Public Group Announcements and Trust in Doxastic Logic

Elise Perrotin[1], Rustam Galimullin[2], Quentin Canu[2], and Natasha Alechina[3]

[1]: Institut de Recherche en Informatique de Toulouse
[2]: Ecole Normale Supérieure Paris-Saclay
[3]: University of Nottingham

October 18, 2019

Friend 1

Writing a paper for LORI

Friend 1

Writing a paper for LORI

Friend 2

"The paper will not be accepted"

Friend 1

Writing a paper for LORI

Friend 2

"The paper will not be accepted"

Friend 3

"Friend 1 thinks that their paper will be accepted"

Friend 1

Writing a paper for LORI

Friend 2

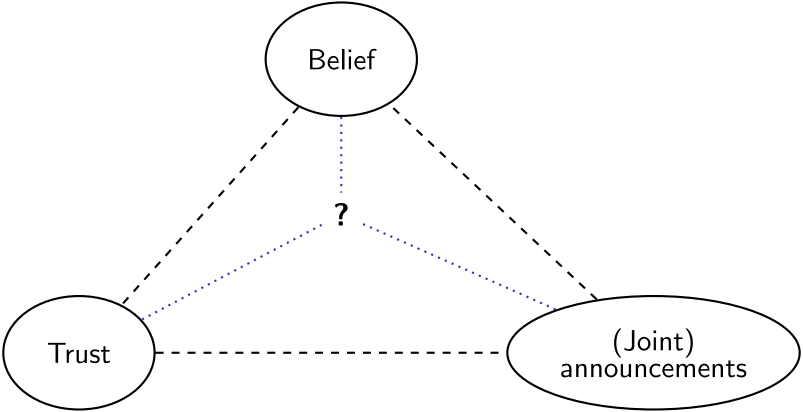
"The paper will not be accepted"

Friend 3

"Friend 1 thinks that their paper will be accepted"

You

???

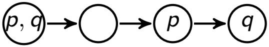

Belief
?
Trust
(Joint)
announcements

Beliefs and announcements

How should we represent beliefs?

Two standard approaches:

- Possible worlds (DEL)



- Belief bases (AGM)

$$BB_a = \{q, \neg p\}$$

Belief bases

- Compact
- Less expressive, but allows us to focus on actual beliefs
- Multiagent: one belief base per agent
- Belief revision (Dynamic Doxastic Logic)
- Contradictory announcements: paraconsistent logic, belief merging

Belief bases

- Compact
- Less expressive, but allows us to focus on actual beliefs
- Multiagent: one belief base per agent
- Belief revision (Dynamic Doxastic Logic)
- Contradictory announcements: paraconsistent logic, belief merging

▶ We're actually going to be using simplified belief bases, in which identifying contradictions is fairly straightforward.

A simple notion of trust

Our meaning of trust: belief that what the other believes is true (i.e. trust in the *reliability*/*expertise* of other agents).
We suppose that agents are honest: they believe in what they announce.
Our trust is:

- Binary (does not depend on the topic of a formula + no gradual trust)
- Memoryless
- Can only be lost

SDDLT: Syntactic Dynamic Doxastic Logic with Trust

If $a, b \in Agt$ and $p \in \mathcal{P}$:

$$\mathcal{L}_B \ni \varepsilon ::= \top \mid p \mid \neg p \mid B_a \varepsilon$$
$$\mathcal{L}_{\text{SDDLT}} \ni \varphi ::= T_{a,b} \mid \varepsilon \mid \neg\varphi \mid \varphi \wedge \varphi \mid [A]\varphi$$

where an announcement A is a partial function $Agt \to 2^{\mathcal{L}_B}$.

Definition (Belief states and trust models)

A belief state is a tuple $s = (\{BB_a^s\}_{a \in Agt}, \{T_a^s\}_{a \in Agt})$, where:

- $BB_a^s \subset \mathcal{L}_B$ (*belief base* of a at s)
- $T_a^s \subseteq Agt$ (*trust set* of a at s)

A *trust model* is a pair (S, V) where S is the set of all belief states and $V \subseteq \mathcal{P}$ is the actual state of the world.

We require $\top \in BB_a^s$ and $a \in T_a^s$.

Trust and belief

Each agent can reason depending on who they trust:

$$\varepsilon_1 \to_a^s \varepsilon_2 \text{ iff } \varepsilon_1 = B_{a_1} \dots B_{a_k} \varepsilon_2 \text{ for some } a_1, \dots, a_k \text{ in } T_a^s.$$

BB_a^s must be closed under $\to_a^s$.

