Trust-based belief change

Emiliano Lorini\textsuperscript{1} \quad Guifei Jiang\textsuperscript{1,2} \quad Laurent Perrussel\textsuperscript{1}

\textsuperscript{1}IRIT – Université de Toulouse – France
\textsuperscript{2}AIRG – University of Western Sydney – Penrith – Australia

ECAI-2014 talk
Communication among agents

- Agents are autonomous
  - Two agents may react differently while facing new information
    - *Deciding to adapt*
    - *Adapting belief*
  - Input $\langle \text{sender}, \text{content} \rangle$

- 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.*
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.

⇒ Needs for a modular definition of the interplay
Motivation (3/3)

How do we do that?

- How to represent belief($i$, $content$, $degree$) and trust($i$, $j$, $content$, $degree$) (Static aspect)
  - Epistemic Logic with a trust operator
  - Extension with degree

- How to represent inform($i$, $j$, $content$) and revise($k$, inform($i$, $j$, $content$)) (Dynamic aspect)
  - Dynamic Epistemic Logic
  - Iterated Belief Change a la Spohn.

⇒ DL-BT: Dynamic Logic of Graded Belief and Trust
## Plan

1. **Motivation**
2. **DL-BT: Syntax**
3. **L-BT: Semantics**
   - Structure
   - Truth conditions
4. **DL-BT Semantics**
   - Change Policies
   - Additive Policy
   - Compensatory Policy
5. **Axiomatics**
   - L-BT: Proof theory
   - DL-BT: Reduction axioms
   - Change Policies: Reduction axioms
6. **Conclusion**
DL-BT: Syntax - Knowledge and Belief operators (1/2)

- $K_i \varphi$: agent $i$ knows $\varphi$

  What is possible for agent $i$?

- $B_i^{\geq \alpha} \varphi$: agent $i$ believes that $\varphi$ is true with strength at least $\alpha$

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

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

  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:

$$\neg K_{bill}luigi\_best\_burger \land B_{bill}^{\geq 1}luigi\_best\_burger$$
Knowledge Shortcut

\[ \hat{K}_i \varphi = \text{def} \neg K_i \neg \varphi \]

Belief shortcut

\[ B_i \varphi = \text{def} B_i^{\geq 1} \varphi \]
\[ \hat{B}_i \varphi = \text{def} \neg B_i \neg \varphi \]
\[ U_i \varphi = \text{def} \neg B_i \varphi \land \neg B_i \neg \varphi \]
\[ B_i^{\alpha} \varphi = \text{def} B_i^{\geq \alpha} \varphi \land \neg B_i^{\geq (\alpha + 1)} \varphi \]
\[ B_i^{\max} \varphi = \text{def} B_i^{\geq \max} \varphi \]
\[ B_i^0 \varphi = \text{def} \neg B_i \varphi \]

Example:

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

\[ B_{\text{bill}} \text{\_luigi\_best\_burger} \]

Bill only weakly believes that Luigi is the best one:

\[ B_{\text{bill}}^1 \text{\_luigi\_best\_burger} \]
**DL-BT: Syntax - Trust Operator**

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

- **Shortcut**

  $$ T_{i,j} \varphi = \text{def} \bigvee_{\alpha \in \text{Num}\setminus\{0\}} T_{i,j}^\alpha \varphi $$

**Example:**
Bill strongly trusts NYT about Luigi:

$$ T_{\text{bill}, \text{nyt} \_ \text{luigi}_\_ \text{best}_\_ \text{burger}}^4 $$

**L-BT: Logic of Graded Belief and Trust**

E. Lorini, G. Jiang, L. Perrussel
Interplay between trust and change may be specific to each agent: $f$ is a policy change function.

$[\ast_i^f \psi ] \varphi$: after agent $i$ has publicly announced that $\psi$ is true and each agent $j$ has revised her beliefs according to the trust-based belief change policy $f(j)$, $\varphi$ is true.

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

$$[\ast_{nyt}^{f} luigi\_best\_burger ]B_{bill} luigi\_best\_burger$$

**DL-BT: L-BT logic + revision operator**
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: \( \kappa \) function
  - 0 is the best value
  - \( \kappa(w, i) \): how exceptional is \( w \) for agent \( i \)

Example:

<table>
<thead>
<tr>
<th>2</th>
<th>( w_3, w_4, w_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( w_2 )</td>
</tr>
<tr>
<td>0</td>
<td>( w_0, w_1 )</td>
</tr>
</tbody>
</table>

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

Non normal operator for handling trust about contradicting statements.

