The Logic of Acceptance:
a bottom-up perspective on institutional reality

Emiliano Lorini
(some parts based on joint works
with Andreas Herzig and Tiago de Lima)

Institut de Recherche en Informatique de Toulouse (IRIT), France

Formal Models of Norm Change 2
Amsterdam, 18-19 January 2010

Bottom-up perspective on institutions (cont.)

"*Only because institutions are anchored in peoples minds do they ever become behaviorally relevant. The elucidation of the internal aspect is the crucial step in adequately explaining the emergence, evolution, and effects of institutions*" [North, 2004].

Bottom-up perspective on institutions (cont.)

- an institution (rules, norms, etc.) is anchored in the agents' attitudes, and
- its dynamics depend on the dynamics of the attitudes of its members

What are the attitudes on the top of which institutions are built?

Acceptance *qua* members of an institution

- Agents can identify themselves as members of the same institution and recognize mutually as members of the same institution (Gilbert, 1989)
- Agents can accept things *qua* members of the same institution
- The existence and dynamics of norms of an institution depend on their acceptances by the institution members and on how these acceptances change over time

Example

The existence and dynamics of the rules of a certain game (e.g. chess) depend on their acceptances by the game players

Individual acceptance vs. collective acceptance

- Individual acceptance: a certain agent i accepts that something is true, *qua* member of a certain institution
- Collective acceptance: the agents in C accept that something is true, *qua* member of the same institution

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.

Our proposal

- A modal logic of acceptance in institutional contexts
- Analysis of three operations on acceptance and related institution dynamics:
  - **Announcement in an institutional context**: assertions and declarations in institutional contexts (e.g. voting for an agreement, collective decision/judgment)
  - **Acceptance shift**: form of paradigm shift for acceptance (radical change in an institution, institution creation)
  - **Assignment on acceptance**: effects of norm promulgation and derogation on acceptance

Acceptance logic (AL)

Normative concepts

Making AL dynamic I: announcement in an institutional context

Making AL dynamic II: acceptance shifting

Making AL dynamic III: assignments on acceptance

Extending Acceptance logic with beliefs

Acceptance logic (AL)

Acceptance Logic (AL)

- AGT: a finite set of agents;
- P: a countable set of atomic formulas;
- X: a finite set of institutional contexts.

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

Language:

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

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

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

Acceptance Logic (AL)

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

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

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

$\mathbf{U}\varphi$
'φ is universally true'.

AL models

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

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

$\mathcal{A}_{G,x}(w) = \{v | (w, v) \in \mathcal{A}_{G,x}\}$: the worlds *accepted* by the agents in G while functioning as members of institution x at world w (G'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 $(w, v) \in \mathcal{A}_{H,y}$ and $(v, u) \in \mathcal{A}_{G,x}$ then $(w, u) \in \mathcal{A}_{G,x}$.

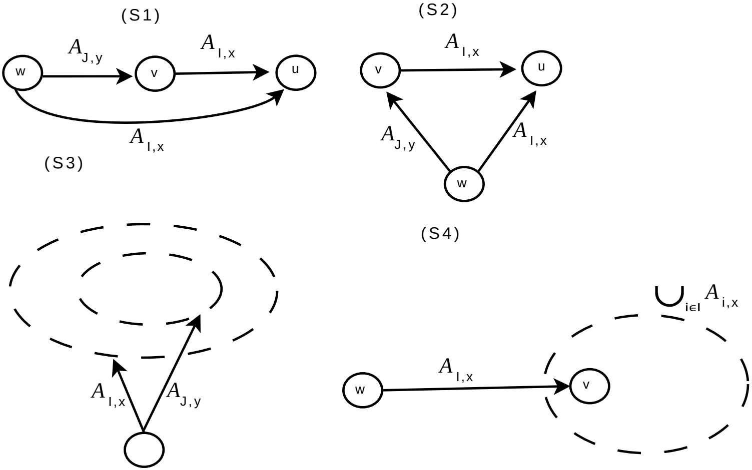
(S.2) If $(w, v) \in \mathcal{A}_{H,y}$ and $(w, u) \in \mathcal{A}_{G,x}$ then $(v, u) \in \mathcal{A}_{G,x}$.

(S.3) If $\mathcal{A}_{G,x}(w) \neq \emptyset$ then $\mathcal{A}_{H,x} \subseteq \mathcal{A}_{G,x}(w)$.