Example

If $T_a^s = \{b, c\}$ and a believes $B_b B_c B_d p$, then a should also believe $B_c B_d p$ and $B_d p$, but not necessarily p.

We denote by $Cl_a^s(B)$ the closure of a set B under $\to_a^s$.

Belief consistency

Definition

If $\varepsilon_1 = B_{a_1} \dots B_{a_n} p$ and $\varepsilon_2 = B_{a_1} \dots B_{a_n} \neg p$, then we write $\varepsilon_1 = \overline{\varepsilon_2}$ and $\varepsilon_2 = \overline{\varepsilon_1}$.

Belief bases must be *consistent*, that is, they should not contain both ε and $\overline{\varepsilon}$ for any formula ε of $\mathcal{L}_B$.

Identifying contradictions

Some notions about contradicitions:

- ε_1 and ε_2 are *contradictory according to a at s* if $Cl_a^s(\{\varepsilon_1, \varepsilon_2\})$ is inconsistent (e.g. $B_c B_d p$, $B_b B_d \neg p$ when $b, c \in T_a^s$);
- ε *contradicts a's beliefs at s* if it contradicts a formula of BB_a^s;
- ε is *supported by a's beliefs at s* if $\overline{\varepsilon}$ contradicts them;
- ε is *neutral w.r.t. a's beliefs at s* if it neither contradicts nor is supported by them.

Identifying contradictions

Theorem

Let ε and ε' be two formulas of $\mathcal{L}_B$, a an agent, and s a belief state.

1. *ε and ε' are contradictory according to a at s iff $\min_a^s(\varepsilon) = \overline{\min_a^s(\varepsilon')}$.*
2. *ε contradicts a's beliefs at s iff $\overline{\min_a^s(\varepsilon)} \in BB_a^s$, and ε is supported by a's beliefs at s iff $\min_a^s(\varepsilon) \in BB_a^s$.*

where $\min_a^s(\varepsilon)$ is the shortest suffix ε' of ε such that $\varepsilon \rightarrow_a^s \varepsilon'$.

Updating trust

If b announces ε_b and c announces ε_c:

- a learns that $B_b\varepsilon_b$ and $B_c\varepsilon_c$
- If $B_b\varepsilon_b$ and $B_c\varepsilon_c$ are contradictory according to a:
  - If $B_b\varepsilon_b$ contradicts a's beliefs (equivalent to $B_c\varepsilon_c$ being supported by a's beliefs), then b is no longer trusted;
  - if $B_c\varepsilon_c$ contradicts a's beliefs, c is no longer trusted;
  - otherwise, a loses trust in both b and c.

Updating trust

Definition (Update of trust)

After an announcement A by group G:

$$T_a^{s\cdot[A]} = T_a^s \setminus \{b \in G \mid \exists c \in G, \varepsilon_b \in A(b), \varepsilon_c \in A(c), \min_a^s(B_b\varepsilon_b) = \overline{\min_a^s(B_c\varepsilon_c)} \text{ and } \min_a^s(B_b\varepsilon_b) \notin BB_a^s\}$$

Updating beliefs

- Add to BB_a^s all $B_b\varepsilon_b$ where $\varepsilon_b \in A(b)$;
- Give priority to the new information: remove formulas of BB_a^s which still contradict the announcement after updating trust.

Definition (Update of beliefs)

After an announcement A by group G:

$$BB_a^{s\cdot[A]} = Cl_a^{s\cdot[A]}((BB_a^s \setminus \{\varepsilon \in BB_a^s \mid \exists b \in G, \exists \varepsilon_b \in A(b), \min_a^{s\cdot[A]}(\varepsilon) = \min_a^{s\cdot[A]}(\overline{B_b\varepsilon_b})\}) \cup \{B_b\varepsilon_b \mid b \in G \text{ and } \varepsilon_b \in A(b)\})$$

Semantics

Definition (Semantics of SDDLT)

If (M, s) is a pointed trust model and A is an announcement by a group $G \subseteq Agt$:

$$
\begin{aligned}
(M, s) &\models B_a\varepsilon && \text{iff } \varepsilon \in BB_a^s \\
(M, s) &\models T_{a,b} && \text{iff } b \in T_a^s \\
(M, s) &\models [A]\varphi && \text{iff } ((M, s) \models B_G A \Rightarrow (M, s \cdot [A]) \models \varphi)
\end{aligned}
$$

where $s \cdot [A] = (\{BB_a^{s \cdot [A]}\}_{a \in Agt}, \{T_a^{s \cdot [A]}\}_{a \in Agt})$ and $B_G A := \bigwedge_{g \in G} \bigwedge_{\varepsilon_g \in A(g)} B_g \varepsilon_g$.

