Logics of individual and collective intentionality
(Lecture IV: Group belief)

Andreas Herzig, **Emiliano Lorini**

ESSLLI’09
Bordeaux, 20-25 July 2009

Introduction

The logic of distributed belief

The logic of common belief

The logic of collective acceptance

Introduction

Beliefs of agents and beliefs of groups

- Bill believes that φ (Bob is a honest person, Earth must be protected, etc.)
- Bill and Mary believe that φ.
  - Both Mary and Bill believe that φ?
  - Mary and Bill, by pooling their beliefs together, can infer φ?
  - Bill believes that φ, Mary believes that φ, Bill believes that Mary believes that φ, Mary believes that Bill believes that φ, and so on.
  - Mary and Bill, *qua* members of the same group, believe/accept that φ?

Reductionist vs. non-reductionistic approaches

Suppose J is a set of agents

- **Reductionistic approaches**: group belief can be reduced to an aggregate of individual beliefs.
  - E.g. summative view (Quinton, 1975): A group J believes φ $\approx$ all or most of the agents in J believe φ.
  - but also common belief, distributed belief.
- **Non-reductionistic approaches** (Gilbert, 1989; Tuomela, 1992; Tuomela, 2002): the concept of *constituted group* + irreducibility of group belief.
  - A group J believes $\varphi \approx$ the agents in J are functioning as members of the same group and they accept φ to stand as the view of the group.
  - Group J may believe φ even though nobody in J individually believes φ.

Reductionist vs. non-reductionistic approaches (cont.)

Engel (1998) prefers the term **collective acceptance** in order to refer to a non-reductionistic notion of group belief

The logic of distributed belief

The concept of distributed belief

$\Rightarrow$ A set of agents I has a distributed belief that φ if and only if by pooling their beliefs together the agents in I can deduce φ, even though it may be the case that anybody in I believes φ

Example (Two detectives)

- 1 believes that the murderer is a tall person and is uncertain whether the murderer is a young person,
- 2 believes that the murderer is young and is uncertain whether the murderer is tall.

Therefore, 1 and 2 have a distributed belief that the murderer is tall and young even though nobody believes this.

Language

- AGT: a set of agents (or individuals);
- ATM: a set of atomic formulas;
- $2^{AGT*} = 2^{AGT} \setminus \emptyset$.

Language:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathbf{B}_i\varphi \mid \mathbf{DB}_I\varphi$$

p ranges over ATM, i ranges over AGT, I ranges over 2^{AGT*}

$\Rightarrow$ $\mathbf{B}_i\varphi$: agent i believes that φ.
$\Rightarrow$ $\mathbf{DB}_I\varphi$: the agents in I have a distributed belief that φ.

Models

Doxastic models are tuples $\langle W, \mathcal{B}, \mathcal{V} \rangle$ where:

- W is a non-empty set of worlds (or states);
- $\mathcal{B}$ yields a serial, transitive and Euclidian accessibility relation $\mathcal{B}_i \subseteq W \times W$ for every $i \in AGT$.
  - Serial: for every $w \in W$ there exists v such that $(w, v) \in \mathcal{B}_i$.
  - Transitive: if $(w, v) \in \mathcal{B}_i$ and $(v, u) \in \mathcal{B}_i$ then $(w, u) \in \mathcal{B}_i$.
  - Euclidian: if $(w, v) \in \mathcal{B}_i$ and $(w, u) \in \mathcal{B}_i$ then $(v, u) \in \mathcal{B}_i$.
- $\mathcal{V} : ATM \to 2^W$.

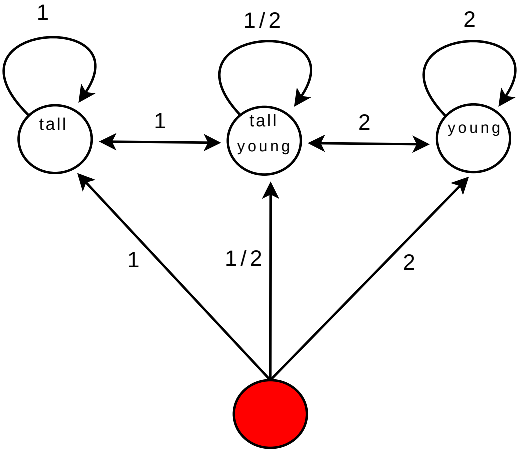
For every $w \in W$, $\mathcal{B}_i(w) = \{v | (w, v) \in \mathcal{B}_i\}$ is the set of worlds that are *possible* for agent i at w (i's information state)

Truth conditions

- $\mathcal{M}, w \models p$ iff $w \in \mathcal{V}(p)$
- $\mathcal{M}, w \models \neg\varphi$ iff not $\mathcal{M}, w \models \varphi$
- $\mathcal{M}, w \models \varphi \vee \psi$ iff $\mathcal{M}, w \models \varphi$ or $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models \mathbf{B}_i\varphi$ iff $\mathcal{M}, v \models \varphi$ for all $(w, v) \in \mathcal{B}_i$
- $\mathcal{M}, w \models \mathbf{DB}_I\varphi$ iff $\mathcal{M}, v \models \varphi$ for all $(w, v) \in \mathcal{D}_I$

where
$\Rightarrow \mathcal{D}_I = \{(w, v) | (w, v) \in \bigcap_{i \in I} \mathcal{B}_i\}$

Example: detectives



$\mathbf{B}_1 \mathit{tall} \wedge \mathbf{B}_2 \mathit{young} \wedge \neg \mathbf{B}_1 \mathit{young} \wedge \neg \mathbf{B}_2 \mathit{tall} \wedge \mathbf{DB}_{\{1,2\}}(\mathit{tall} \wedge \mathit{young})$
is true at the red world

A complete axiomatization of distributed belief logic

(ProTau) All tautologies of propositional calculus

($\mathsf{K}_{\mathbf{B}_i}$) $(\mathbf{B}_i\varphi \wedge \mathbf{B}_i(\varphi \rightarrow \psi)) \rightarrow \mathbf{B}_i\psi$

($\mathsf{D}_{\mathbf{B}_i}$) $\neg(\mathbf{B}_i\varphi \wedge \mathbf{B}_i\neg\varphi)$

($4_{\mathbf{B}_i}$) $\mathbf{B}_i\varphi \rightarrow \mathbf{B}_i\mathbf{B}_i\varphi$

($5_{\mathbf{B}_i}$) $\neg\mathbf{B}_i\varphi \rightarrow \mathbf{B}_i\neg\mathbf{B}_i\varphi$

($\mathsf{K}_{\mathbf{DB}_I}$) $(\mathbf{DB}_I\varphi \wedge \mathbf{DB}_I(\varphi \rightarrow \psi)) \rightarrow \mathbf{DB}_I\psi$

($4_{\mathbf{DB}_I}$) $\mathbf{DB}_I\varphi \rightarrow \mathbf{DB}_I\mathbf{DB}_I\varphi$

