Trust-based belief change

Emiliano Lorini[1] Guifei Jiang[1,2] **Laurent Perrussel**[1]

[1] IRIT – Université de Toulouse – France

[2] AIRG – University of Western Sydney – Penrith – Australia

ECAI-2014 talk

Motivation (1/3)

Communication among agents

- Agents are autonomous
  - Two agents may react differently while facing new information
    - *Deciding to adapt*
    - *Adapting belief*
  - Input $\langle sender, content \rangle$
- Receiver adapts its belief with respect to its *trust* in the *sender* about the *content*.
  1. Should it trust the sender (about the content)?
  2. Up to which "trust degree" it should consider new information?

***Our goal**: Exhibiting the interplay between trust and belief.*

Motivation (2/3)

Typical Interplay between trust and belief change

- Impact of the trust degree

  If NY Times informs Bill that Luigi's Burger is the best burger restaurant (lb) in NYC and Bill strongly trusts NYT about lb, then he should strongly believe lb

- Cumulative impact of the trust degree

  Jane trusts Trip Advisor, Hotels.com and Ebookers about Pine Hotel quality in a reasonable way. Trip Advisor, Hotels.com and Ebookers informs Jane that Pine Hotel-NYC is a bad hotel. Jane may strongly believe that Pine Hotel is a bad hotel.

$\Rightarrow$ Needs for a modular definition of the interplay

Motivation (3/3)

How do we do that?

- How to represent $\texttt{belief}(i, \mathit{content}, \mathit{degree})$ and $\texttt{trust}(i, , j, \mathit{content}, \mathit{degree})$ (Static aspect)
  - Epistemic Logic with a trust operator
  - Extension with degree
- How to represent $\texttt{inform}(i, j, \mathit{content})$ and $\texttt{revise}(k, \texttt{inform}(i, j, \mathit{content}))$ (Dynamic aspect)
  - Dynamic Epistemic Logic
  - Iterated Belief Change a la Spohn.

$\Rightarrow$ DL-BT: Dynamic Logic of Graded Belief and Trust

Plan

DL-BT: Syntax - Knowledge and Belief operators (1/2)

- $\mathsf{K}_i\varphi$: agent i knows φ

  What is possible for agent i?

- $\mathsf{B}_i^{\geq\alpha}\varphi$: agent i believes that φ is true with strength at least α

  "What is possible" is structured as an epistemic state.

- Scale for beliefs: numerical scale $Num = \{0, \ldots, \mathsf{max}\}$

  Scale is finite and can be viewed as an encoding of a qualitative scale

Example:

$Num = \{0, 1, 2, 3, 4, 5\}$ s.t. 0 stands for 'null' and 5 for 'very high'.
Bill believes (at least weakly) that Luigi's Burger is the best one in NYC:

$$\neg\mathsf{K}_{bill}\, luigi\_best\_burger \wedge \mathsf{B}_{bill}^{\geq 1}\, luigi\_best\_burger$$

DL-BT: Syntax - Knowledge and Belief operators (2/2)

- Knowledge Shortcut

$$\widehat{\mathsf{K}}_i\varphi =_{def} \neg\mathsf{K}_i\neg\varphi$$

- Belief shortcut

$$\mathsf{B}_i\varphi =_{def} \mathsf{B}_i^{\geq 1}\varphi \qquad \widehat{\mathsf{B}}_i\varphi =_{def} \neg\mathsf{B}_i\neg\varphi$$

$$\mathsf{U}_i\varphi =_{def} \neg\mathsf{B}_i\varphi \wedge \neg\mathsf{B}_i\neg\varphi \qquad \mathsf{B}_i^{\alpha}\varphi =_{def} \mathsf{B}_i^{\geq\alpha}\varphi \wedge \neg\mathsf{B}_i^{\geq(\alpha+1)}\varphi$$

$$\mathsf{B}_i^{\max}\varphi =_{def} \mathsf{B}_i^{\geq\max}\varphi \qquad \mathsf{B}_i^{0}\varphi =_{def} \neg\mathsf{B}_i\varphi$$

