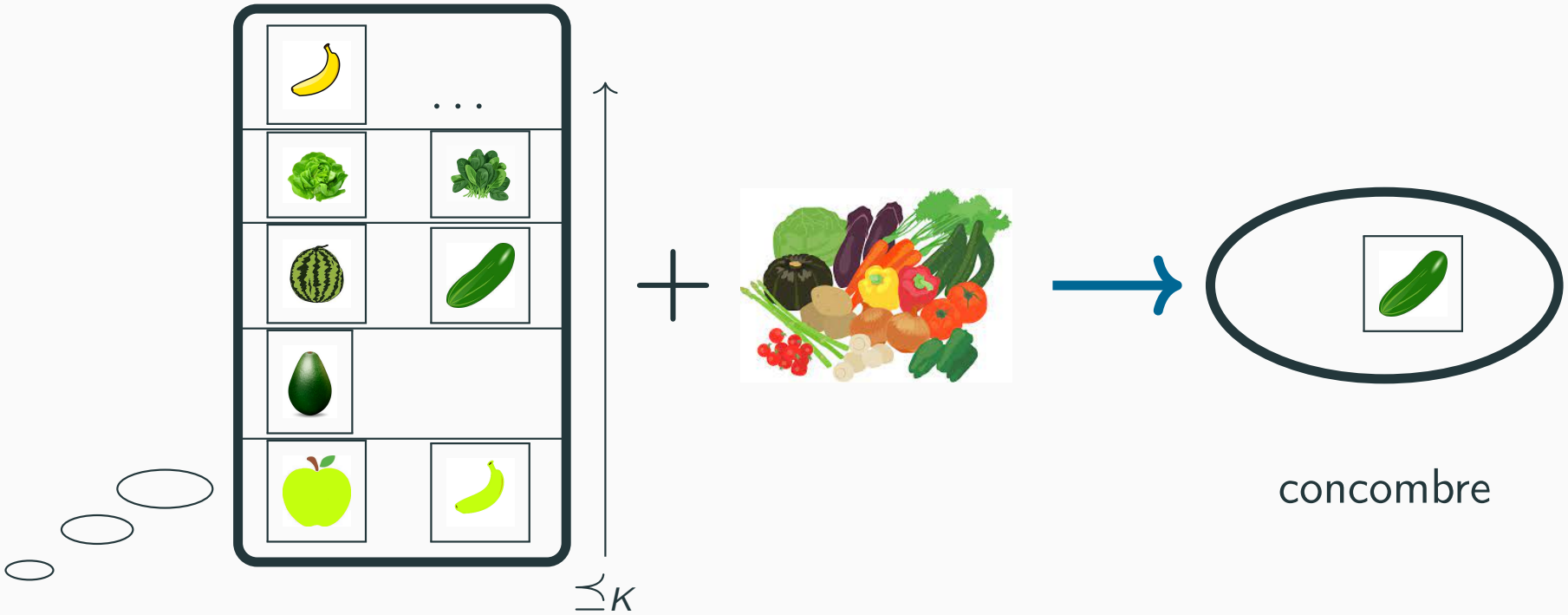
Belief Change Theory: Fruits pas mur

- Il y a une seule chose dans la boite, c'est un fruit rond ou long, il est vert.
- J'apprends que c'est un légume.

K: initial belief about the world
$fruit, rond \oplus long, vert$

Revision by *vegetable* →

new belief about the world
concombre



Théorème de représentation pour la révision

(R1): $K \circ \alpha \models \alpha$

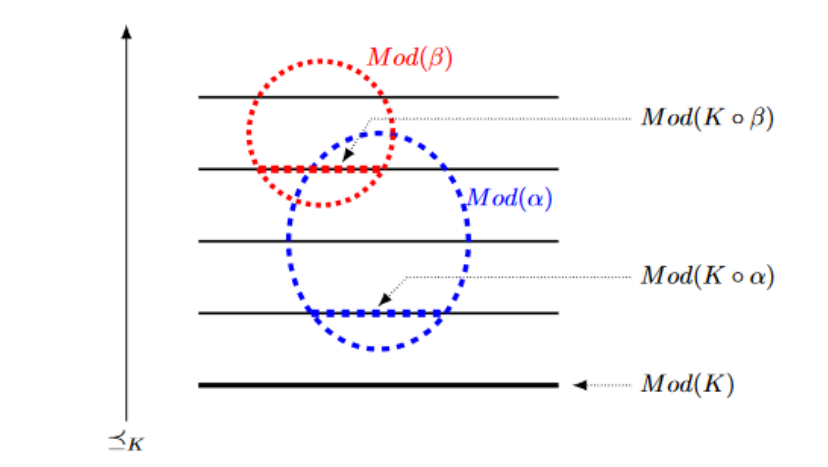
(R2): If $K \wedge \alpha$ consistent, then $K \circ \alpha \equiv K \wedge \alpha$

(R3): If α consistent, then $K \circ \alpha$ consistent

(R4): If $K_1 \equiv K_2$ and $\alpha_1 \equiv \alpha_2$
then $K_1 \circ \alpha_1 \equiv K_2 \circ \alpha_2$

(R5): $(K \circ \alpha) \wedge \psi \models K \circ (\alpha \wedge \psi)$

(R6): If $(K \circ \alpha) \wedge \psi$ is consistent
then $K \circ (\alpha \wedge \psi) \models (K \circ \alpha) \wedge \psi$



Théorème de représentation Révision *(KaMe91a)*

$\circ$ satisfies the postulates (R1) - (R6) iff there is a faithful assignment s.t.:

$$Mod(K \circ \alpha) = \min(Mod(\alpha), \preceq_K)$$

Faithful assignment = mapping from K to a total pre-order $\preceq_K$ "proximity to K" s.t.:

1) If $\omega \models K$ and $\omega' \models K$ then $\omega =_K \omega'$.
2) If $\omega \models K$ and $\omega' \not\models K$ then $\omega \prec_K \omega'$.
3) If $K \equiv K'$ then $\preceq_K = \preceq_{K'}$.