($5_{\mathbf{DB}_I}$) $\neg\mathbf{DB}_I\varphi \rightarrow \mathbf{DB}_I\neg\mathbf{DB}_I\varphi$

($\mathsf{Int}_{\mathbf{B}_i,\mathbf{DB}_{\{i\}}}$) $\mathbf{B}_i\varphi \leftrightarrow \mathbf{DB}_{\{i\}}\varphi$

($\mathsf{Mon}_{\mathbf{DB}_I}$) $\mathbf{DB}_I\varphi \rightarrow \mathbf{DB}_J\varphi$ if $I \subseteq J$

(MP) If φ and $\varphi \rightarrow \psi$ then ψ

($\mathsf{Nec}_{\mathbf{B}_i}$) If φ then $\mathbf{B}_i\varphi$

($\mathsf{Nec}_{\mathbf{DB}_I}$) If φ then $\mathbf{DB}_I\varphi$

A remark

We might have $\bigcap_{i \in I} \mathcal{B}_i(w) = \emptyset$ for $|I| > 1$

$$\mathbf{DB}_I \bot \text{ is consistent for } |I| > 1$$

where $\mathbf{DB}_I \bot$ means that the agent in I do not have a distributed belief

The logic of common belief

The concept of common belief

Common belief that φ in a set of agents $I \approx$
the agents in I mutually believe that φ for every order $k \geq 0$.

every agent in I believes φ, every agent in I believes that every agent in I believes φ, and so on ad infinitum.

- Aumann (1976, 1999) gives the first mathematical characterization of a similar concept using set theory: common knowledge (common K is always truthful).
- Theories of common B/common K using doxastic/epistemic logic can be found in Bacharach (1992), Bicchieri (1989), Fagin et al. (1995).

The concept of common belief (cont.)

Common belief is a fundamental concept for the explanation of team activity, coordination, communication.

- Joint/team activity involves common belief (Grosz & Kraus, 1996).
- The concept of convention, as a solution to coordination problems, is classically defined in terms of common belief (Lewis, 1969).
- Common belief justifies the plausibility of Equilibrium notions in game theory like Nash Equilibrium, Iterated Strict Dominance, Rationalizability (Battigalli & Bonanno, 1999).
- Common belief has been used to define the concept of common ground in a conversation (Stalnaker, 2001) as a fundamental basis for discourse understanding and definite reference (Clark & Marshall, 1981; Schiffer, 1972).

Language

Language:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathbf{B}_i\varphi \mid \mathbf{CB}_I\varphi$$

p ranges over ATM, i ranges over AGT, I ranges over 2^{AGT*}

$\Rightarrow$ $\mathbf{CB}_I\varphi$: there is a common belief that φ in I.

Further concepts

- Agent i has a doubt about φ.
- $\mathbf{Doubt}_i\varphi \overset{\text{def}}{=} \neg\mathbf{B}_i\varphi \wedge \neg\mathbf{B}_i\neg\varphi$
- Everybody in I believes φ.
- $\mathbf{EB}_I\varphi \overset{\text{def}}{=} \bigwedge_{i\in I}\mathbf{B}_i\varphi$

Models

Doxastic models are tuples $\langle W, \mathcal{B}, \mathcal{V} \rangle$ where:

- W is a non-empty set of worlds (or states);
- $\mathcal{B}$ yields a serial, transitive and Euclidian accessibility relation $\mathcal{B}_i \subseteq W \times W$ for every $i \in AGT$.
- $\mathcal{V} : ATM \to 2^W$.

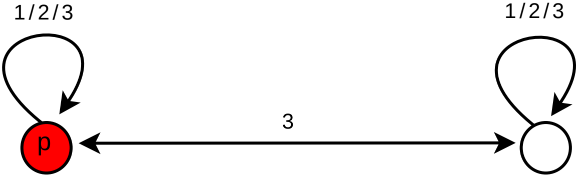
For every $w \in W$, $\mathcal{B}_i(w) = \{v | (w, v) \in \mathcal{B}_i\}$ is the set of worlds that are *possible* for agent i at w (i's information state)

Truth conditions

- $\mathcal{M}, w \models \mathbf{CB}_I\varphi$ iff $\mathcal{M}, v \models \varphi$ for all $(w, v) \in \mathcal{B}_I^+$

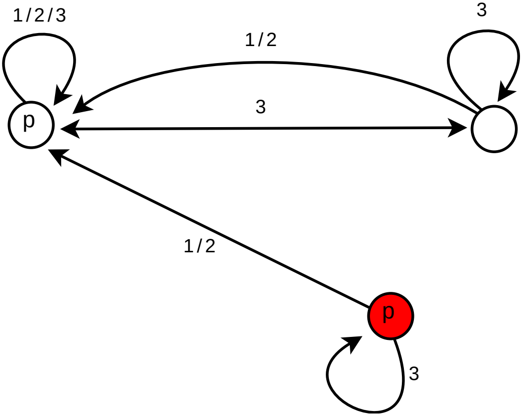
where
$\Rightarrow \mathcal{B}_I = \bigcup_{i \in I} \mathcal{B}_i$
$\Rightarrow \mathcal{B}_I^+$ is the transitive closure of $\mathcal{B}_I$

Example 1: a real secret



$\mathbf{EB}_{\{1,2\}}p \wedge \mathbf{Doubt}_3 p \wedge \mathbf{CB}_{\{1,2\}}(p \wedge \mathbf{Doubt}_3 p)$
is true at the red world

Example 2: an apparent secret



$\mathbf{EB}_{\{1,2,3\}}p \wedge \mathbf{CB}_{\{1,2\}}(p \wedge \mathbf{Doubt}_3 p) \wedge \mathbf{B}_3\mathbf{CB}_{\{1,2\}}(p \wedge \mathbf{Doubt}_3 p)$
is true at the red world

A complete axiomatization of common belief logic

(ProTau)	All tautologies of propositional calculus
($\mathrm{K}_{\mathbf{B}_i}$)	$(\mathbf{B}_i\varphi \wedge \mathbf{B}_i(\varphi \to \psi)) \to \mathbf{B}_i\psi$
($\mathrm{D}_{\mathbf{B}_i}$)	$\neg(\mathbf{B}_i\varphi \wedge \mathbf{B}_i\neg\varphi)$
($4_{\mathbf{B}_i}$)	$\mathbf{B}_i\varphi \to \mathbf{B}_i\mathbf{B}_i\varphi$
($5_{\mathbf{B}_i}$)	$\neg\mathbf{B}_i\varphi \to \mathbf{B}_i\neg\mathbf{B}_i\varphi$
($\mathrm{K}_{\mathbf{CB}_I}$)	$(\mathbf{CB}_I\varphi \wedge \mathbf{CB}_I(\varphi \to \psi)) \to \mathbf{CB}_I\psi$
(FixPoint)	$\mathbf{CB}_I\varphi \to \mathbf{EB}_I(\varphi \wedge \mathbf{CB}_I\varphi)$
(MP)	If φ and $\varphi \to \psi$ then ψ
($\mathrm{Nec}_{\mathbf{B}_i}$)	If φ then $\mathbf{B}_i\varphi$
($\mathrm{Nec}_{\mathbf{CB}_I}$)	If φ then $\mathbf{CB}_I\varphi$
(Induction)	If $\varphi \to \mathbf{EB}_I(\varphi \wedge \psi)$ then $\varphi \to \mathbf{CB}_I\psi$