Example:
Bill believes that Luigi's Burger is the best one in NYC:

$$\mathsf{B}_{bill}\mathit{luigi\_best\_burger}$$

Bill only weakly believes that Luigi is the best one:

$$\mathsf{B}_{bill}^{1}\mathit{luigi\_best\_burger}$$

DL-BT: Syntax - Trust Operator

- $\mathsf{T}^{\alpha}_{i,j}\varphi$: agent i trusts agent j's judgement on formula φ with strength α.
  1. $\alpha \geqslant 0$
  2. Trust degree is exactly α (and not a lower bound)
  3. Scale is shared with the belief scale
- Shortcut

$$\mathsf{T}_{i,j}\varphi =_{def} \bigvee_{\alpha \in Num \setminus \{0\}} \mathsf{T}^{\alpha}_{i,j}\varphi$$

Example:
Bill strongly trusts NYT about Luigi:

$$\mathsf{T}^{4}_{bill,nyt} luigi\_best\_burger$$

DL-BT: Syntax - Revision Operator

- Interplay between trust and change may be specific to each agent: f is a policy change function.
- $[*_i^f \psi]\varphi$: after agent i has publicly announced that ψ is true and each agent j has revised her beliefs according to the trust-based belief change policy $f(j)$, φ is true.

Example: Bill believes that Luigi is the best after NYT announces it (w.r.t. some $f(bill)$):

$$[*_{nyt}^f luigi\_best\_burger]\mathsf{B}_{bill}\, luigi\_best\_burger$$

DL-BT: L-BT logic + revision operator

L-BT: Semantics - Structure (1/3)

- Kripke semantics
- For each agent i:
  - possible states: equivalence relation $\mathcal{E}_i$

    Example: $\mathcal{E}_i(w_0) = \{w_0, w_1, w_2, w_3, w_4, w_5\}$
  - ranking of the states: κ function
    - 0 is the best value
    - $\kappa(w, i)$: how exceptional is w for agent i

Example:

2	w_3, w_4, w_5
1	w_2
0	w_0, w_1

- for any state and agent, there is always a possible state with a 0 value (consistency).

L-BT: Semantics - Structure (2/3)

- Neighbourhood semantics for trust

  Non normal operator for handling trust about contradicting statements.

- For each agent i and state w:
  - j-trustable states: function $\mathcal{N}_{i,j}(w, \alpha)$

    Example:

$$\mathcal{N}_{i,j}(w_0, 1) = \{\{w_1\}\}$$
$$\mathcal{N}_{i,j}(w_0, 2) = \{\{w_3, w_4\}\}$$
$$\mathcal{N}_{i,j}(w_0, 3) = \{\{w_0, w_2\}\{w_0, w_5\}\}$$

- constraint on $\mathcal{N}_{i,j}$
  - no two states with different degrees
  - two equivalent possible states must lead to the trustable states (and values)
  - trustable states must be possible states

L-BT: Semantics - Structure (3/3)

- Valuations $\mathcal{V}$ on each state
- To sum up, a model

$$M = (W, \{\mathcal{E}_i\}_{i \in Agt}, \kappa, \{\mathcal{N}_{i,j}\}_{i,j \in Agt}, \mathcal{V})$$

- Exceptionality degree of a formula φ (w.r.t. some w and i):
  - $\kappa_{w,i}(\varphi) = \min_{v \in \|\varphi\|_{w,i}} \kappa(v, i)$ if $\|\varphi\|_{w,i} \neq \emptyset$
  - $\kappa_{w,i}(\varphi) = \max$ if $\|\varphi\|_{w,i} = \emptyset$

L-BT: Semantics - Truth conditions

Truth conditions are defined with respect to some model M and a state w.

$$M, w \models \mathsf{K}_i \varphi \textit{ iff } \forall v \in \mathcal{E}_i(w) : M, v \models \varphi$$

$$M, w \models \mathsf{B}_i^{\geq \alpha} \varphi \textit{ iff } \kappa_{w,i}(\neg\varphi) \geq \alpha$$

$$M, w \models \mathsf{T}_{i,j}^{\alpha} \varphi \textit{ iff } \|\varphi\|_M \in \mathcal{N}_{i,j}(w, \alpha)$$

DL-BT: - Embedding policies

- Policy: how to change degrees (κ values).
- Input: an L-BT model

$$M = (W, \{\mathcal{E}_i\}_{i \in Agt}, \kappa, \{\mathcal{N}_{i,j}\}_{i,j \in Agt}, \mathcal{V})$$

- Function f maps a policy to each agent.
- Output: new degrees

$\kappa^{*^f_i \varphi}(w, j)$: κ values for agent j revised w.r.t. M, f and initial announcement φ by i.