(S.4) If $v \in \mathcal{A}_{G,x}(w)$ then $v \in \bigcup_{i \in G} \mathcal{A}_{i,x}(v)$.

Constraints on AL models (cont.)

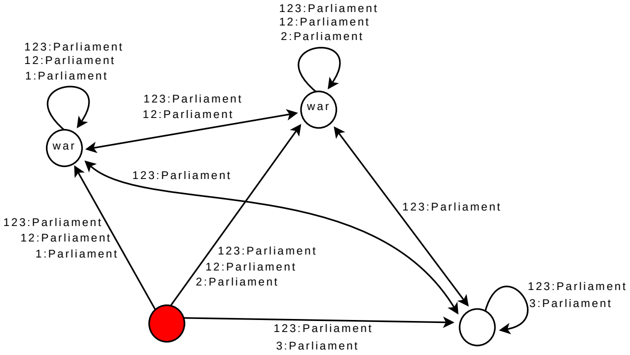
For $J \subseteq I$



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 $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models \mathbf{A}_{G:x}\varphi$ iff $\mathcal{M}, v \models \varphi$ for all $(w, v) \in \mathcal{A}_{G,x}$

Example: agreements and disagreements



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

is true at the red world

A complete axiomatization of AL

All K-principles for every $\mathbf{A}_{G:x}$.

All S5-principles for $\mathbf{U}$.

(4*) $\mathbf{A}_{G:x}\varphi \to \mathbf{A}_{H:y}\mathbf{A}_{G:x}\varphi$ if $H \subseteq G$

(5*) $\neg\mathbf{A}_{G:x}\varphi \to \mathbf{A}_{H:y}\neg\mathbf{A}_{G:x}\varphi$ if $H \subseteq G$

($\mathrm{Inc}_{\mathbf{A}}$) $(\widehat{\mathbf{A}}_{G:x}\top \wedge \mathbf{A}_{G:x}\varphi) \to \mathbf{A}_{H:x}\varphi$ if $H \subseteq G$

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

($\mathrm{Inc}_{\mathbf{A},\mathbf{U}}$) $\mathbf{U}\varphi \to \mathbf{A}_{G:x}\varphi$

Theorem

AL *is decidable.*

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 (Inc$_{\mathbf{A}}$) we infer

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

Example 2

Agents 1, 2 and 3 accept that 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 (4*) we infer:

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

$\Rightarrow$ Every group accepts (the validity of) other groups' acceptances

Some properties

Theorem

- $\vdash \mathbf{A}_{G:x}\widehat{\mathbf{A}}_{G:x}\top$
- $\vdash \mathbf{A}_{G:x}\varphi \rightarrow \mathbf{A}_{G:x}\mathbf{A}_{H:x}\varphi$ *if* $H \subseteq G$
- $\vdash \mathbf{A}_{H:y}\mathbf{A}_{G:x}\varphi \leftrightarrow (\mathbf{A}_{H:y}\bot \vee \mathbf{A}_{G:x}\varphi)$ *if* $H \subseteq G$
- $\vdash \mathbf{A}_{H:y}\neg\mathbf{A}_{G:x}\varphi \leftrightarrow (\mathbf{A}_{H:y}\bot \vee \neg\mathbf{A}_{G:x}\varphi)$ *if* $H \subseteq G$
- $\vdash \mathbf{A}_{G:x}(\mathbf{A}_{G:x}\varphi \rightarrow \varphi)$
- $\vdash (\bigwedge_{i\in G} \mathbf{A}_{G:x}\mathbf{A}_{i:x}\varphi) \rightarrow \mathbf{A}_{G:x}\varphi$

Some invalid properties

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

$\Rightarrow$ Institution membership is 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.)

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

$\Rightarrow$ Institution membership is 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 G and H such that $H \subseteq G$ and $|G \setminus H| < |H|$ (i.e. H represents the majority of agents in G):

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

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

Theorem

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

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

Normative concepts

Attitude-dependent institutional facts

Institutional facts are characterized by two features (Tuomela, 2002):

- **PERFORMATIVITY**: if the members of an institution accept a certain fact to be true then, the fact becomes true according to the institution
- **REFLEXIVITY**: if a certain fact is true according to an institution then, the members of the institution accept that the fact is true

Example

If the members of a certain institution accept a certain piece of paper as money, then, this piece of paper is money for the institution (PERFORMATIVITY).
If it is true that a certain piece of paper is money according to the institution then, the members of the institution accept the piece of paper as money (REFLEXIVITY).