Some theorems

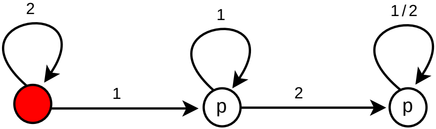
- $\vdash \mathbf{CB}_I\varphi \rightarrow \mathbf{EB}_I\varphi$
- $\vdash \neg(\mathbf{CB}_I\varphi \wedge \mathbf{CB}_I\neg\varphi)$
- proof:
  1. $\vdash (\mathbf{CB}_I\varphi \rightarrow \mathbf{EB}_I\varphi) \wedge (\mathbf{CB}_I\neg\varphi \rightarrow \mathbf{EB}_I\neg\varphi)$ by Fixpoint Axiom
  2. $\vdash (\mathbf{CB}_I \wedge \mathbf{CB}_I\neg\varphi) \rightarrow \mathbf{EB}_I\bot$ from 1 by prop. calculus
  3. $\vdash \mathbf{EB}_I\bot \rightarrow \bot$ by Axiom $\mathrm{D}_{\mathbf{B}_i}$
  4. $\vdash (\mathbf{CB}_I \wedge \mathbf{CB}_I\neg\varphi) \rightarrow \bot$ from 2,3
- $\vdash \mathbf{CB}_I\varphi \rightarrow \mathbf{CB}_J\varphi$ if $J \subseteq I$
- proof:
  1. $\vdash \mathbf{EB}_I\varphi \rightarrow \mathbf{EB}_J\varphi$ by prop. calculus
  2. $\vdash \mathbf{CB}_I\varphi \rightarrow \mathbf{EB}_J(\varphi \wedge \mathbf{CB}_I\varphi)$ from 1 by Fixpoint Axiom
  3. $\vdash \mathbf{CB}_I\varphi \rightarrow \mathbf{CB}_J\varphi$ from 2 by Induction Rule

Some theorems (cont.)

- $\vdash \mathbf{CB}_I\varphi \to \mathbf{CB}_I\mathbf{CB}_I\varphi$
- proof:
  1. $\vdash \mathbf{CB}_I\varphi \to \mathbf{EB}_I\mathbf{CB}_I\varphi$ by Fixpoint Axiom
  2. $\vdash \mathbf{CB}_I\varphi \to \mathbf{EB}_I(\mathbf{CB}_I\varphi \land \mathbf{CB}_I\varphi)$ from 1
  3. $\vdash \mathbf{CB}_I\varphi \to \mathbf{CB}_I\mathbf{CB}_I\varphi$ from 2 by Induction Rule
- $\vdash \mathbf{CB}_I\varphi \to \bigwedge_{1\leq k\leq n} \mathbf{EB}_I^k\varphi$
- $\vdash \mathbf{EB}_I\mathbf{CB}_I\varphi \to \mathbf{CB}_I\varphi$
- proof:
  1. $\vdash \mathbf{EB}_I\mathbf{CB}_I\varphi \to \mathbf{EB}_I\mathbf{EB}_I(\varphi \land \mathbf{CB}_I\varphi)$
     by Fixpoint Axiom, Axiom K and Necessitation for $\mathbf{B}_i$
  2. $\vdash \mathbf{EB}_I\mathbf{EB}_I(\varphi \land \mathbf{CB}_I\varphi) \to \mathbf{EB}_I(\varphi \land \mathbf{EB}_I\mathbf{CB}_I\varphi)$
     by Theorem $\mathbf{EB}_I\mathbf{EB}_I\varphi \to \mathbf{EB}_I\varphi$
  3. $\mathbf{EB}_I\mathbf{CB}_I\varphi \to \mathbf{EB}_I(\varphi \land \mathbf{EB}_I\mathbf{CB}_I\varphi)$ from 1,2
  4. $\mathbf{EB}_I\mathbf{CB}_I\varphi \to \mathbf{CB}_I\varphi$ from 3 by Induction Rule

Some invalid properties

- $\not\models \neg\mathbf{CB}_I\varphi \rightarrow \mathbf{CB}_I\neg\mathbf{CB}_I\varphi$
- $\not\models \mathbf{B}_i\mathbf{CB}_I\varphi \rightarrow \mathbf{CB}_I\varphi$



$\neg\mathbf{CB}_{\{1,2\}}p \wedge \neg\mathbf{CB}_{\{1,2\}}\neg\mathbf{CB}_{\{1,2\}}p \wedge \mathbf{B}_1\mathbf{CB}_{\{1,2\}}p$
is true at the red world

Common belief in coordination problems

Example

The agents in AGT have to move an attack against an enemy. The attack will be successful iff it is a coordinated attack (everybody attacks the enemy). We assume that:

1. an agent attacks iff he believes that the others also attack,
2. the agents have a common belief about this.

$$\begin{array}{ll} (\mathit{Hyp1}) & \bigwedge_{i \in AGT} (attack_i \leftrightarrow \mathbf{B}_i \bigwedge_{j \in AGT} attack_j) \\ (\mathit{Hyp2}) & \mathbf{CB}_{AGT} \mathit{Hyp1} \end{array}$$

$\mathit{Hyp1}$ and $\mathit{Hyp2}$ imply that everybody attacks iff there is common belief that everybody attacks:

$$\vdash (\mathit{Hyp1} \wedge \mathit{Hyp2}) \rightarrow (\bigwedge_{i \in AGT} attack_i \leftrightarrow (\mathbf{CB}_{AGT} \bigwedge_{i \in AGT} attack_i))$$