For each agent $i$ and state $w$:

- $j$-trustable states: function $N_{i,j}(w, \alpha)$

**Example:**

- $N_{i,j}(w_0, 1) = \{ \{ w_1 \} \}$
- $N_{i,j}(w_0, 2) = \{ \{ w_3, w_4 \} \}$
- $N_{i,j}(w_0, 3) = \{ \{ w_0, w_2 \}, \{ w_0, w_5 \} \}$

**constraint on $N_{i,j}$**

- no two states with different degrees
- two equivalent possible states must lead to the trustable states (and values)
- trustable states must be possible states
Valuations $\mathcal{V}$ on each state

To sum up, a model

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

Exceptionality degree of a formula $\varphi$ (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 $\|\varphi\|_{w,i} = \emptyset$
Truth conditions are defined with respect to some model $M$ and a state $w$.

\[ M, w \models K_i \varphi \iff \forall v \in \mathcal{E}_i(w) : M, v \models \varphi \]

\[ M, w \models B_{i}^{\geq \alpha} \varphi \iff \kappa_{w, i}(\neg \varphi) \geq \alpha \]

\[ M, w \models T_{i, j}^{\alpha} \varphi \iff \| \varphi \|_M \in \mathcal{N}_{i, j}(w, \alpha) \]
DL-BT: Embedding policies

- **Policy**: how to change degrees ($\kappa$ values).
- **Input**: an L-BT model
  \[ M = (W, \{E_i\}_{i \in \text{Agt}}, \kappa, \{N_{i,j}\}_{i,j \in \text{Agt}}, V) \]
- **Function** $f$ maps a policy to each agent.
- **Output**: new degrees
  \[ \kappa^*_i f \varphi(w,j): \kappa \text{ values for agent } j \text{ revised w.r.t. } M, f \text{ and initial announcement } \varphi \text{ by } i. \]

- ... and a new model
  \[ M^*_i f \varphi = (W, \{E_i\}_{i \in \text{Agt}}, \kappa^*_i f \varphi, \{N_{i,j}\}_{i,j \in \text{Agt}}, V) \]
- **Semantics** of the revision
  \[ M, w \models [*_i f \varphi] \psi \iff M^*_i f \varphi, w \models \psi \]
Agent $i$ announces $\varphi$

Agent $j$ changes its belief only if it trusts $i$ about $\varphi$

If so ($T_{j,i} \varphi$ holds at $w$)

- $\varphi$ also holds on $w$: agent $j$ decreases the degree of exceptionality of $w$:
  \[
  \kappa_{i}^{\text{add}} \varphi(w,j) = \kappa(w,j) - \kappa_{w,j}(\varphi)
  \]

- $\varphi$ does not hold on $w$: agent $j$ increases the degree of exceptionality of $w$ w.r.t. to its degree of trust ($T_{j,i}^\alpha \varphi$ holds at $w$):
  \[
  \kappa_{i}^{\text{add}} \varphi(w,j) = \kappa(w,j) + \alpha \text{ bounded to Max}
  \]
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 $\kappa$ for an $lbr$ state
  \[
  \kappa^{*_{\text{add}}}_{\text{nayt}}(w_0, bill) = \kappa(w_0, bill) - \kappa_{w_0, bill}(lbr) = 0 - 0 = 0
  \]
- revising $\kappa$ for an $\neg lbr$ state
  \[
  \kappa^{*_{\text{add}}}_{\text{nayt}}(w_1, bill) = \kappa(w_1, bill) + \alpha = 1 + 2 = 3 \Rightarrow 2
  \]
DL-BT Semantics: - Additive policy (3/4)

Key Properties

- Syntax independence: if $\models \varphi_1 \leftrightarrow \varphi_2$ then
  
  $$
  \models [*_i \varphi_1] \psi \leftrightarrow [*_i \varphi_2] \psi
  $$

- Keeping beliefs (objective formulas only)
  
  $$
  \models (B^\alpha_j \psi \land \neg T_{j,i} \varphi) \rightarrow [*_i \varphi] B^\alpha_j \psi
  $$

  $$
  \models T_{j,i} \psi \rightarrow [*_i \psi] B_j \psi
  $$

- Increasing degrees (objective formulas only)
  
  $$
  \models (T^\alpha_j \varphi \land B^\beta_j \varphi) \rightarrow [*_i \varphi] B^{\alpha+\beta}_j \varphi
  $$

- Priority to last input (objective formulas only)
  
  $$
  \models T_{j,i_2} \psi \rightarrow [*_i \neg \psi][*_i_2 \psi] B_j \psi
  $$
Example

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

\[ Hyp1 = \text{def } U_{\text{Bill stockUp}} \]

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

\[ Hyp2 = \text{def } T^3_{\text{Bill, Mary stockUp}} \land T^1_{\text{Bill, Jack stockUp}} \]

- Mary and Jack announces that stock will go up:

\[ Hyp1 \land Hyp2 \rightarrow [\ast^f_{\text{Mary stockUp}}][\ast^{f'}_{\text{Jack stockUp}}]B^4_{\text{Bill stockUp}}. \]

- But, it also holds

\[ Hyp1 \land Hyp2 \rightarrow [\ast^f_{\text{Mary stockUp}}][\ast^{f'}_{\text{Jack stockUp}}]B_{\text{Bill stockUp}}. \]
Avoiding the priority to last announcement

Agent \( j \) changes its belief only if it trusts \( i \) about \( \varphi \)

If so (\( T_{j;i} \varphi \) holds at \( w \))