- ... and a new model

$$M^{*^f_i \varphi} = (W, \{\mathcal{E}_i\}_{i \in Agt}, \kappa^{*^f_i \varphi}, \{\mathcal{N}_{i,j}\}_{i,j \in Agt}, \mathcal{V})$$

- Semantics of the revision

$$M, w \models [*^f_i \varphi]\psi \text{ iff } M^{*^f_i \varphi}, w \models \psi$$

DL-BT Semantics: - Additive policy (1/4)

- Agent i announces φ
- Agent j changes its belief only if it trusts i about φ
- If so ($\mathsf{T}_{j,i}\varphi$ holds at w)
  - φ also holds on w: agent j decreases the degree of exceptionality of w:

$$\kappa^{*}{}_{i}^{add}\varphi(w,j) = \kappa(w,j) - \kappa_{w,j}(\varphi)$$

  - φ does not hold on w: agent j increases the degree of exceptionality of w w.r.t. to its degree of trust ($\mathsf{T}^{\alpha}_{j,i}\varphi$ holds at w):

$$\kappa^{*}{}_{i}^{add}\varphi(w,j) = \kappa(w,j) + \alpha \textit{ bounded to } \mathsf{Max}$$

DL-BT Semantics: - Additive policy (2/4)

Example

- Suppose three states w_0, w_1 and w_2, 2 agents *bill* and *nyt* and statement *Luigi_Best_Restaurant* (lbr).
- *lbr* holds in w_0 but not in w_1 and w_2.
- $\kappa(w_0, bill) = 0; \kappa(w_1, bill) = 1; \kappa(w_2, bill) = 2$
- $\kappa_{w_0,bill}(lbr) = 0; \kappa_{w_1,bill}(lbr) = 0; \kappa_{w_2,bill}(lbr) = 0$
- $\kappa_{w_0,bill}(\neg lbr) = 1; \kappa_{w_1,bill}(\neg lbr) = 1; \kappa_{w_2,bill}(\neg lbr) = 1$
- $\mathcal{N}_{bill,nyt}(2, w_0) = \mathcal{N}_{bill,nyt}(2, w_1) = \mathcal{N}_{bill,nyt}(2, w_2) = \{\{w_0\}\}$
- revising κ for an *lbr* state

$$\kappa^{*^{add}_{nyt}\, lbr}(w_0, bill) = \kappa(w_0, bill) - \kappa_{w_0,bill}(lbr) = 0 - 0 = 0$$

- revising κ for an $\neg lbr$ state

$$\kappa^{*^{add}_{nyt}\, lbr}(w_1, bill) = \kappa(w_1, bill) + \alpha = 1 + 2 = 3 \Rightarrow 2$$

DL-BT Semantics: - Additive policy (3/4)

Key Properties

- Syntax independence: if $\models \varphi_1 \leftrightarrow \varphi_2$ then

$$\models [*_i^f \varphi_1]\psi \leftrightarrow [*_i^f \varphi_2]\psi$$

- Keeping beliefs (objective formulas only)

$$\models (\mathsf{B}_j^\alpha \psi \wedge \neg \mathsf{T}_{j,i}\varphi) \rightarrow [*_i^f \varphi]\mathsf{B}_j^\alpha \psi$$
$$\models \mathsf{T}_{j,i}\psi \rightarrow [*_i^f \psi]\mathsf{B}_j \psi$$

- Increasing degrees (objective formulas only)

$$\models (\mathsf{T}_{j,i}^\alpha \varphi \wedge \mathsf{B}_j^\beta \varphi) \rightarrow [*_i^f \varphi]\mathsf{B}_j^{\alpha+\beta} \varphi$$

- Priority to last input (objective formulas only)

$$\models \mathsf{T}_{j,i_2}\psi \rightarrow [*_{i_1}^f \neg\psi][*_{i_2}^f \psi]\mathsf{B}_j \psi$$