Théorème de représentation pour la mise à jour

U1 $(K \diamond \alpha)$ implies α.
U2 If K implies α then $(K \diamond \alpha)$ is equivalent to K
U3 If K and α consistent then $(K \diamond \alpha)$ consistent
U4 If $K_1 \leftrightarrow K_2$ and $\alpha_1 \leftrightarrow \alpha_2$ then $(K_1 \diamond \alpha_1) \leftrightarrow (K_2 \diamond \alpha_2)$.
U5 $(K \diamond \alpha) \wedge \psi$ implies $(K \diamond (\alpha \wedge \psi))$.
U6 If $(K \diamond \alpha_1)$ implies α_2 and $(K \diamond \alpha_2)$ implies α_1 then $(K \diamond \alpha_1) \leftrightarrow (K \diamond \alpha_2)$.
U7 If K is complete then $(K \diamond \alpha_1) \wedge (K \diamond \alpha_2)$ implies $(K \diamond (\alpha_1 \vee \alpha_2))$.
U8 $(K1 \vee K2) \diamond \alpha$ equivalent to $(K1 \diamond \alpha) \vee (K2 \diamond \alpha)$.
U9 If $(K \diamond \alpha_1) \wedge \alpha_2$ consistent then $(K \diamond (\alpha_1 \wedge \alpha_2))$ implies $((K \diamond \alpha_1) \wedge \alpha_2))$.

Théorème de représentation Mise à jour *(KaMe91b)*

$\diamond$ satisfies the postulates (U1) - (U9) iff there is a faithful assignment (mapping each ω to $\preceq_\omega$ "proximity to ω" which favors ω):

$$Mod(K \diamond \alpha) = \bigcup_{\omega \models K} \min(Mod(\alpha), \preceq_\omega)$$

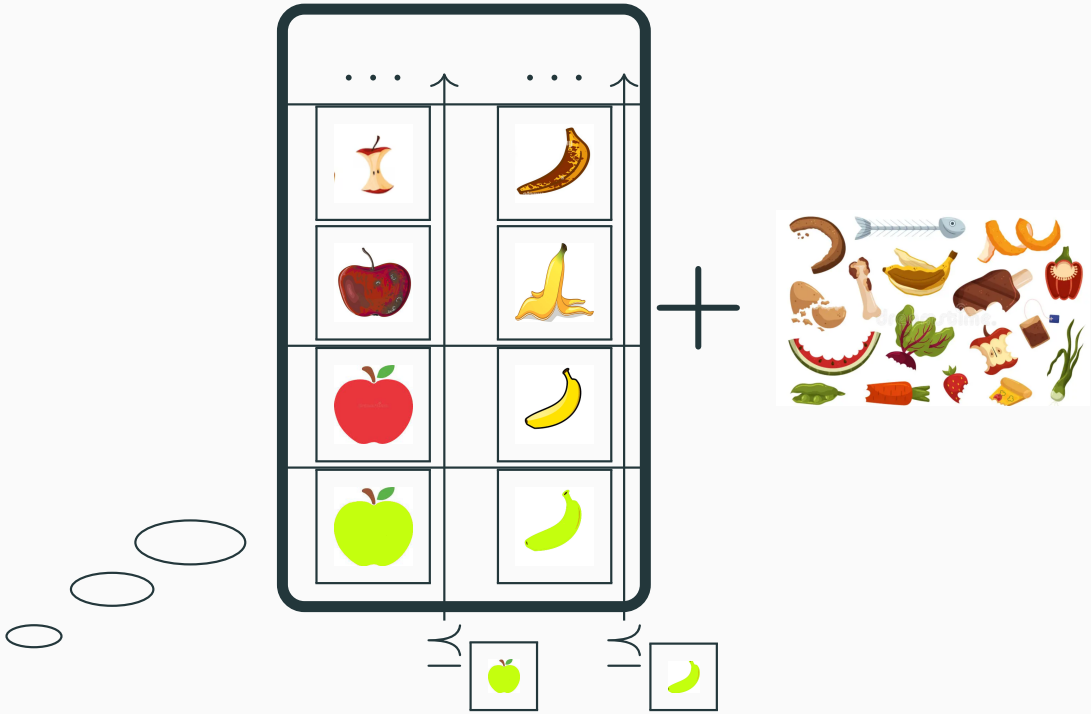
Belief Change Theory: Fruits pas mur

- Il y a une seule chose dans la boite, c'est un fruit rond ou long, il est vert.
- J'apprends qu'il ne reste que des épluchures.

K: initial state of the world
fruit, *rond* $\oplus$ *long*, *vert*

$\xrightarrow{\text{Update by } \textit{epluchures}}$

new state of the world



Belief Change Theory: Fruits pas mur

- Il y a une seule chose dans la boite, c'est un fruit rond ou long, il est vert.
- J'apprends qu'il ne reste que des épluchures.