Proof: left-to-right direction

We note $\chi = \bigwedge_{i \in AGT} attack_i$

1. $(\mathit{Hyp1} \wedge \mathit{Hyp2} \wedge \chi) \rightarrow$
   $(\mathbf{EB}_{AGT}\chi \wedge \mathbf{CB}_{AGT}(\chi \leftrightarrow \mathbf{EB}_{AGT}\chi))$
2. $(\mathbf{EB}_{AGT}\chi \wedge \mathbf{CB}_{AGT}(\chi \leftrightarrow \mathbf{EB}_{AGT}\chi)) \rightarrow$
   $(\mathbf{EB}_{AGT}\chi \wedge \mathbf{EB}_{AGT}((\chi \leftrightarrow \mathbf{EB}_{AGT}\chi) \wedge \mathbf{CB}_{AGT}(\chi \leftrightarrow \mathbf{EB}_{AGT}\chi)))$
   by Fixpoint Axiom
3. $(\mathbf{EB}_{AGT}\chi \wedge \mathbf{EB}_{AGT}((\chi \leftrightarrow \mathbf{EB}_{AGT}\chi) \wedge \mathbf{CB}_{AGT}(\chi \leftrightarrow \mathbf{EB}_{AGT}\chi))) \rightarrow$
   $(\mathbf{EB}_{AGT}\chi \wedge \mathbf{EB}_{AGT}\mathbf{EB}_{AGT}\chi \wedge \mathbf{EB}_{AGT}\mathbf{CB}_{AGT}(\chi \leftrightarrow \mathbf{EB}_{AGT}\chi))$
   by Theorem $\mathbf{EB}_{AGT}\chi \rightarrow \mathbf{EB}_{AGT}\mathbf{EB}_{AGT}\chi$
4. $(\mathbf{EB}_{AGT}\chi \wedge \mathbf{EB}_{AGT}\mathbf{EB}_{AGT}\chi \wedge \mathbf{EB}_{AGT}\mathbf{CB}_{AGT}(\chi \leftrightarrow \mathbf{EB}_{AGT}\chi)) \rightarrow$
   $\mathbf{EB}_{AGT}(\chi \wedge \mathbf{EB}_{AGT}\chi \wedge \mathbf{CB}_{AGT}(\chi \leftrightarrow \mathbf{EB}_{AGT}\chi))$
5. $(\mathbf{EB}_{AGT}\chi \wedge \mathbf{CB}_{AGT}(\chi \leftrightarrow \mathbf{EB}_{AGT}\chi)) \rightarrow$
   $\mathbf{EB}_{AGT}(\chi \wedge \mathbf{EB}_{AGT}\chi \wedge \mathbf{CB}_{AGT}(\chi \leftrightarrow \mathbf{EB}_{AGT}\chi))$
   from 2,3,4
6. $(\mathbf{EB}_{AGT}\chi \wedge \mathbf{CB}_{AGT}(\chi \leftrightarrow \mathbf{EB}_{AGT}\chi)) \rightarrow \mathbf{CB}_{AGT}\chi$
   from 5 by Induction Rule
7. $(\mathit{Hyp1} \wedge \mathit{Hyp2}) \rightarrow (\chi \rightarrow \mathbf{CB}_{AGT}\chi)$
   from 1,6

The dynamics of common belief

- Common belief can be created by means of public announcements.
    - A certain fact φ is perceived/observed by every agent.
    - All agents have a common belief that φ has been perceived/observed by every agent.

$\Rightarrow$ $\varphi!$: public announcement of φ.

$\Rightarrow$ $[\varphi!]\psi$: ψ holds after the public announcement of φ.

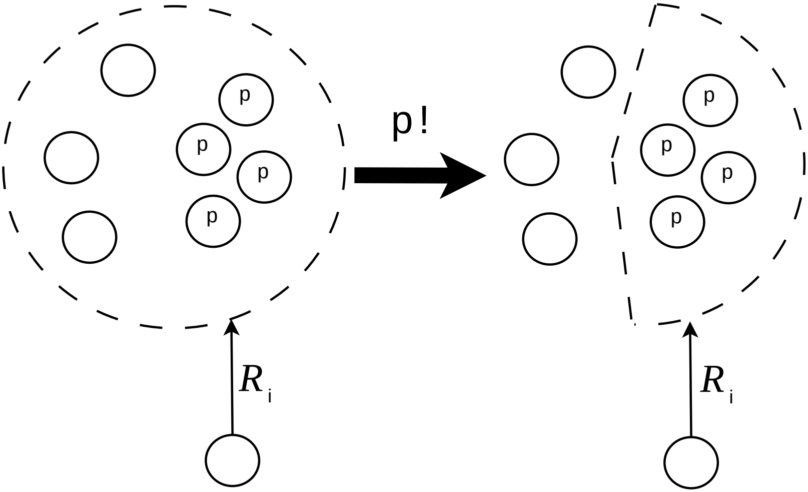
The dynamics of common belief (cont.)

$$\mathcal{M}, w \models [\varphi!]\psi \text{ iff } \mathcal{M}^{\varphi!}, w \models \psi$$

The updated model $\mathcal{M}^{\varphi!}$ is the tuple $\langle W^{\varphi!}, \mathcal{B}^{\varphi!}, \mathcal{V}^{\varphi!} \rangle$ where:

- $W^{\varphi!} = W$;
- For every $i \in AGT$, $\mathcal{B}_i^{\varphi!} = \{(w, v) \in \mathcal{B}_i | M, v \models \varphi\}$;
- For every $p \in ATM$, $\mathcal{V}^{\varphi!}(p) = \mathcal{V}(p)$.

The dynamics of common belief (cont.)

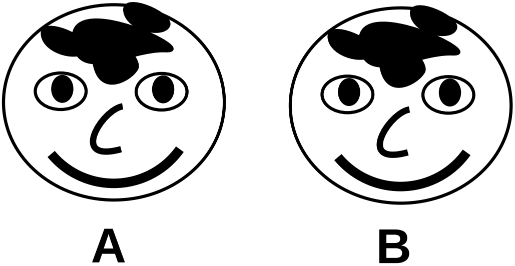


A remark

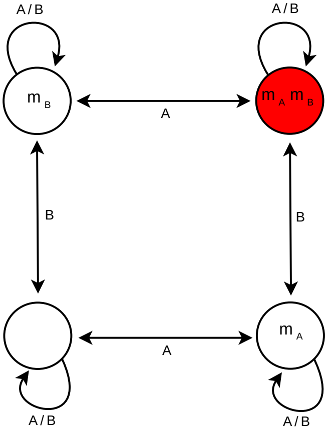
Axiom D for $\mathbf{B}_i$ and the corresponding propriety of seriality for every $\mathcal{B}_i$ must be removed when adding announcements. Indeed:

$$\models \mathbf{B}_i \neg p \rightarrow [p!]\mathbf{B}_i \bot$$

The muddy children problem



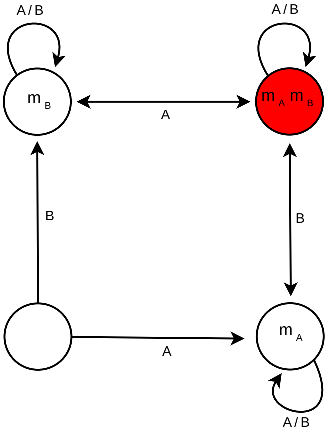
The muddy children problem (cont.)



Father says: “at least one of you is muddy!”

$\Rightarrow m_A \vee m_B$!