DL-BT Semantics: - Additive policy (4/4)

Example

- Bill has to decide whether he buys a certain stock and he is currently uncertain:

$$Hyp1 =_{def} \mathsf{U}_{Bill} stockUp$$

- Bill trusts *fairly* Mary's judgement on *stockUp*, and Bill trusts *very weakly* Jack's judgement on *stockUp*:

$$Hyp2 =_{def} \mathsf{T}^{3}_{Bill,Mary} stockUp \wedge \mathsf{T}^{1}_{Bill,Jack} stockUp$$

- Mary and Jack announces that stock will go up:

$$Hyp1 \wedge Hyp2 \rightarrow [*^{f}_{Mary} stockUp][*^{f'}_{Jack} stockUp]\mathsf{B}^{4}_{Bill} stockUp.$$

- But, it also holds

$$Hyp1 \wedge Hyp2 \rightarrow [*^{f}_{Mary} \neg stockUp][*^{f'}_{Jack} stockUp]\mathsf{B}_{Bill} stockUp.$$

DL-BT Semantics: - Compensatory policy (1/3)

- Avoiding the priority to last announcement
- Agent j changes its belief only if it trusts i about φ
- If so ($\mathsf{T}_{j,i}\varphi$ holds at w)
  - φ also holds on w: agent j decreases the degree of exceptionality of w w.r.t. to the degree α of trust

$$\kappa^{*}{}_{i}^{add}\varphi(w,j) = \kappa(w,j) - \alpha \textit{ bounded to } 0.$$

  - φ does not hold on w: agent j increases the degree of exceptionality of w w.r.t. to its degree of trust ($\mathsf{T}^{\alpha}_{j,i}\varphi$ holds at w):

$$\kappa^{*}{}_{i}^{add}\varphi(w,j) = \kappa(w,j) + \alpha \textit{ bounded to } \mathsf{Max}$$

Case is limited to the situation where $\widehat{\mathsf{B}}_i\varphi$ also holds

DL-BT Semantics: - Compensatory policy (2/3)

Key Properties

- Cumulative effect

$$\models (\mathsf{T}^{\alpha}_{j,i}\varphi \wedge \mathsf{B}^{\beta}_{j}\varphi) \rightarrow [*^{f}_{i}\varphi]\mathsf{B}^{\alpha+\beta}_{j}\varphi$$

- Compensatory effect (one shot)

$$\models (\mathsf{T}^{\alpha}_{j,i}\neg\varphi \wedge \mathsf{B}^{\beta}_{j}\varphi) \rightarrow [*^{f}_{i}\neg\varphi]\mathsf{B}^{\beta-\alpha}_{j}\varphi$$

- multiple change and compensatory effect

$$\models (\mathsf{T}^{\alpha_1}_{j,i_1}\varphi \wedge \mathsf{T}^{\alpha_2}_{j,i_2}\neg\varphi \wedge \mathsf{U}_j\varphi) \rightarrow [*^{f}_{i_1}\varphi][*^{f'}_{i_2}\neg\varphi]\mathsf{B}^{\alpha_1-\alpha_2}_{j}\varphi$$

DL-BT Semantics: - Compensatory policy (3/3)

- Bill has to decide whether he buys a certain stock and he is currently uncertain:

$$Hyp1 =_{def} \mathsf{U}_{Bill} stockUp$$

- Suppose that Mary and Jack provide contradictory information about proposition *stockUp*.
- Bill trusts *fairly* Mary's judgement on *stockUp* while Bill trusts *very weakly* Jack's judgement on $\neg stockUp$:

$$Hyp2 =_{def} \mathsf{T}^{3}_{Bill,Mary} stockUp \wedge \mathsf{T}^{1}_{Bill,Jack} \neg stockUp$$

- Now, assume Mary announces that *stockUp* is true and Jack announces that *stockUp* is false ($f(Bill), f'(Bill) = \{comp\}$).

$$\models (Hyp1 \wedge Hyp2') \rightarrow [*^{f}_{Mary} stockUp][*^{f'}_{Jack} \neg stockUp]\mathsf{B}^{2}_{Bill} stockUp.$$