K: initial state of the world
fruit, *rond* $\oplus$ *long*, *vert*

$\xrightarrow{\text{Update by } \textit{epluchures}}$

new state of the world
peau de banane ou trognon

peau de banane $\vee$ trognon

Comprendre des Histoires : l'extrapolation de croyances

Belief change: résumé Révision et mise à jour

Révision

- monde statique
- K: croyances initiales sur le monde
- α: nouvelle information sur le monde
- $K \circ \alpha$: nouvel état de croyances

Belief change: résumé Révision et mise à jour

Révision

- monde statique
- K: croyances initiales sur le monde
- α: nouvelle information sur le monde
- $K \circ \alpha$: nouvel état de croyances

Mise à jour

- monde dynamique
- K: croyances initiales sur le monde à t=0
- α: nouvelle information sur le monde à t=1
- $K \diamond \alpha$: nouvel état de croyances à t=1

Exemple: scenario 1

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on observe que b fonctionne et que c est cassé	$\neg\, \text{panne(b)} \wedge \text{panne(c)}$

Exemple: scenario 1

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on observe que b fonctionne et que c est cassé	$\neg\, \text{panne(b)} \wedge \text{panne(c)}$

Raisonnement:

- les changements sont rares
- donc à $t = 0$: b fonctionnait et a était en panne.

Exemple: scenario 1

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on observe que b fonctionne et que c est cassé	$\neg\, \text{panne(b)} \wedge \text{panne(c)}$

Raisonnement:

- les changements sont rares
- donc à $t = 0$: b fonctionnait et a était en panne.

t	états de croyance
0	$\text{panne(a)} \wedge \neg\, \text{panne(b)} \wedge \neg\, \text{panne(c)}$
1	$\text{panne(a)} \wedge \neg\, \text{panne(b)} \wedge \text{panne(c)}$

Exemple: scenario 2

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on remplace b par un composant neuf et on casse c	$\neg\, \text{panne(b)} \wedge \text{panne(c)}$

Exemple: scenario 2

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on remplace b par un composant neuf et on casse c	$\neg\, \text{panne(b)} \wedge \text{panne(c)}$

Raisonnement:

- en 0: panne(a) ou panne(b)

Exemple: scenario 2

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on remplace b par un composant neuf et on casse c	$\neg\, \text{panne(b)} \wedge \text{panne(c)}$

Raisonnement:

- en 0: panne(a) ou panne(b)
- on ne peut rien dire

Exemple: scenario 2

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on remplace b par un composant neuf et on casse c	$\neg\, \text{panne}(b) \wedge \text{panne}(c)$

Raisonnement:

- en 0: panne(a) ou panne(b)
- on ne peut rien dire

t	états de croyance
0	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	$\neg\, \text{panne}(b) \wedge \text{panne}(c)$

Exercice: scenario 2 = mise-à-jour ou révision?

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on remplace b par un composant neuf et on casse c	$\neg\, \text{panne}(\text{b}) \wedge \text{panne}(\text{c})$

Exercice: scenario 2 = mise-à-jour ou révision?

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on remplace b par un composant neuf et on casse c	$\neg\, \text{panne(b)} \wedge \text{panne(c)}$

- K: connaissances initiales, on apprend α

Exercice: scenario 2 = mise-à-jour ou révision?

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on remplace b par un composant neuf et on casse c	$\neg\, \text{panne}(\text{b}) \wedge \text{panne}(\text{c})$

- K: connaissances initiales, on apprend α
- ici α est une action: changement du monde

Exercice: scenario 2 = mise-à-jour ou révision?

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on remplace b par un composant neuf et on casse c	$\neg\, \text{panne}(\text{b}) \wedge \text{panne}(\text{c})$

- K: connaissances initiales, on apprend α
- ici α est une action: changement du monde
- Raisonnement = mise à jour

Exercice: scenario 2 = mise-à-jour ou révision?

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on remplace b par un composant neuf et on casse c	$\neg\, \text{panne}(b) \wedge \text{panne}(c)$

- K: connaissances initiales, on apprend α
- ici α est une action: changement du monde
- Raisonnement = mise à jour

t	états de croyance
0	$(\text{panne}(a) \oplus \text{panne}(b) \wedge \neg\, \text{panne}(c)$
1	$\neg\, \text{panne}(b) \wedge \text{panne}(c)$

Exercice: scenario 1 Mise-à-jour ou révision?

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on observe que b fonctionne et que c est en panne	$\neg\, \text{panne(b)} \wedge \text{panne(c)}$

Exercice: scenario 1 Mise-à-jour ou révision?

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on observe que b fonctionne et que c est en panne	$\neg\, \text{panne(b)} \wedge \text{panne(c)}$

Raisonnement: α est une observation

Exercice: scenario 1 Mise-à-jour ou révision?

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on observe que b fonctionne et que c est en panne	$\neg\, \text{panne(b)} \wedge \text{panne(c)}$

Raisonnement: α est une observation

Raisonnement = extrapolation (révision particulière)

Exercice: scenario 1 Mise-à-jour ou révision?

t	Connaissances	Représentation
0	a ou b en panne mais pas les deux et c fonctionne	$(\text{panne}(a) \oplus \text{panne}(b)) \wedge \neg\, \text{panne}(c)$
1	on observe que b fonctionne et que c est en panne	$\neg$ panne(b) $\wedge$ panne(c)

Raisonnement: α est une observation

Raisonnement = extrapolation (révision particulière)

t	états de croyance
0	panne(a) $\wedge \neg$ panne(b) $\wedge \neg$ panne(c)
1	panne(a) $\wedge \neg$ panne(b) $\wedge$ panne(c)