Attitude-dependent institutional facts (cont.)

For formal institutions PERFORMATIVITY and REFLEXIVITY should be defined wrt authorities/leaders/legislators:

- **PERFORMATIVITY** (formal case): if the authorities of an institution accept a certain fact to be true then, the fact becomes true according to the institution
- **REFLEXIVITY** (formal case): if a certain fact is true according to an institution then, the authorities of the institution accept that the fact is true

Attitude-dependent institutional facts (cont.)

$$Auth : X \longrightarrow 2^{AGT}$$

- $Auth(x)$ is the set of authorities/leaders/legislators of institution x
- if $Auth(x) = \emptyset$ then x is an informal institution, otherwise x is a formal one

Definition

Institutional truth (informal case). If $Auth(x) = \emptyset$:
$[x]\varphi \stackrel{def}{=} \bigwedge_{G \in 2^{AGT*}} \mathbf{A}_{G:x}\varphi$

Definition

Institutional truth (formal case). If $Auth(x) \neq \emptyset$:
$[x]\varphi \stackrel{def}{=} \mathbf{A}_{Auth(x):x}\varphi$

- $[x]\varphi$ means "φ is true according to institution x"

Attitude-dependent institutional facts (cont.)

Theorem

- $\vdash ([x](\varphi \rightarrow \psi) \wedge [x]\varphi) \rightarrow [x]\psi$
- *From* $\vdash \varphi$ *infer* $\vdash [x]\varphi$
- $\vdash [x]\varphi \rightarrow [x][x]\varphi$ *if* $Auth(x) \neq \emptyset$
- $\vdash \neg[x]\varphi \rightarrow [x]\neg[x]\varphi$ *if* $Auth(x) \neq \emptyset$

For the legal case, $[x]$ is K45 (it embeds Grossi et al.'s logic, 2006)

Counts-as

Definition

Counts-as (Searle, 1995). $\varphi \overset{x}{\rhd} \psi \overset{def}{=} [x](\varphi \to \psi)$

- $\varphi \overset{x}{\rhd} \psi$ is meant to stand for: "φ counts as ψ in the context of institution x"
  - *Classificatory rule* in the sense of Grossi et al. (2006)

Example

In the language of gesture, the nodding gesture "counts as" an endorsement of what the speaker is suggesting: $[gesture](nodding \to yes)$.

Counts-as (cont.)

[Jones and Sergot, 1996]

Theorem

- *From* $\vdash (\varphi_2 \leftrightarrow \varphi_3)$ *infer* $\vdash (\varphi_1 \overset{x}{\rhd} \varphi_2 \leftrightarrow \varphi_1 \overset{x}{\rhd} \varphi_3)$
- *From* $\vdash (\varphi_1 \leftrightarrow \varphi_3)$ *infer* $\vdash (\varphi_1 \overset{x}{\rhd} \varphi_2 \leftrightarrow \varphi_3 \overset{x}{\rhd} \varphi_2)$
- $\vdash (\varphi_1 \overset{x}{\rhd} \varphi_2 \wedge \varphi_1 \overset{x}{\rhd} \varphi_3) \rightarrow (\varphi_1 \overset{x}{\rhd} (\varphi_2 \wedge \varphi_3))$
- $\vdash (\varphi_1 \overset{x}{\rhd} \varphi_2 \wedge \varphi_3 \overset{x}{\rhd} \varphi_2) \rightarrow ((\varphi_1 \vee \varphi_3) \overset{x}{\rhd} \varphi_2)$
- $\vdash (\varphi_1 \overset{x}{\rhd} \varphi_2 \wedge (\varphi_1 \wedge \varphi_2) \overset{x}{\rhd} \varphi_3) \rightarrow (\varphi_1 \overset{x}{\rhd} \varphi_3)$

Obligations and permissions

Definition

Obligation. $\mathbf{O}_x\varphi \stackrel{def}{=} \neg\varphi \stackrel{x}{\rhd} \mathrm{V}$