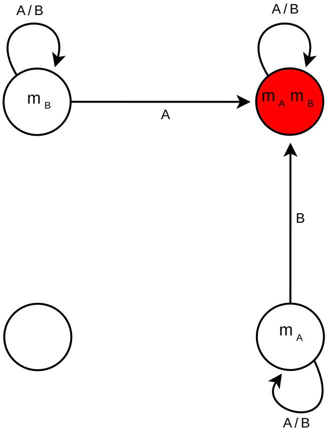
The muddy children problem (cont.)



Father asks: “are you muddy?”
A says: “I don’t know!” and B says: “I don’t know!”

$\Rightarrow \mathbf{Doubt}_A m_A \wedge \mathbf{Doubt}_B m_B!$

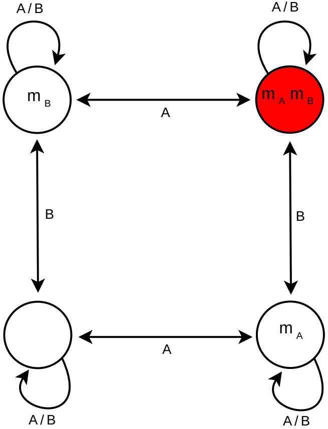
The muddy children problem (cont.)



1 and 2 have reached a common belief that $m_A \wedge m_B$

$\Rightarrow \mathbf{CB}_{\{1,2\}}(m_A \wedge m_B)$ is true at the red world

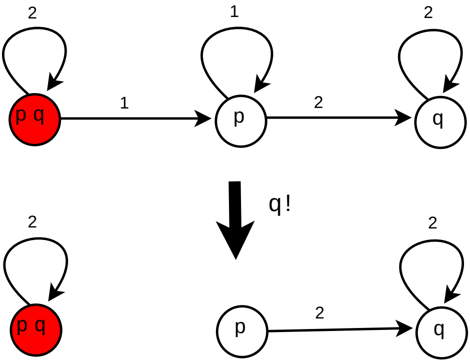
The muddy children problem (cont.)



$[m_A \vee m_B!][\mathbf{Doubt}_A m_A \wedge \mathbf{Doubt}_B m_B!]\ \mathbf{CB}_{\{1,2\}}(m_A \wedge m_B)$
is true at the red world

The “unforeseen” appearance of common belief

$[q!]\mathbf{CB}_{\{1,2\}}p \wedge \neg\mathbf{CB}_{\{1,2\}}(q \rightarrow p)$ is true at the red world



Concluding remarks

Is common belief a good candidate for a concept of proper group belief?

- No notion of ‘group’ *stricto sensu* involved in the notion of common belief: ‘group of agents’ $\neq$ ‘set of agents’.
  - Agents i and j might have a common belief that ‘$2+2=4$’ even if i and j do not know each other and are not members of the same group.
- Individual belief and group belief should be independent.
  - At the end of the 80s, the Communist Party of Ruritania believed/accepted that capitalist countries will soon perish, but none of its members really believed so (Tuomela, 1992).
  - Common belief in I implies individual belief for every agent in I ($\vdash \mathbf{CB}_I\varphi \rightarrow \mathbf{B}_i\varphi$, if $i \in I$).

...towards a notion of *collective acceptance*

The logic of collective acceptance

Acceptance *qua* members of a group

Acceptance logic (Lorini & Longin, 2008; Lorini et al., 2009; Herzig, de Lima, Lorini, 2009) allows to reason about the following two aspects:

- Certain agents identify themselves as members of the same group, or organization, or team, or institution, etc. and recognize mutually as members of the same group, or organization, or team, or institution, etc. (Gilbert, 1989) and,
- they accept certain things *qua* members of the same group, or organization, or team, or institution, etc. (Tuomela, 2007).

Individual acceptance vs. collective acceptance

- Individual acceptance: a certain agent i accepts that something is true, *qua* member of a certain group (or organization, or team, or institution, etc.)
- Collective acceptance: the agents in C accept that something is true, *qua* member of the same group (or organization, or team, or institution, etc.)

Example

Agent i, *qua* lawyer, accepts that his client is innocent.

Example

Three agents i, j and z accept that their mission is to protect the Earth *qua* members of Greenpeace.

Acceptance Logic (AL)

- AGT: a finite set of agents;
- ATM: a countable set of atomic formulas;
- X: a finite set of social contexts (group, organization, team, institution, etc.).

We note $2^{AGT*} = 2^{AGT} \setminus \emptyset$

Language:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathbf{A}_{I:x}\varphi$$

where I ranges over 2^{AGT*} and x ranges over X

$$\widehat{\mathbf{A}}_{I:x}\varphi =_{def} \neg\mathbf{A}_{I:x}\neg\varphi$$

Acceptance Logic (AL)

$\mathbf{A}_{I:x}\varphi$
'if the agents in I function together as members of x then they accept φ' (or 'the agents in I accept that φ while functioning together as members of x').

$\widehat{\mathbf{A}}_{I:x}\top$
'the agents in I are functioning together as members of x'.

$\widehat{\mathbf{A}}_{I:x}\top \wedge \mathbf{A}_{I:x}\varphi$
'the agents in I accept that φ *qua* members of x'.

AL Models

Acceptance models are tuples $\langle W, \mathcal{A}, \mathcal{V} \rangle$ where:

- W is a non-empty set of worlds (or states);
- $\mathcal{A}$ yields an accessibility relation $\mathcal{A}_{I,x} \subseteq W \times W$ for every $I \in 2^{AGT*}$ and $x \in X$.
- $\mathcal{V} : ATM \rightarrow 2^{W}$.

$\mathcal{A}_{I,x}(w) = \{v | (w,v) \in \mathcal{A}_{I,x}\}$: the worlds *accepted* by the agents in I while functioning as members of group x at world w (I's acceptance state in the context x)

Remark

$\neq$ common belief, the accessibility relations for collective acceptances are not computed from the accessibility relations for individuals.

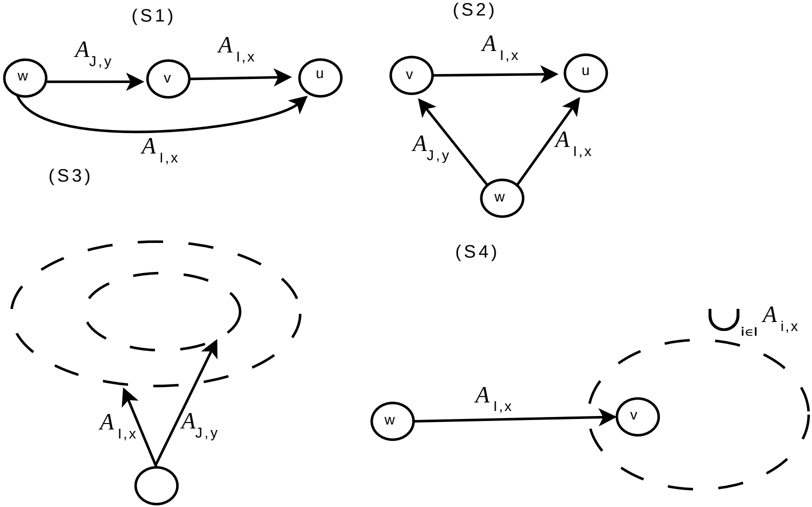
Constraints on AL models

For every $x, y \in X$ and $I, J \in 2^{AGT*}$ such that $J \subseteq I$:

- (S.1) If $(w, v) \in \mathcal{A}_{J,y}$ and $(v, u) \in \mathcal{A}_{I,x}$ then $(w, u) \in \mathcal{A}_{I,x}$.
- (S.2) If $(w, v) \in \mathcal{A}_{J,y}$ and $(w, u) \in \mathcal{A}_{I,x}$ then $(v, u) \in \mathcal{A}_{I,x}$.
- (S.3) If $\mathcal{A}_{I,x}(w) \neq \emptyset$ then $\mathcal{A}_{J,x} \subseteq \mathcal{A}_{I,x}(w)$.
- (S.4) If $v \in \mathcal{A}_{I,x}(w)$ then $v \in \bigcup_{i \in I} \mathcal{A}_{i,x}(v)$.

Constraints on AL models (cont.)

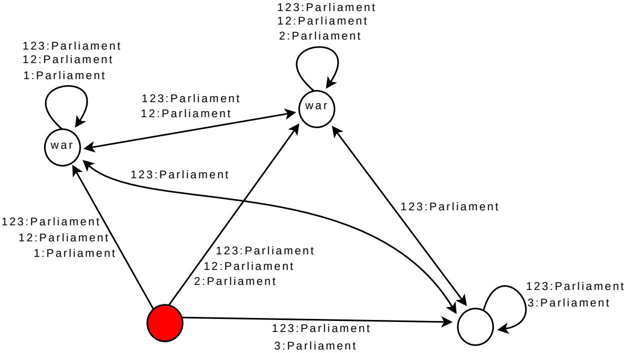
For $J \subseteq I$



Truth conditions

- $\mathcal{M}, w \models p$ iff $w \in \mathcal{V}(p)$
- $\mathcal{M}, w \models \neg\varphi$ iff not $\mathcal{M}, w \models \varphi$
- $\mathcal{M}, w \models \varphi \vee \psi$ iff $\mathcal{M}, w \models \varphi$ or $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models \mathbf{A}_{I:x}\varphi$ iff $\mathcal{M}, v \models \varphi$ for all $(w, v) \in \mathcal{A}_{I,x}$

Example: agreements and disagreements



$$
\begin{aligned}
&\widehat{\mathbf{A}}_{1:Parliament}\top \wedge \mathbf{A}_{1:Parliament}\,war\wedge\\
&\widehat{\mathbf{A}}_{2:Parliament}\top \wedge \mathbf{A}_{2:Parliament}\,war\wedge\\
&\widehat{\mathbf{A}}_{3:Parliament}\top \wedge \mathbf{A}_{3:Parliament}\neg war\wedge\\
&\widehat{\mathbf{A}}_{12:Parliament}\top \wedge \mathbf{A}_{12:Parliament}\,war\wedge\\
&\neg\mathbf{A}_{123:Parliament}\,war \wedge \neg\mathbf{A}_{1,2,3:Parliament}\neg war
\end{aligned}
$$

is true at the red world

A complete axiomatization of AL

(ProTau) All tautologies of propositional calculus

($\mathsf{K}_{\mathbf{A}_{I:x}}$) $(\mathbf{A}_{I:x}\varphi \wedge \mathbf{A}_{I:x}(\varphi \rightarrow \psi)) \rightarrow \mathbf{A}_{I:x}\psi$

(PIntr) $\mathbf{A}_{I:x}\varphi \rightarrow \mathbf{A}_{J:y}\mathbf{A}_{I:x}\varphi$ if $J \subseteq I$

(NIntr) $\neg\mathbf{A}_{I:x}\varphi \rightarrow \mathbf{A}_{J:y}\neg\mathbf{A}_{I:x}\varphi$ if $J \subseteq I$

(Incl) $(\widehat{\mathbf{A}}_{I:x}\top \wedge \mathbf{A}_{I:x}\varphi) \rightarrow \mathbf{A}_{J:x}\varphi$ if $J \subseteq I$

(Unanim) $\mathbf{A}_{I:x}(\bigwedge_{i\in I}\mathbf{A}_{i:x}\varphi \rightarrow \varphi)$

(MP) If φ and $\varphi \rightarrow \psi$ then ψ

($\mathsf{Nec}_{\mathbf{A}_{I:x}}$) If φ then $\mathbf{A}_{I:x}\varphi$

Example 1

Agent 1 and agent 2, *qua* Clue players, accept that Mrs Red is the killer (noted r):

$$\widehat{\mathbf{A}}_{12:Clue}\top \wedge \mathbf{A}_{12:Clue}r$$

By axiom (Incl) we infer

$$\mathbf{A}_{1:Clue}r \wedge \mathbf{A}_{2:Clue}r$$

Example 2

Agents 1, 2 and 3 accept that the President of Republic is the supreme authority while functioning as French citizens:

$$\mathbf{A}_{123:France} PresAuth$$

Agents 1, 2 and 3 accept that the Pope is the supreme authority while functioning as Catholics:

$$\mathbf{A}_{123:Cath} PopeAuth$$

By Axiom (PIntr) we infer:

$$\mathbf{A}_{12:Cath}\mathbf{A}_{123:France} PresAuth \wedge \mathbf{A}_{12:France}\mathbf{A}_{123:Cath} PopeAuth$$

⇒ Every group accepts (the validity of) other groups' acceptances

Example 3: discursive dilemma (Pettit, 2001)

A three-member committee c has to judge whether a student can be admitted to a PhD program in Logic and Computation. According to the admission rule used for deciding: a student can be admitted (adm) iff he is good in mathematics ($math$) and in English writing (Eng). That is,

$$adm \leftrightarrow (math \wedge Eng).$$

The three-member committee uses a majority rule to decide on the issue.

Example 3: discursive dilemma (cont.)

	$math$	Eng	$adm \leftrightarrow (math \wedge Eng)$	adm
Judge 1	yes	yes	yes	yes
Judge 2	yes	no	yes	no
Judge 3	no	yes	yes	no
Majority	yes	yes	yes	no/yes ($\perp$)

Table: doctrinal paradox

Example 3: discursive dilemma (cont.)

A The three judges publicly agree on the admission rule: $\mathbf{A}_{123:c}(adm \leftrightarrow (math \wedge Eng)) \wedge \widehat{\mathbf{A}}_{123:c}\top$.

B Judge 1 says that he accepts $math \wedge Eng$: $\mathbf{A}_{123:c}\mathbf{A}_{1:c}(math \wedge Eng)$.

C Judge 2 says that he accepts $math \wedge \neg Eng$: $\mathbf{A}_{123:c}\mathbf{A}_{2:c}(math \wedge \neg Eng)$.