Extrapolation de croyances

Raisonner avec des observations et des changements imprévus

Observations $\longrightarrow$ **Croyances**
(fiables)
Scenario $\longrightarrow$ **Scenario complété**

t :	1	2	3	4	5
Σ :	a	$a \vee c$	b	$\neg a \vee \neg b$	$\neg c$

$\xrightarrow{\text{Extrapolation}}$

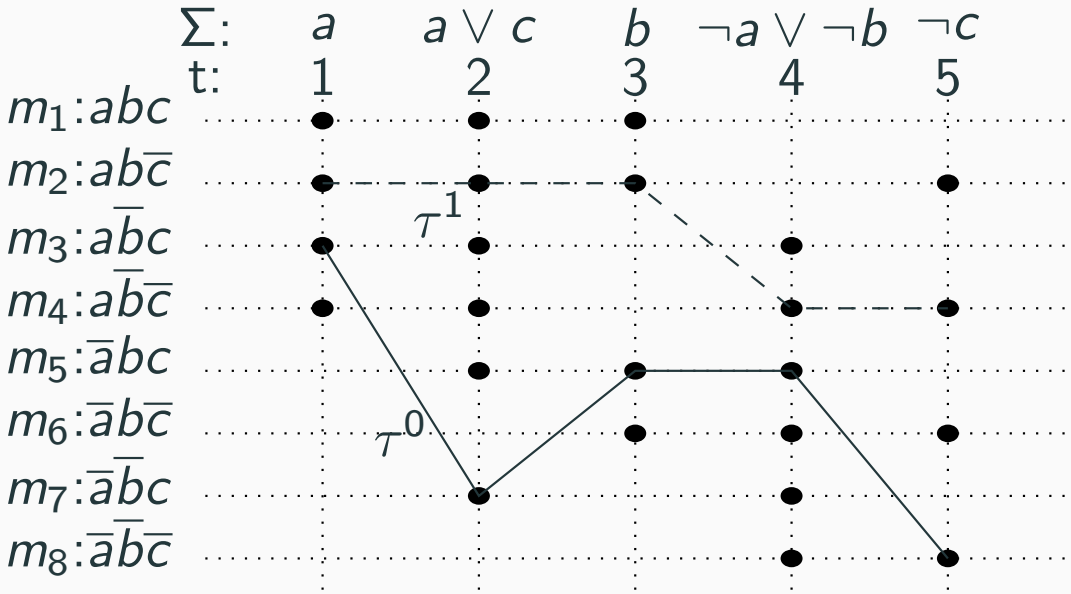
1	2	3	4	5
a	a	b	$(a \oplus b)$	$(a \oplus b)$
$\wedge b$	$\wedge b$	$\wedge c$	$\wedge \neg c$	$\wedge \neg c$
$\wedge \neg c$	$\wedge \neg c$			

Sémantique de l'extrapolation de croyances

- Un scenario Σ = séquence de formules $(\varphi_1, \ldots, \varphi_N)$
- Une trajectoire τ = séquence d'interprétations $(\omega_1, \ldots, \omega_N)$
- $\tau \models \Sigma$ ssi $\omega_1 \models \varphi_1$ et ... et $\omega_N \models \varphi_N$
- $Ch(\tau)$ ensemble (fluent,instant) t.q. fluent change à cet instant

$$Ch(\tau) = \{\langle f, t\rangle | t \in [2, N], \omega_{t-1} \models \neg f \text{ et } \omega_t \models f\}$$

Exemples deux trajectoires



$Ch(\tau^0) = \{\langle \neg a, 2\rangle, \langle b, 3\rangle, \langle \neg b, 5\rangle, \langle \neg c, 5\rangle\}$
$Ch(\tau^1) = \{\langle \neg b, 4\rangle\}$

Sémantique de l'extrapolation de croyances (suite)

- $\preccurlyeq$ pré-ordre (réflexif et transitif) sur des trajectoires
- Extrapolation : trajectoires minimales pour $\preccurlyeq$
- Pré-ordre sur les trajectoires
  - Minimiser le nombre de changements

$$\tau \preccurlyeq_{nc} \tau' \quad \text{ssi} \quad |Ch(\tau)| \leq |Ch(\tau')|$$

  - Minimiser les ensembles de changements

$$\tau \preccurlyeq_{csi} \tau' \quad \text{ssi} \quad Ch(\tau) \subseteq Ch(\tau')$$

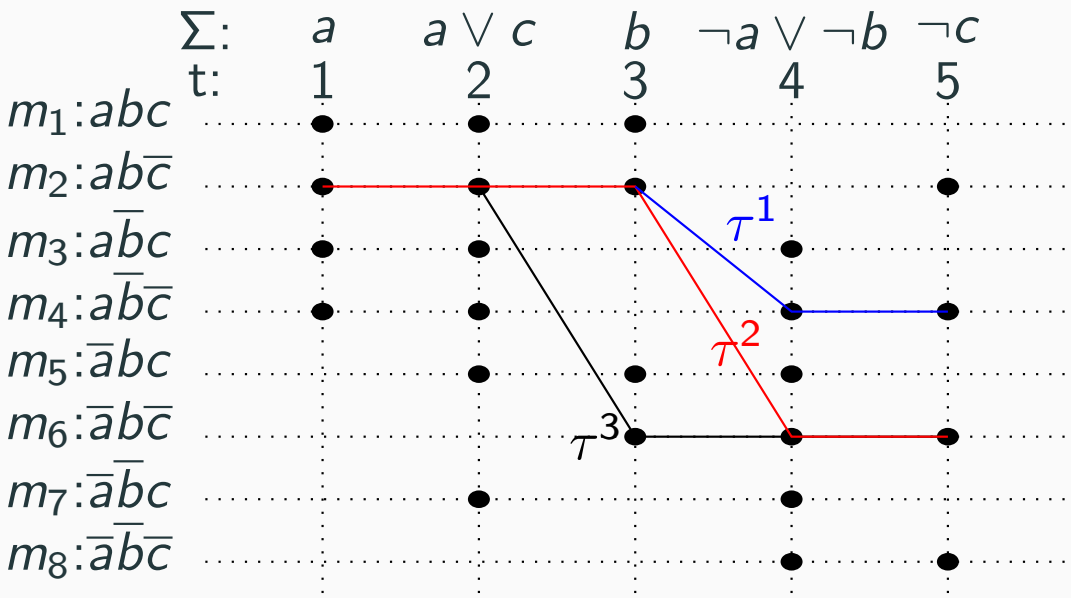
  - Minimiser les pénalités de changement
    - $k : \mathcal{F} \to \mathbb{N}^*$ pénalité du changement de $\neg f$ à f
    - $k(\tau)$= somme des coûts des fluents qui changent dans τ