Example

Example

If $Agt = \{a, b, c, d, e\}$, $T_a^s = \{a, b, c, d\}$ and $BB_a^s = \{p\}$:

- If b announces $B_d p$ and c announces $\neg p$, then b's statement is supported by a's beliefs: $T_a^{s \cdot A_1} = \{a, b, d\}$ and $BB_a^{s \cdot A_1} = \{p, B_b B_d p, B_d p, B_c \neg p\}$.
- If b announces q and c announces $\neg q$, a loses trust in both b and c: $T_a^{s \cdot A_2} = \{a, d\}$ and $BB_a^{s \cdot A_2} = \{p, B_b q, B_c \neg q\}$.
- If c announces $\neg p$ and e announces p, there is no conflict according to a, and c's announcement takes precedence over a's previous beliefs: $T_a^{s \cdot A_3} = \{a, b, c, d\}$ and $BB_a^{s \cdot A_3} = \{\neg p, B_c \neg p, B_e p\}$.

Group and coalition announcements

If we restrict the depth of nestings of belief operators in announcements to m, we can write:

$$\langle G\rangle\varphi \leftrightarrow \bigvee_{A\in PA(G,\varphi)}\Big(\bigwedge_{\substack{g\in G\\ \varepsilon_g\in A(g)}} B_g\varepsilon_g \wedge [A]\varphi\Big)$$

$$\langle\!\![G]\!\!\rangle\varphi \leftrightarrow \bigvee_{A_G\in PA(G,\varphi)} \bigwedge_{A_{\overline{G}}\in PA(\overline{G},\varphi)} \Big(\bigwedge_{\substack{g\in G\\ \varepsilon_g\in A_G(g)}} B_g\varepsilon_g \wedge \Big(\bigwedge_{\substack{g'\in\overline{G}\\ \varepsilon_{g'}\in A_{\overline{G}}(g')}} B_{g'}\varepsilon_{g'} \to [A_{Agt}]\varphi\Big)\Big)$$

where $\overline{G} = Agt \setminus G$ and $A_{Agt} = A_{\overline{G}} \cup A_G$ and $PA(G,\varphi) = \{A \in Ann_G \mid \forall b \in G, \forall\varepsilon \in A(b), \mathrm{depth}(\varepsilon) \leqslant m \text{ and } \mathrm{lit}(\varepsilon) \in \mathrm{Lit}(\varphi)\}$.

Axiomatization: the static language

$$\text{Propositional tautologies} \quad \text{(A0)}$$
$$B_a\top \quad \text{(A1)}$$
$$\varepsilon \to \neg\overline{\varepsilon} \quad \text{(A2)}$$
$$T_{a,b} \to (B_aB_b\varepsilon \to B_a\varepsilon) \quad \text{(A3)}$$
$$T_{a,a} \quad \text{(A4)}$$
$$\vdash \varphi, \varphi \to \psi \Rightarrow \vdash \psi \quad \text{(I0)}$$

Theorem

The axiom system $(A0) - (A4) + (I0)$ is sound and complete with regard to the static logic $SDDLT^*$.

Axioms: the full language

First reduction axioms, where A is an announcement by group G:

$$[A]\neg\varphi \leftrightarrow (B_G A \rightarrow \neg[A]\varphi) \tag{A5}$$

$$[A](\varphi \wedge \psi) \leftrightarrow [A]\varphi \wedge [A]\psi \tag{A6}$$

$$[A]p \leftrightarrow (B_G A \rightarrow p) \tag{A7}$$

$$[A]T_{a,b} \leftrightarrow T_{a,b} \text{ for } b \notin G \tag{A8}$$

Reduction axiom for trust

Recall:

$$T_a^{s\cdot[A]} = T_a^s \setminus \{b \in G \mid \exists c \in G, \varepsilon_b \in A(b), \varepsilon_c \in A(c), \\ \min_a^s(B_b\varepsilon_b) = \overline{\min_a^s(B_c\varepsilon_c)} \\ \text{and } \min_a^s(B_b\varepsilon_b) \notin BB_a^s\}$$

We just need to know how to express $\min_a^s(\varepsilon_b) = \overline{\min_a^s(\varepsilon_c)}$ and $\min_a^s(\varepsilon_b) \notin BB_a^s$.

Lemma

Let ε_1 and ε_2 be formulas of $\mathcal{L}_B$, a be an agent, and s a state. Define ε_0 as $\varepsilon_0 = \max(\varepsilon_1, \overline{\varepsilon_2})$. We have that $\min_a^s(\varepsilon_1) = \overline{\min_a^s(\varepsilon_2)}$ iff $\mathrm{lit}(\varepsilon_1) = \mathrm{lit}(\overline{\varepsilon_2})$, $\varepsilon_1 \rightarrow_a^s \varepsilon_0$ and $\varepsilon_2 \rightarrow_a^s \overline{\varepsilon_0}$.