D Judge 3 says that he accepts $\neg math \wedge Eng$: $\mathbf{A}_{123:c}\mathbf{A}_{3:c}(\neg math \wedge Eng)$.

E Majority rule is used. For every J such that $J \subseteq \{123\}$ and $|J| \geq 2$ and for every φ such that $\varphi \in \{math, \neg math, Eng, \neg Eng, adm, \neg adm\}$:

$$\mathbf{A}_{123:c}(\bigwedge_{i \in J} \mathbf{A}_{i:c}\varphi \rightarrow \varphi)$$

We can prove that $(A \wedge B \wedge C \wedge D \wedge E) \rightarrow \bot$!

Some theorems

1. $\vdash \mathbf{A}_{I:x}\widehat{\mathbf{A}}_{I:x}\top$
2. $\vdash \mathbf{A}_{J:y}\mathbf{A}_{I:x}\varphi \leftrightarrow (\mathbf{A}_{J:y}\bot \vee \mathbf{A}_{I:x}\varphi)$ if $J \subseteq I$
3. $\vdash \mathbf{A}_{J:y}\neg\mathbf{A}_{I:x}\varphi \leftrightarrow (\mathbf{A}_{J:y}\bot \vee \neg\mathbf{A}_{I:x}\varphi)$ if $J \subseteq I$
4. $\vdash \mathbf{A}_{I:x}(\mathbf{A}_{I:x}\varphi \rightarrow \varphi)$
5. $\vdash (\bigwedge_{i\in I} \mathbf{A}_{I:x}\mathbf{A}_{i:x}\varphi) \rightarrow \mathbf{A}_{I:x}\varphi$

Some invalid properties

- $\not\models \widehat{\mathbf{A}}_{I:x}\top \rightarrow \widehat{\mathbf{A}}_{J:x}\top$ if $J \subseteq I$

$\Rightarrow$ Constituted groups are not closed under subsets

Example

Eleven players $\{1,\ldots,11\}$ constitute a football team while $\{1,\ldots,10\}$ do not constitute a football team.

Some invalid properties (cont.)

- $\not\models (\widehat{\mathbf{A}}_{I:x}\top \wedge \widehat{\mathbf{A}}_{J:x}\top) \rightarrow \widehat{\mathbf{A}}_{I\cup J:x}\top$

$\Rightarrow$ Constituted groups are not closed under set union

Example

$\{1, 2\}$ recognize mutually as owners of a property, $\{3, 4\}$ recognize mutually as owners of the same property, $\{1, 2, 3, 4\}$ do not recognize mutually as owners of the property.

Some remarks: beyond unanimity

Take two specific sets of agents I and J such that $J \subseteq I$ and $|I \setminus J| < |J|$ (i.e. J represents the majority of agents in I):

$$\text{(Majority)} \qquad \mathbf{A}_{I:x}((\bigwedge_{i \in J} \mathbf{A}_{i:x}\varphi) \rightarrow \varphi)$$

Suppose $Leader(x) \subseteq AGT$ is the set of leaders of group x:

$$\text{(Leader)} \qquad \mathbf{A}_{I:x}((\bigwedge_{i \in Leader(x)} \mathbf{A}_{i:x}\varphi) \rightarrow \varphi)$$

Some remarks: beyond unanimity (cont.)

Taking (Majority) as a logical axiom might be dangerous...

Theorem

Suppose (Majority) is valid for any I, J such that $J \subseteq I$ and $|I \setminus J| < |J|$ then, for $i \neq j$ we have:

$$(\mathbf{A}_{AGT:x}\mathbf{A}_{\{i,j\}:x}\varphi \wedge \bigwedge_{I \in 2^{AGT*}} \widehat{\mathbf{A}}_{I:x}\top) \rightarrow \mathbf{A}_{AGT:x}\varphi$$

Extending Acceptance Logic with beliefs

Language of $\mathsf{AL}+\mathsf{B}$ (Acceptance Logic with beliefs):

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathbf{A}_{I:x}\varphi \mid \mathbf{B}_i\varphi$$

where I ranges over 2^{AGT*}, x ranges over X and i ranges over AGT

AL + B Models

AL + B models are tuples $\langle W, \mathcal{A}, \mathcal{B}, \mathcal{V} \rangle$ where:

- $\langle W, \mathcal{A}, \mathcal{V} \rangle$ is a AL model;
- $\mathcal{B}$ yields a doxastic (serial, transitive and Euclidian) accessibility relation $\mathcal{B}_i \subseteq W \times W$ for every $i \in AGT$.

Interaction principles between acceptance and belief

- $\mathbf{A}_{I:x}\varphi \rightarrow \mathbf{B}_i\mathbf{A}_{I:x}\varphi$ if $i \in I$
- $\neg\mathbf{A}_{I:x}\varphi \rightarrow \mathbf{B}_i\neg\mathbf{A}_{I:x}\varphi$ if $i \in I$

The three principles correspond to the following constraints on AL $+$ B models:

(S.1) If $(w, v) \in \mathcal{B}_i$ and $(v, u) \in \mathcal{A}_{I,x}$ then $(w, u) \in \mathcal{A}_{I,x}$.

(S.2) If $(w, v) \in \mathcal{B}_i$ and $(w, u) \in \mathcal{A}_{I,x}$ then $(v, u) \in \mathcal{A}_{I,x}$.

Discussion

$\Rightarrow$ Acceptances are public

By the interaction principles $\mathbf{A}_{I:x}\varphi \rightarrow \mathbf{B}_i\mathbf{A}_{I:x}\varphi$ and $\neg\mathbf{A}_{I:x}\varphi \rightarrow \mathbf{B}_i\neg\mathbf{A}_{I:x}\varphi$ (with $i \in I$) we can infer:

$$\mathbf{A}_{I:x}\varphi \leftrightarrow \bigwedge_{1 \leq k \leq n} \mathbf{EB}_I^k\mathbf{A}_{I:x}\varphi$$

$$\neg\mathbf{A}_{I:x}\varphi \leftrightarrow \bigwedge_{1 \leq k \leq n} \mathbf{EB}_I^k\neg\mathbf{A}_{I:x}\varphi$$

Discussion (cont.)

⇒ Acceptance and belief might be incompatible: some agents can privately disbelieve something they accept while functioning as members of a given group (or organization, or team, or institution, etc.)

Example

At the end of the 80s, the Communist Party of Ruritania accepted that capitalist countries will soon perish but none of its members really believed so (Tuomela, 1992):

$$\mathbf{A}_{I:CPR} ccwp \wedge \bigwedge_{i \in I} \neg \mathbf{B}_i ccwp$$

should be satisfiable.

Discussion (cont.)