$$\tau \preccurlyeq_{k} \tau' \quad \text{ssi} \quad k(\tau) \leq k(\tau')$$

  - Pré-ordre chronologique

$$\tau \preccurlyeq_{chr} \tau' \quad \text{ssi} \quad \begin{cases} Ch(\tau,1) \subset Ch(\tau',1) \\ \text{ou } Ch(\tau,1) = Ch(\tau',1) \text{ et } Ch(\tau,2) \subset Ch(\tau',2) \\ \text{ou } \ldots \end{cases}$$

Minimiser le nombre de changements



$Ch(\tau^1) = \{\langle \neg b, 4\rangle\}$
$Ch(\tau^2) = \{\langle \neg a, 3\rangle\}$
$Ch(\tau^3) = \{\langle \neg a, 4\rangle\}$

$$Extrap_{\preccurlyeq_{nc}}(\Sigma) = (a \wedge b \wedge \neg c, a \wedge b \wedge \neg c, b \wedge \neg c, (a \oplus b) \wedge \neg c, (a \oplus b) \wedge \neg c)$$

Sémantique de l'extrapolation de croyances (suite)

- $\preccurlyeq$ pré-ordre (réflexive et transitive) sur des trajectoires
- Extrapolation : trajectoires minimales pour $\preccurlyeq$
- Pré-ordre sur les trajectoires
  - Minimiser le nombre de changements

$$\tau \preccurlyeq_{nc} \tau' \quad \text{ssi} \quad |Ch(\tau)| \leq |Ch(\tau')|$$

  - Minimiser les ensembles de changements

$$\tau \preccurlyeq_{csi} \tau' \quad \text{ssi} \quad Ch(\tau) \subseteq Ch(\tau')$$

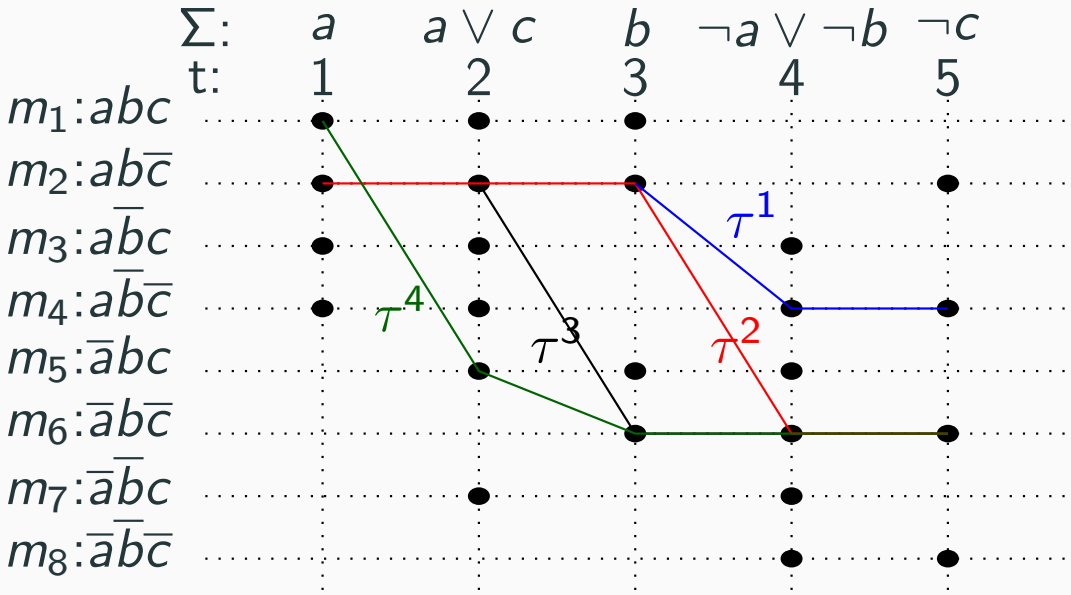
  - Minimiser les pénalités de changement
    - $k : \mathcal{F} \to \mathbb{N}^*$ pénalité du changement de $\neg f$ à f
    - $k(\tau)$= somme des coûts des fluents qui changent dans τ

$$\tau \preccurlyeq_{k} \tau' \quad \text{ssi} \quad k(\tau) \leq k(\tau')$$

  - Pré-ordre chronologique

$$\tau \preccurlyeq_{chr} \tau' \quad \text{ssi} \quad \begin{cases} Ch(\tau,1) \subset Ch(\tau',1) \\ \text{ou } Ch(\tau,1) = Ch(\tau',1) \text{ et } Ch(\tau,2) \subset Ch(\tau',2) \\ \text{ou } \ldots \end{cases}$$

Minimisation chronologique/antichronologique



Minimales chronologique pour changements:

$Ch(\tau^1) = \{\langle \neg b, 4\rangle\}$ et $Ch(\tau^3) = \{\langle \neg a, 4\rangle\}$

$$Extrap_{\preccurlyeq_{chr}}(\Sigma) = (a \wedge b \wedge \neg c, a \wedge b \wedge \neg c, a \wedge b \wedge \neg c, (a \oplus b) \wedge \neg c, (a \oplus b) \wedge \neg c)$$