Contradiction of ε_1 and ε_2 according to a is expressed by:

$$CO(a, \varepsilon_1, \varepsilon_2) = \bot$$

if $lit(\varepsilon_1) \neq \overline{lit(\varepsilon_2)}$, and

$$CO(a, \varepsilon_1, \varepsilon_2) = \Big(\bigwedge_{\alpha \in Agt(\varepsilon_1 \setminus \max(\varepsilon_1, \overline{\varepsilon_2}))} T_{a,\alpha} \Big) \wedge \Big(\bigwedge_{\beta \in Agt(\varepsilon_2 \setminus \max(\varepsilon_2, \overline{\varepsilon_1}))} T_{a,\beta} \Big)$$

otherwise.

Lemma

If ε is a formula of $\mathcal{L}_B$, a is an agent, and s is a state, we have that $\min_a^s(\varepsilon) \notin BB_a^s$ iff for all $\mu \preceq \varepsilon$, $\varepsilon \rightarrow_a^s \mu$ implies $\mu \notin BB_a^s$.

The condition about previous beliefs is expressed by:

$$PB(a, \varepsilon) = \bigwedge_{\mu \preceq \varepsilon}((\bigwedge_{\alpha \in Agt(\varepsilon \setminus \mu)} T_{a,\alpha}) \rightarrow \neg B_a \mu)$$

Reduction axiom for trust, when $b \in G$:

$$[A] T_{a,b} \leftrightarrow T_{a,b} \wedge \neg \bigvee_{\varepsilon_b \in A(b)} \bigvee_{\substack{c \in G \\ \varepsilon_c \in A(c)}} (CO(a, B_b\varepsilon_b, B_c\varepsilon_c) \wedge PB(a, \varepsilon_b)) \quad \text{(A9)}$$

Reduction axiom for beliefs

We can construct a reduction axiom for beliefs in a similar way:

$$[A]B_a\varepsilon \leftrightarrow Ann(A, \varepsilon, a) \vee (B_a\varepsilon \wedge NC(A, \varepsilon, a)) \text{ (A10)}$$

where

$$Ann(A, \varepsilon, a) = \bigvee_{\substack{b \in G \\ \varepsilon_b \in A(b): \varepsilon \preceq B_b\varepsilon_b}} \bigwedge_{\alpha \in Agt(B_b\varepsilon_b \setminus \varepsilon)} [A]T_{a,\alpha}$$

and

$$NC(A, \varepsilon, a) = \bigwedge\nolimits_{\mu \preceq \overline{\varepsilon}}((\bigwedge\nolimits_{\alpha \in Agt(\overline{\varepsilon} \setminus \mu)}[A]T_{a,\alpha}) \rightarrow \neg Ann(A, \mu, a))$$

Discussion: on the choice of dynamics

We have made choices in order to attempt to find a balance between non-trivial trust dynamics and clarity of the formalization. However, more complex dynamics could be envisioned, so long as we know how to express relevant properties as formulas.

Example (So agents don't believe false announcements about their own beliefs)

After announcement A by a group G:

$$T_a^{s\cdot[A]} = T_a^s \setminus (C(a,s,A) \cup \{b \in G \mid \exists c \in G \setminus C(a,s,A), \varepsilon_b \in A(b), \varepsilon_c \in A(c), \min_a^s(B_b\varepsilon_b) = \overline{\min_a^s(B_c\varepsilon_c)} \text{ and } \min_a^s(B_b\varepsilon_b) \notin BB_a^s\})$$

where $C(a,s,A) = \{b \in G \mid \exists \varepsilon \notin BB_a^s, B_a\varepsilon \in A(b)\}$

Conclusion

- Logic SDDLT: framework for reasoning about the evolution of trust and beliefs as triggered by joint announcements
- Sound and complete axiomatization using reduction axioms
- Operators for coalition announcements are definable in SDDLT when the depth of nesting of belief operators is bounded.

Conclusion

- Logic SDDLT: framework for reasoning about the evolution of trust and beliefs as triggered by joint announcements
- Sound and complete axiomatization using reduction axioms
- Operators for coalition announcements are definable in SDDLT when the depth of nesting of belief operators is bounded.
- Possible expansions:
  - Announcements: private announcements, dishonest agents
  - Trust: degrees of trust, gaining trust (memory?)
  - Beliefs: more complex formulas $\rightarrow$ paraconsistent logic / belief merging to deal with contradictory announcements