- "φ is something obligatory within the institutional context x" (noted $\mathbf{O}_x\varphi$) if and only if "$\neg\varphi$ counts as a violation in the context of institution x"
- V is an atom for *violation* (Anderson, 1958)

Definition

Permission. $\mathbf{P}_x\varphi \stackrel{def}{=} \neg\mathbf{O}_x\neg\varphi$

Making AL dynamic I: announcement in an institutional context

From a Static Point of View to a Dynamic Point of View

- Announcements are commonly studied in combination with knowledge and belief, such as in Dynamic Epistemic Logic (DEL)
- We here combine announcements with acceptances

AL with Announcements

AL$^+$ extends AL by formulas $[x!\psi]\varphi$

$x!\psi$
'it is publicly announced in the institution x that ψ'.

$[x!\psi]\varphi$
'φ holds after the announcement of ψ in x'.

$x!\mathbf{A}_{i:x}p$
'it is publicly announced in institution x that i accepts p in x'.

$\Rightarrow$ It approximates i's action of announcing p in the context x: an assertion in the sense of Speech Act theory.

Semantics of AL$^+$

Models and the satisfaction relation are as before plus:

$$\langle W, \mathcal{A}, \mathcal{V} \rangle, w \models [x!\psi]\varphi$$

iff

$$\langle W, \mathcal{A}^{x!\psi}, \mathcal{V} \rangle, w \models \varphi$$

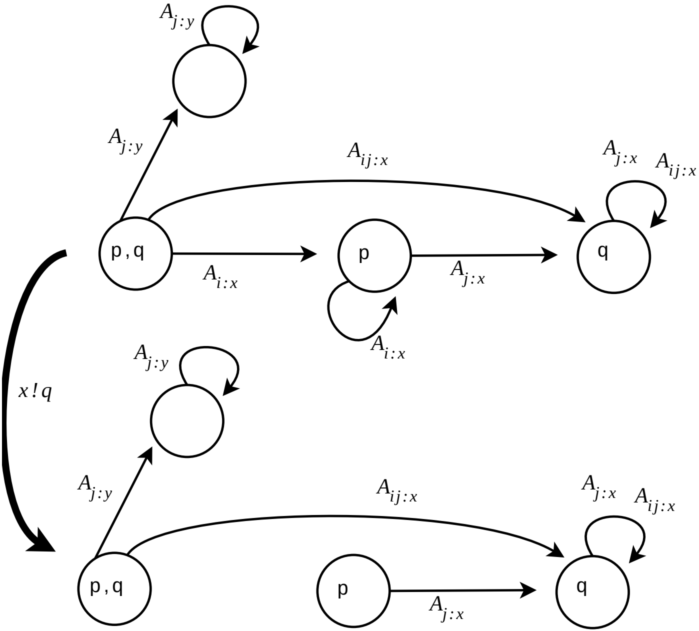
where:

- $\mathcal{A}^{x!\psi}(G, y)(w) = \mathcal{A}(G, y)(w)$, if $y \neq x$.
- $\mathcal{A}^{x!\psi}(G, y)(w) = \mathcal{A}(G, y)(w) \cap \|\psi\|_{\mathcal{M}}$, if $y = x$.

Theorem

If $\mathcal{M}$ is an acceptance model then $\mathcal{M}^{x!\psi}$ is an acceptance model.

Example



$A_{j:y}$
$A_{j:y}$
$A_{ij:x}$
$A_{j:x}$
$A_{ij:x}$
p , q
p
q
$A_{i:x}$
$A_{j:x}$
$A_{i:x}$
$x!q$
$A_{j:y}$
$A_{j:y}$
$A_{ij:x}$
$A_{j:x}$
$A_{ij:x}$
p , q
p
q
$A_{j:x}$

Axiomatization of AL^+

All axioms and inference rules of logic AL.

(R.1) $[x!\psi]p \leftrightarrow p$

(R.2) $[x!\psi]\neg\varphi \leftrightarrow \neg[x!\psi]\varphi$

(R.3) $[x!\psi](\varphi_1 \wedge \varphi_2) \leftrightarrow ([x!\psi]\varphi_1 \wedge [x!\psi]\varphi_2)$

(R.4) $[x!\psi]\mathbf{A}_{G:y}\varphi \leftrightarrow \mathbf{A}_{G:y}[x!\psi]\varphi$ if $y \neq x$

(R.5) $[x!\psi]\mathbf{A}_{G:y}\varphi \leftrightarrow \mathbf{A}_{G:y}(\psi \rightarrow [x!\psi]\varphi))$ if $y = x$

(R.6) Rule of replacement of proved equivalence

Theorem

This is a sound and complete axiomatization w.r.t. the class of acceptance models.