Minimales anti-chronologiques pour changements:

$Ch(\tau^2) = \{\langle \neg a, 3\rangle\}$ et $Ch(\tau^4) = \{\langle \neg a, 2\rangle, \langle \neg c, 3\rangle\}$

$$Extrap_{\preccurlyeq_{achr}}(\Sigma) = (a \wedge b, (a \oplus c) \wedge b, \neg a \wedge b \wedge \neg c, \neg a \wedge b \wedge \neg c, \neg a \wedge b \wedge \neg c$$

Résumé de extrapolation theory *(DuLa11)*

- Given a *scenario* Σ (sequence of formulas)
- an "inertial" pre-order on trajectories

  $\Rightarrow$ inertial= static trajectories are more plausible

  $\Rightarrow$ example = number of changes in the trajectory

- compute the more plausible trajectories for Σ
- $Extrap(\Sigma)$ is the temporal formula representing the most plausible trajs
- five postulates (E1 : $Extrap(\Sigma) \models \Sigma$...)
- several proposals for pre-orders (inclusion of changes, chronological mini of changes, number of time points of change...)

Extrapolation étendue (basée sur des événements) *(DdSC08)*

- Événement= opération change cours normal évolution du monde
  - cours normal: supposé décrit par lois statiques et dynamiques dérogeables
  - lois résumées par des fonctions qui quantifient degré de surprise d'occurrence d'événements et de transitions
- Une formule temporelle mixte :
  - variables représentant des faits
  - variables represéntant des occurrences d'événements
- Le coût d'une trajectoire= somme pour chaque instant de
  - degré de surprise de l'occurrence des événements
  - degré de surprise d'atteindre la situation suivante
- Extrapolation étendue = trajectoires (mixtes) les moins surprenantes

Problem d'attribution de causalités

- Given
  - a knowledge base K (dynamic laws (or action theory) + static laws)
  - a scenario Σ (facts and event occurrences)
  - a formula φ true at the end of the scenario
- Compute the "causes" of the formula φ.

Quelle est la cause de φ_N ?

φ_t is a cause for ψ_N given a scenario Σ and a knowledge base K if

1. Counterfactual
   - ψ true in Σ: $\Sigma \models \psi_N$
   - φ_t is true in Σ: $\Sigma \models \varphi_t$
   - ψ is false when Σ "modified" such that $\neg\varphi_t$
2. Manipulability: φ_t *event occurrence* or (fact with $t = 0$)

$\Rightarrow$ how to define the "modification" of Σ by a formula?

- encode the question : "what would have happen if $\neg\varphi_t$?"

Quelle est la cause de φ_N ?

φ_t is a cause for ψ_N given a scenario Σ and a knowledge base K if

1. Counterfactual
   - ψ true in Σ: $\Sigma \models \psi_N$
   - φ_t is true in Σ: $\Sigma \models \varphi_t$
   - ψ is false when Σ "modified" such that $\neg\varphi_t$
2. Manipulability: φ_t *event occurrence* or (fact with $t = 0$)

$\Rightarrow$ how to define the "modification" of Σ by a formula?

- encode the question : "what would have happen if $\neg\varphi_t$?"
- It is an update of Σ by $\neg\varphi_t$!!

▶ proximity relation of trajectories

Les histoires drôles

Joke

Example

A man just got hit by a car. The driver gets out of the car and says, "You're lucky we're just in front of a doctor's office.
- Yes! Except the doctor is me!"

From a French collection of "funny" stories: *(Negre70)*

Surprising story

Definition (simple story)

simple story $=_{def}$ *pair* (α, β) *of formulas of .*

- Listening to a story makes an agent revise her beliefs.
- *(KaMe91a)* Revision involves a family $\preceq_K$ of pre-orders for each initial knowledge K (KaMe91b: update involves a family $\preceq_\omega$ for each world ω)
- Agent characterized by $\begin{cases} \circ \text{ a belief base } K \text{ and} \\ \circ \text{ a revision mechanism } \circ \end{cases}$
- Here listener's knowledge $\Sigma = (P, \Delta)$ playing the role of both K and $\circ$:
  $\begin{cases} \circ \text{ a consistent set } P \text{ of propostional formulas: integrity constraints} \\ \circ \text{ a consistent set } \Delta \text{ of default rules } \alpha_i \rightsquigarrow \beta_i\text{: induces pre-order } \preceq_\Delta \end{cases}$
- $Mod(K \circ \varphi)$ is computed by $min(Mod(P \cup \{\varphi\}), \preceq_\Delta)$

Formalization of the joke

Definition (surprising story)

story (α, β) *surprising wrt* $(K, \circ)$ *iff*$_{def}$

$$(K \circ \alpha) \text{ consistent and } (K \circ \alpha) \cup (K \circ (\alpha \wedge \beta)) \vdash \bot$$

Example

$\alpha = injured \wedge doctorNearby$

$\beta = injured \wedge doctorHimself$

$$\Sigma = \left| \begin{array}{l} injured \wedge doctorNearby \rightsquigarrow treatedRapidly \\ injured \wedge \neg treatingDoctor \rightsquigarrow \neg treatedRapidly \\ doctorHimself \rightsquigarrow doctorNearby \\ injured \wedge doctorHimself \rightsquigarrow \neg treatingDoctor \end{array} \right.$$

$$\left. \begin{array}{ll} K \circ \alpha \models & treatedRapidly \\ K \circ (\alpha \wedge \beta) \models & \neg treatingDoctor \wedge \neg treatedRapidly \end{array} \right\} \vdash \bot$$

hence the story is surprising.