- \( \varphi \) also holds on \( w \): agent \( j \) decreases the degree of exceptionality of \( w \) w.r.t. to the degree \( \alpha \) of trust

\[
\kappa^{\text{add}}_{i} \varphi(w, j) = \kappa(w, j) - \alpha \text{ bounded to } 0.
\]

- \( \varphi \) does not hold on \( w \): agent \( j \) increases the degree of exceptionality of \( w \) w.r.t. to its degree of trust (\( T_{j;i}^{\alpha} \varphi \) holds at \( w \)):

\[
\kappa^{\text{add}}_{i} \varphi(w, j) = \kappa(w, j) + \alpha \text{ bounded to } \text{Max}
\]

Case is limited to the situation where \( \hat{B}_{i} \varphi \) also holds
**Key Properties**

- **Cumulative effect**

\[ \models (T_{j,i}^\alpha \varphi \land B_j^\beta \varphi) \rightarrow [*_i^f \varphi] B_j^{\alpha + \beta} \varphi \]

- **Compensatory effect (one shot)**

\[ \models (T_{j,i}^\alpha \neg \varphi \land B_j^\beta \varphi) \rightarrow [*_i^f \neg \varphi] B_j^{\beta - \alpha} \varphi \]

- **Multiple change and compensatory effect**

\[ \models (T_{j,i_1}^\alpha \varphi \land T_{j,i_2}^\alpha \neg \varphi \land U_j \varphi) \rightarrow [*_i^f \varphi][*_i' \neg \varphi] B_j^{\alpha_1 - \alpha_2} \varphi \]
Bill has to decide whether he buys a certain stock and he is currently uncertain:

\[ Hyp1 = \text{def } U_{Bill} stockUp \]

Suppose that Mary and Jack provide contradictory information about proposition \( stockUp \).

Bill trusts \textit{fairly} Mary’s judgement on \( stockUp \) while Bill trusts \textit{very weakly} Jack’s judgement on \( \neg stockUp \):

\[ Hyp2 = \text{def } T^3_{Bill, Mary} stockUp \land T^1_{Bill, Jack} \neg stockUp \]

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

\[ |= (Hyp1 \land Hyp2') \rightarrow [^f_{Mary} stockUp][^f'_{Jack} \neg stockUp]B^2_{Bill} stockUp. \]
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 $B_i^{\geq \alpha}$
- Interplay axioms between $K_i$ and $B_i^{\geq \alpha}$
  - $B_i^{\geq \alpha} \varphi \rightarrow K_i B_i^{\geq \alpha} \varphi$
  - $\neg B_i^{\geq \alpha} \varphi \rightarrow K_i \neg B_i^{\geq \alpha} \varphi$
  - $K_i \varphi \rightarrow B_i^{\geq \alpha} \varphi$
  - $B_i^{\geq \alpha+1} \varphi \rightarrow B_i^{\geq \alpha} \varphi$
Proof Theory for L-BT logic (2/2)

- Graded Trust axioms
  - $T_{i,j}^\alpha \varphi \rightarrow \neg T_{i,j}^\beta \varphi$ if $\alpha \neq \beta$
  - $T_{i,j}^\alpha \varphi \rightarrow K_i T_{i,j}^\alpha \varphi$
  - $T_{i,j}^\alpha \varphi \rightarrow \hat{K}_i \varphi$

- Inference rules for Graded Trust
  - From $\varphi \leftrightarrow \psi$ infer $T_{i,j}^\alpha \varphi \leftrightarrow T_{i,j}^\alpha \psi$
Reduction axioms for DL-BT logic

- Reduction axioms are similar to DEL:
  
  \[
  \begin{align*}
  [*_j^f \phi] p & \iff p \\
  [*_j^f \phi] \neg \psi & \iff \neg [*_j^f \phi] \psi \\
  [*_j^f \phi](\psi_1 \land \psi_2) & \iff ([*_j^f \phi] \psi_1 \land [*_j^f \phi] \psi_2) \\
  [*_j^f \phi] K_i \psi & \iff K_i[*_j^f \phi] \psi \\
  [*_j^f \phi] T_{i,k}^\alpha \psi & \iff T_{i,k}^\alpha [*_j^f \phi] \psi
  \end{align*}
  \]

Theorem: Soundness and completeness for DL-BT
Additive policy

\[ [\ast_j \phi]B_i^{\geq \alpha} \psi \leftrightarrow \left( \neg T_{i,j} \phi \rightarrow B_i^{\geq \alpha} [\ast_j \phi] \psi \right) \land \]
\[ \land_{\beta \in \text{Num} \setminus \{0\}, \gamma_1 \in \text{Num}} \left( (T_{i,j}^\beta \phi \land B_i^{\gamma_1} \neg \phi) \rightarrow \right. \]
\[ (B_i^{\geq \alpha + \gamma_1} (\phi \rightarrow [\ast_j \phi] \psi) \land \]
\[ B_i^{\geq \alpha - \beta} (\neg \phi \rightarrow [\ast_j \phi] \psi)))) \right) \]

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