$\Rightarrow$ Collective acceptances *could be* built by the expression of unanimous opinions to the other members of the group

In certain situations the principle

$$\mathbf{A}_{I:x}(\bigwedge_{i \in I} \mathbf{B}_i \varphi \rightarrow \varphi)$$

sounds reasonable

Example

WHO members accept that if each of them expresses the opinion that ‘swine flu’ should be considered to be pandemic then ‘swine flu’ is pandemic:

$$\mathbf{A}_{I:WHO}(\bigwedge_{i \in I} \mathbf{B}_i pandemic \rightarrow pandemic)$$

Suppose WHO members express unanimous opinions on the issue:

$$\mathbf{A}_{I:WHO}(\bigwedge_{i \in I} \mathbf{B}_i pandemic)$$

It follows that that the WHO members accept that ‘swine flu’ is pandemic:

$$\mathbf{A}_{I:WHO} pandemic$$

Discussion (cont.)

The formula

$$\mathbf{A}_{I:x}(\bigwedge_{i \in I} \mathbf{B}_i \varphi \rightarrow \varphi)$$

cannot be taken as a logical axiom which is valid for every institution x, for every set of agents I and for every formula φ

Counterexample: symbolic game between two children (Piaget, 1951)

Two children are playing a game which consists in ‘changing the natural order of things’ through imagination.

1 and 2 could accept *qua* players of the game that a broom is a horse ($BisH$) and ‘riding’ it, i.e.

$$\mathbf{A}_{\{1,2\}:game} BisH \wedge \neg \mathbf{A}_{\{1,2\}:game} \bot,$$

even if they accept that each of them believes that the broom is not a horse, i.e.

$$\mathbf{A}_{\{1,2\}:game}(\mathbf{B}_2 \neg BisH \wedge \mathbf{B}_2 \neg BisH).$$

The previous two formulas are inconsistent with the formula

$$\mathbf{A}_{\{1,2\}:game}((\mathbf{B}_1 \neg BisH \wedge \mathbf{B}_2 \neg BisH) \rightarrow \neg BisH).$$

Discussion (cont.)

$\Rightarrow$ Belief aims at truth, while acceptance does not necessarily so (Engel, 1998)

The following is a theorem of doxastic logic:

$\vdash \mathbf{B}_i(\mathbf{B}_i\varphi \rightarrow \varphi)$
Proof:

1. $\vdash \neg\mathbf{B}_i\varphi \rightarrow \mathbf{B}_i\neg\mathbf{B}_i\varphi$ Axiom 5 for $\mathbf{B}_i$
2. $\vdash \mathbf{B}_i\varphi \vee \mathbf{B}_i\neg\mathbf{B}_i\varphi$ From 1
3. $\vdash \mathbf{B}_i(\varphi \vee \neg\mathbf{B}_i\varphi)$ From 2 by standard modal principles for $\mathbf{B}_i$
4. $\vdash \mathbf{B}_i(\mathbf{B}_i\varphi \rightarrow \varphi)$ From 3

Discussion (cont.)

In contrast, the formula $\mathbf{B}_i(\mathbf{A}_{i:x}\varphi \rightarrow \varphi)$ should not be valid

Example

Consider the lawyer who at court accepts his client is innocent, and believes so, i.e. $\mathbf{B}_{i_1}\mathbf{A}_{i_1:court}innocent$, while privately believing the contrary, i.e. $\mathbf{B}_{i_1}\neg innocent$. If $\mathbf{B}_i(\mathbf{A}_{i:x}\varphi \rightarrow \varphi)$ was valid then this would entail $\mathbf{B}_{i_1}\mathbf{A}_{i_1:court}innocent \rightarrow \mathbf{B}_{i_1}innocent$.

References

- Aumann, R. J. (1976). *Agreeing to Disagree*. The Annals of Statistics, 4(6): 1236-1239.
- Aumann, R. J. (1999). *Interactive epistemology I: Knowledge*. International Journal of Game Theory, 28(3): 263-300.
- Battigalli, P., Bonanno, G. (1999). *Recent results on belief, knowledge and the epistemic foundations of game theory*. Research in Economics, 53:149-225.
- Bacharach, M. (1992). *The Acquisition of Common Knowledge*. In Bicchieri C., Dalla Chiara M.L. (Eds.), Knowledge, Belief and Strategic Interaction, Cambridge University Press.
- Bicchieri, C. (1989). *Self-refuting theories of strategic interaction: a paradox of common knowledge*. Erkenntnis, 30:69-85.
- Clark, H. H., Marshall C. (1981). *Definite reference and mutual knowledge*. In A. Joshi, B. Webber, and I. Sag (Eds.), Elements of Discourse Understanding, Cambridge University Press.
- van Ditmarsch, H., van der Hoek, W., and Kooi, B. (2007). *Dynamic Epistemic Logic*, volume 337 of Synthese Library Series. Springer.
- Engel, P. (1998). *Believing, holding true, and accepting*. Philosophical Explorations, 1(2):140-151.

References

- Fagin, R., Halpern, J., Moses, Y., and Vardi, M. (1995). *Reasoning about Knowledge*. MIT Press.
- Gaudou, B., Longin, D., Lorini, E., and Tummolini, L. (2008). *Anchoring Institutions in Agents' Attitudes: Towards a Logical Framework for Autonomous MAS*. In Proc. of the Seventh International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'08), pages 728-735. ACM Press.
- Gilbert, M. (1989). *On Social Facts*. Routledge.
- Grosz, B.J., Kraus, S. (1996). *Collaborative Plans for Complex Group Action*. Artificial Intelligence, 86:269-357.
- Herzig, A., de Lima, T., Lorini, E. (2009). *On the dynamics of institutional agreements*. Synthese, to appear.
- Lewis, D. K. (1969). *Convention: a philosophical study*. Harvard University Press.
- Lorini, E., Longin, D., Gaudou, B., Herzig, A. (2009). *The logic of acceptance: grounding institutions on agents' attitudes*. Journal of Logic and Computation, to appear (available at http://logcom.oxfordjournals.org/cgi/content/short/exn103v3).

References

- Lorini, E., Longin, D. (2008). *A logical account of institutions: from acceptances to norms via legislators*. In Proc. of the Eleventh International Conference on Principles on Principles of Knowledge Representation and Reasoning (KR 2008), AAAI Press.
- Meyer, J.-J. C., Van der Hoek, W. (1995). *Epistemic Logic for AI and Theoretical Computer Science*. Cambridge University Press.
- Pettit, P. (2001). *Deliberative democracy and the discursive dilemma*. Philosophical Issues, 11:268-299.
- Quinton, A. (1975). *Social Objects*. In Proc. of the Aristotelian Society, 75, pages 38-48.
- Schiffer, S. (1972). *Meaning*. Oxford University Press.
- Stalnaker, R. (2001). *Common Ground*. Linguistics and Philosophy, 25(5-6):701-721.
- Tuomela, R. (1992). *Group beliefs*. Synthese, 91:285-318.
- Tuomela, R. (2002). *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge University Press.
- Tuomela, R. (2007). The Philosophy of Sociality. Oxford University Press.