Revealing punchline

revealing punchline β: β completely admissible + β explains the situation α

Definition (revealing punchline)

the punchline β in (α, β) is revealing given $(K, \circ)$ iff$_{def}$

$$(K \circ \beta) \text{ consistent and } K \circ \beta \models \alpha$$

Example

$\alpha = \mathit{injured} \wedge \mathit{doctorNearby}$,
$\beta = \mathit{injured} \wedge \mathit{doctorHimself}$,

$$\Sigma = \left| \begin{array}{l} \mathit{injured} \wedge \mathit{doctorNearby} \rightsquigarrow \mathit{treatedRapidly} \\ \mathit{injured} \wedge \neg\mathit{treatingDoctor} \rightsquigarrow \neg\mathit{treatedRapidly} \\ \mathit{doctorHimself} \rightsquigarrow \mathit{doctorNearby} \\ \mathit{injured} \wedge \mathit{doctorHimself} \rightsquigarrow \neg\mathit{treatingDoctor} \end{array} \right.$$

$$K \circ \mathit{injured} \wedge \mathit{doctorHimself} \vdash \mathit{injured} \wedge \mathit{doctorNearby}$$

hence the punchline is revealing.

Potentially funny story

Definition (potentially funny story)

Potentially funny story = surprising with a revealing punchline.

Inspired by *(Raccah15)* "the punchline makes us fall into an unexpected and unavoidable trap".

Proposition (never give the punchline before the end!)

If (α, β) *is a potentially funny story wrt* $(K, \circ)$
then (β, α) *is not potentially funny wrt* $(K, \circ)$.

More defs and props in *(DuPr23)* (The shortest jokes, Psycho-rigidity, Graduality of surprise/revealing levels, ...)

Integrating incongruity

Incongruity is related to the violation of social norms:

- lack of conformity with current practices
  - drink water from a finger bowl
  - wearing oversized shoes
- be deliberately in disregard with obvious universal knowledge
  - four elephants riddle

Example

How to put four elephants in a Citroën 2CV ?

Integrating incongruity

Incongruity is related to the violation of social norms:

- lack of conformity with current practices
  - drink water from a finger bowl
  - wearing oversized shoes
- be deliberately in disregard with obvious universal knowledge
  - four elephants riddle

Example

How to put four elephants in a Citroën 2CV ?
Answer: two in the front and two in the rear.

Formal handling of incongruity

Definition (incongruity)

The statement (α, β) *is felt incongruous given* $\Sigma = (P, \Delta)$
with norms $P_N \subseteq P$ *and* $\Delta_N \subseteq \Delta$

iff_{def}

There is a norm ρ *in* P_N *or in* Δ_N *s.t.*

1. $\alpha \wedge \beta \mid\!\sim^{m}_{\Sigma \setminus \{\rho\}} \textit{Not } \rho$ *(violation of* ρ*)*
2. $\beta \mid\!\sim^{m}_{\Sigma \setminus \{\rho\}} \alpha$ *(revelation in disregard of* ρ*)*

$$\textit{where Not } \rho = \begin{cases} \neg\rho & \textit{if } \rho \in P \\ \alpha_i \wedge \neg\beta_i & \textit{if } \rho = \alpha_i \rightsquigarrow \beta_i \in \Delta \end{cases}$$

The four elephants riddle

Example

α : $i \wedge e$: *4 elephants are in the 2CV car*

β: $tt \wedge e$: *2 elephants in the front and 2 in the rear of the 2CV*

$\Sigma = (P, \Delta)$ *with* $P = \{R1, R2, R3, R4\}$ *and* $\Delta = \emptyset$ *where*

R1: $h \rightarrow \neg i$ *(Something huge cannot be inside a 2CV car)*

R2: $e \rightarrow h$ *(Elephants are huge)*

R3: $h \rightarrow \neg tt$ *(Something huge cannot be put in the rear and in the front)*

R4: $tt \rightarrow i$ *(Putting in the rear and in the front implies putting inside)*

$P \cup \{\alpha\}$ is inconsistent : violation of law R2

$\{\beta\} \cup (P \setminus \{R2\}) \models \alpha$: revealing in disregard of $R2$

hence the story is felt incongruous by the listener

Conclusion (1)

- Outils théoriques:
  - *Révision* = outil pour raisonner sur des changements de croyances
  - *Mise à jour* = outil pour raisonner sur des évolutions du monde
  - *Extrapolation* = outil pour raisonner sur une histoire, comparaison de trajectoires
  - *Non-monotonicity* = dynamic and static laws may be defeasible + revision
    - connaissances sur les lois du monde
    - relation de proximité entre trajectoires
- Applications (requiert connaissances explicites sur les lois sociales et lois du monde)
  - Causalité dans un scenario (φ_t causes ψ_N) :
    - Contrefactuelles= mise à jour scenario par $\neg\varphi_t$
    - Manipulabilité: causes sont des événements ou (faits à t=0)
    - Diagnostic/extrapolation/théorie action: calcul événements expliquant scenario
  - Comprendre une blague :
    - surprise et révélation : requiert capacité raisonnement sous incomplètude

Conclusion (2)

- incongruité = violation d'une norme + raisonnement au mépris de celle-ci

- Perspectives:
  - extrapolation, causalité et narration computationnelle
  - incongruité graduelle : nature choquante de la violation
  - l'ingrédient de Supériorité : nécessite de raisonner sur les autres agents