Axiomatics

- 2 steps definition
  1. Axiomatics for L-BT logics
  2. Reduction axiom for DL-BT logics
- Additional axioms for the specifications of the policies

Proof Theory for L-BT logic (1/2)

- K,T, 4 and 5 for K_i
- K, D for $\mathsf{B}_i^{\geq\alpha}$
- Interplay axioms between K_i and $\mathsf{B}_i^{\geq\alpha}$
  - $\mathsf{B}_i^{\geq\alpha}\varphi \rightarrow \mathsf{K}_i\mathsf{B}_i^{\geq\alpha}\varphi$
  - $\neg\mathsf{B}_i^{\geq\alpha}\varphi \rightarrow \mathsf{K}_i\neg\mathsf{B}_i^{\geq\alpha}\varphi$
  - $\mathsf{K}_i\varphi \rightarrow \mathsf{B}_i^{\geq\alpha}\varphi$
  - $\mathsf{B}_i^{\geq\alpha+1}\varphi \rightarrow \mathsf{B}_i^{\geq\alpha}\varphi$

Proof Theory for L-BT logic (2/2)

- Graded Trust axioms
  - $\mathsf{T}^{\alpha}_{i,j}\varphi \rightarrow \neg\mathsf{T}^{\beta}_{i,j}\varphi$ if $\alpha \neq \beta$
  - $\mathsf{T}^{\alpha}_{i,j}\varphi \rightarrow \mathsf{K}_i\mathsf{T}^{\alpha}_{i,j}\varphi$
  - $\mathsf{T}^{\alpha}_{i,j}\varphi \rightarrow \widehat{\mathsf{K}}_i\varphi$
- Inference rules for Graded Trust
  - From $\varphi \leftrightarrow \psi$ infer $\mathsf{T}^{\alpha}_{i,j}\varphi \leftrightarrow \mathsf{T}^{\alpha}_{i,j}\psi$

Reduction axioms for DL-BT logic

- Reduction axioms are similar to DEL:

$$
\begin{array}{lcl}
[*^f_j\,\varphi]p & \leftrightarrow & p \\
[*^f_j\,\varphi]\neg\psi & \leftrightarrow & \neg[*^f_j\,\varphi]\psi \\
[*^f_j\,\varphi](\psi_1 \wedge \psi_2) & \leftrightarrow & ([*^f_j\,\varphi]\psi_1 \wedge [*^f_j\,\varphi]\psi_2) \\
[*^f_j\,\varphi]\mathsf{K}_i\psi & \leftrightarrow & \mathsf{K}_i[*^f_j\,\varphi]\psi \\
[*^f_j\,\varphi]\mathsf{T}^\alpha_{i,k}\psi & \leftrightarrow & \mathsf{T}^\alpha_{i,k}[*^f_j\,\varphi]\psi
\end{array}
$$

Theorem: Soundness and completeness for DL-BT

Reduction axioms for Change Policy (Illustration)

- Additive policy

$$
\begin{aligned}
[*^f_j\,\varphi]\mathsf{B}_i^{\geq\alpha}\psi \quad \leftrightarrow \Big(& (\neg \mathsf{T}_{i,j}\varphi \rightarrow \mathsf{B}_i^{\geq\alpha}[*^f_j\,\varphi]\psi) \wedge \\
& \bigwedge_{\beta\in Num\setminus\{0\},\gamma_1\in Num} ((\mathsf{T}^{\beta}_{i,j}\varphi \wedge \mathsf{B}_i^{\gamma_1}\neg\varphi) \rightarrow \\
& (\mathsf{B}_i^{\geq\alpha+\gamma_1}(\varphi \rightarrow [*^f_j\,\varphi]\psi) \wedge \\
& \mathsf{B}_i^{\geq\alpha-\beta}(\neg\varphi \rightarrow [*^f_j\,\varphi]\psi))) \Big)
\end{aligned}
$$

Theorem: Soundness and completeness for DL-BT^{add} (idem for DL-BT^{comp} and thus DL-BT^{add,comp}

Conclusion

- Key results
  - Modular definition of interplay between trust and belief change.
  - Family of logics DL-BT
  - Sound and complete definition for two policies (additive and compensatory)
- Future work
  - In depth definition of the interplay (loop of announcements)
  - New policies
  - Representation theorems