Some properties

Theorem

- $\vdash [x!\varphi]\mathbf{A}_{G:x}\varphi$ *if φ is Boolean*
- $\vdash \mathbf{A}_{G:x}\varphi \rightarrow [x!\neg\varphi]\mathbf{A}_{G:x}\bot$
- $\vdash [x!\psi][x!\psi']\varphi \leftrightarrow [x!\psi'][x!\psi]\varphi$ *if ψ, ψ' are positive formulas*
- $\vdash \varphi \rightarrow [x!\varphi]\varphi$ *if φ is Boolean*
- $\vdash \neg\mathbf{A}_{G:x}p \rightarrow [x!\neg\mathbf{A}_{G:x}p]\neg\mathbf{A}_{G:x}p$
- $\vdash [x!(p \wedge \neg\mathbf{A}_{G:x}p)]\neg(p \wedge \mathbf{A}_{G:x}p)$

Establishing norms by agreement during dialogue

⇒ 1, 2, 3 are the committee members of the Master program in AI at Toulouse University. They have to decide whether Master thesis should be written in French, and use majority to decide. If $H \subseteq \{1,2,3\}$ and $|H| \geq 2$, $P \in \{Fr, \neg Fr\}$:

$$\mathbf{A}_{123:c}(\bigwedge_{i \in H} \mathbf{A}_{i:c} P \to P)$$

⇒ It follows that, after each agent in $\{1,2\}$ *qua* member of the committee declares that thesis should be written in French, 1, 2 and 3 start to accept that thesis should be written in French:

$$[c!\mathbf{A}_{1:c}Fr \wedge \mathbf{A}_{2:c}Fr]\mathbf{A}_{123:c}Fr$$

⇒ If 1, 2 and 3 accept that at least one in $\{1,2\}$ does not accept that thesis should be written in French, the committee is “dissolved” after the announcement:

$$\mathbf{A}_{123:c}\neg(\mathbf{A}_{1:c}Fr \wedge \mathbf{A}_{2:c}Fr) \to [c!\mathbf{A}_{1:c}Fr \wedge \mathbf{A}_{2:c}Fr]\mathbf{A}_{123:c}\bot$$

⇒ Need for acceptance revision/acceptance shifting!

Making AL dynamic II: acceptance shifting

AL with Announcements and Shifting

AL^{++} extends AL^{+} by formulas $[G{:}x{\uparrow}\psi]\varphi$

$G{:}x{\uparrow}\psi$
'the agents in G, *qua* members of x, start to accept ψ (or shift their acceptance to ψ)'

$[G{:}x{\uparrow}\psi]\varphi$
'φ holds after the shifting of G to ψ in the institutional context x'.

The operation $G{:}x{\uparrow}\psi$ can be used for modelling:
⇒ institution formation (From $\mathbf{A}_{G:x}\bot$ To $\widehat{\mathbf{A}}_{G:x}\top \wedge \mathbf{A}_{G:x}\psi$);
⇒ a radical change in an institution (form of revision for collective attitudes or paradigm shift)

Semantics of AL^{++}

Models and the satisfaction relation are as before plus:

$$\langle W, \mathcal{A}, \mathcal{V} \rangle, w \models [G{:}x{\uparrow}\psi]\varphi$$

iff

$$\langle W, \mathcal{A}^{G:x\uparrow\psi}, \mathcal{V} \rangle, w \models \varphi$$