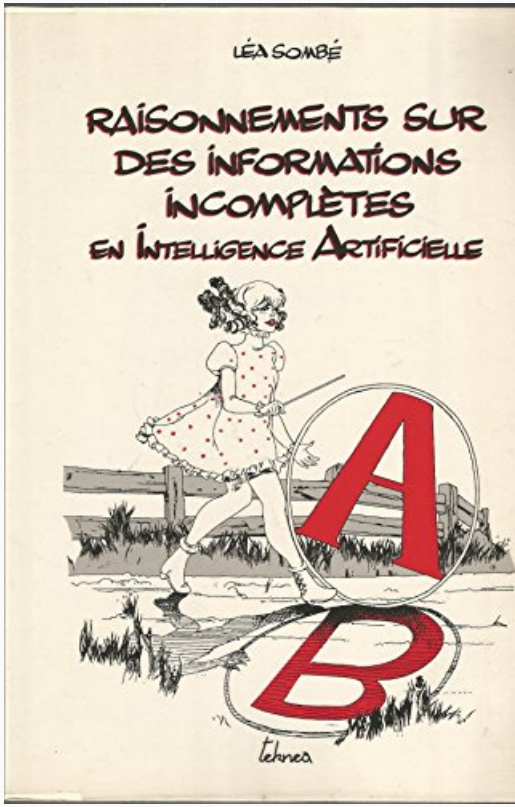
Léa Sombé alias Thea Arbee (avatar utilisé pour dire les histoires drôles de *(DuPr23)*)

No, I have no question, it was perfectly clear.

Only a comment: as I said in one of my previous talks entitled: "Is the superfluous a new modality?",

Incongruity is never superfluous!

See you soon for more incongruous adventures!



- Léa Sombé's avatar was created by Elisabetta Bevacqua (Lab-STICC - COMMEDIA)
- Léa's text was written by Henri Prade
- The assignment of gestures adapted to the text were made by Florence Dupin de Saint-Cyr

Pierre Bisquert and Florence Dupin de Saint-Cyr.
Working Memory for Assessment Under Inconsistency.
In Leila Amgoud and Richard Booth, editors, *NMR 2021 - 19th International Workshop on Non-Monotonic Reasoning*, Proceedings of NMR'2021 - 19th International Workshop on Non-Monotonic Reasoning, pages 149–158, Hanoï, Vietnam, November 2021.

Pierre Bisquert, Florence Dupin de Saint-Cyr, and Philippe Besnard.
Assessing Arguments with Schemes and Fallacies.
In Yuliya Lierler and Stefan Woltran, editors, *Logic Programming and Nonmonotonic Reasoning (LPNMR 2019), Philadelphia, PA, USA, 03/06/2019-07/06/2019*, volume 11481 of *Lecture Notes in Artificial Intelligence*, pages 61–74, https://link.springer.com, juin 2019. Springer.

Anne-Gwenn Bosser, Liana Ermakova, Florence Dupin de Saint-Cyr, Pierre de Loor, Victor Charpenay, Nicolas Pépin-Hermann, Benoît Alcaraz, Jean-Victor Autran, Alexandre Devillers, Juliette Grosset, Aymeric Hénard, and Florian Marchal.
Poetic or Humorous Text Generation: Jam Event at PFIA2022.

In Guglielmo Faggioli, Nicola Ferro, Allan Hanbury, and Martin Potthast, editors, *13th Conference and Labs of the Evaluation Forum (CLEF 2022)*, volume 3180 of *CEUR Workshop Proceedings*, pages 1719–1726, Bologna, Italy, September 2022. CEUR-WS.org.

Florence Dupin de Saint-Cyr, Nicolas Yannick Pepin, Julien Vianey, Nassim Mokhtary, Philippe Morignot, Anne-Gwenn Bosser, and Liana Ermakova. **Drivetogaether: a turnkey collaborative robotic event platform.** In *16th International Conference on Agents and Artificial Intelligence (ICAART 2024), Rome, Italy*, 2024.

Florence Dupin de Saint-Cyr. **Scenario Update Applied to Causal Reasoning (regular paper).** In Gerhard Brewka and Jerome Lang, editors, *International Conference on Principles of Knowledge Representation and Reasoning (KR 2008), Sydney, Australia, 16/09/08-19/09/08*, pages 188–197, http://www.aaai.org/Press/press.php, septembre 2008. AAAI Press.

Florence Dupin de Saint-Cyr and Francis Faux. **Integrating evolutionary prejudices in belief function theory.**

In Zied Bouraoui, Said Jabbour, and Srdjan Vesic, editors, *17th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2023)*, Lecture Notes in Artificial Intelligence (LNCS/LNAI), pages 400–414, Arras, France, September 2023. CRIL, Springer.

Florence Dupin de Saint-Cyr and Jérôme Lang.
Belief extrapolation (or how to reason about observations and unpredicted change).
Artificial Intelligence, 175:760–790, janvier 2011.

Florence Dupin de Saint-Cyr and Henri Prade.
Belief revision and incongruity: is it a joke?
Journal of Applied Non-Classical Logics, Special issue in honour of Philippe Besnard:1–28, August 2023.

Christian Godin.
Encyclopédie conceptuelle et thématique de la philosophie.
Champ Vallon, 2018.
ISBN 9791026708056.

Hirofumi Katsuno and Alberto O Mendelzon.
On the difference between updating a knowledge base and revising it.
KR, 91:387–394, 1991.

Hirofumi Katsuno and Alberto O Mendelzon.
Propositional knowledge base revision and minimal change.
Artificial Intelligence, 52(3):263–294, 1991.

Hervé Nègre.
***Dictionnaire des Histoires Drôles*.**
Fayard, 1970.

Pierre-Yves Raccah.
Humour et métaphore: quelques éléments d'une analogie pour la construction d'un sens inattendu: Illustration sur un corpus de citations de George Bernard Shaw.
Revue de Sémantique et Pragmatique, Presses de l'Univ. d'Orléans, 2015.

A. Tversky and D. Kahneman.

Judgment under uncertainty: Heuristics and biases.
Science, 185(4157):1124–1131, 1974.