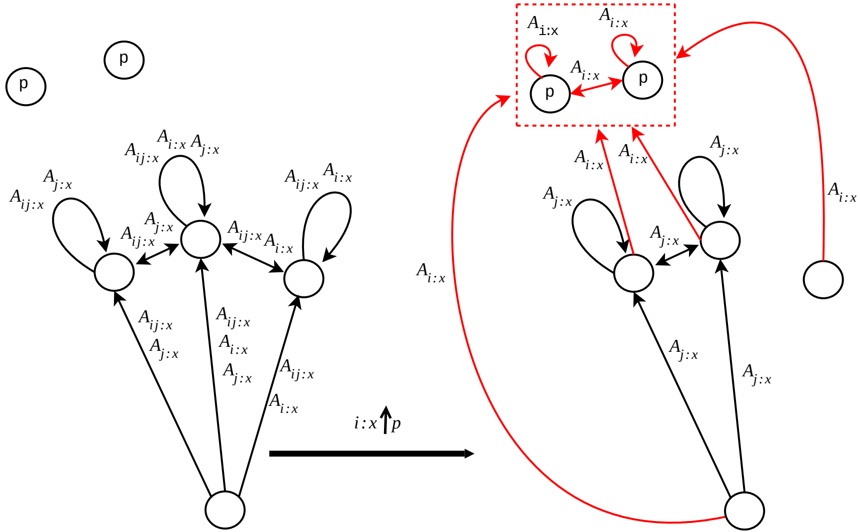
where:

- $\mathcal{A}^{G:x\uparrow\psi}_{H:y}(w) = \mathcal{A}_{H:y}(w)$, if $x \neq y$ or $G \cap H = \emptyset$.
- $\mathcal{A}^{G:x\uparrow\psi}_{H:y}(w) = \emptyset$, if $x = y$, $G \cap H \neq \emptyset$ and $H \not\subseteq G$.
- $\mathcal{A}^{G:x\uparrow\psi}_{H:y}(w) = \|\psi\|_{\mathcal{M}}$, if $x = y$ and $H \subseteq G$.

Theorem

If $\mathcal{M}$ is an acceptance model then $\mathcal{M}^{G:x\uparrow\psi}$ is an acceptance model.

Example



Axiomatization of AL^{++}

All axioms and inference rules of logic AL^{+}.

(R.7) $[G{:}x{\uparrow}\psi]p \leftrightarrow p$

(R.8) $[G{:}x{\uparrow}\psi]\neg\varphi \leftrightarrow \neg[G{:}x{\uparrow}\psi]\varphi$

(R.9) $[G{:}x{\uparrow}\psi](\varphi_1 \wedge \varphi_2) \leftrightarrow ([G{:}x{\uparrow}\psi]\varphi_1 \wedge [G{:}x{\uparrow}\psi]\varphi_2)$

(R.10) $[G{:}x{\uparrow}\psi]\mathbf{A}_{H:y}\varphi \leftrightarrow \mathbf{A}_{H:y}[G{:}x{\uparrow}\psi]\varphi$
if $x \neq y$ or $G \cap H = \emptyset$

(R.11) $[G{:}x{\uparrow}\psi]\mathbf{A}_{H:y}\varphi \leftrightarrow \bot$
if $x = y$, $G \cap H \neq \emptyset$ and $H \not\subseteq G$

(R.12) $[G{:}x{\uparrow}\psi]\mathbf{A}_{H:y}\varphi \leftrightarrow \mathbf{U}(\psi \rightarrow [G{:}x{\uparrow}\psi]\varphi)$
if $x = y$ and $H \subseteq G$

(R.13) $[G{:}x{\uparrow}\psi]\mathbf{U}\varphi \leftrightarrow \mathbf{U}[G{:}x{\uparrow}\psi]\varphi$

Theorem

This is a sound and complete axiomatization w.r.t. the class of acceptance models.

Example: changing paradigm

⇒ Before French Revolution, French citizens accept that the king is the authority:

$$\neg \mathbf{A}_{G:France}\bot \wedge \mathbf{A}_{G:France} kingAuth.$$

⇒ After French revolution, French citizens start to accept the contrary (acceptance shifting):

$$\neg \mathbf{U}\ kingAuth \rightarrow [G{:}France{\uparrow}\neg kingAuth] \\ (\neg \mathbf{A}_{G:France}\bot \wedge \mathbf{A}_{G:France}\neg kingAuth).$$

Some properties

Theorem

- $\vdash [G{:}x{\uparrow}\varphi]\mathbf{A}_{G:x}\varphi$ *if φ is Boolean*
- $\vdash \neg\mathbf{U}\neg\varphi \to [G{:}x{\uparrow}\varphi]\neg\mathbf{A}_{G:x}\bot$
- $\vdash [G{:}x{\uparrow}\varphi]\mathbf{A}_{H:y}\mathbf{A}_{G:x}\varphi$ *if φ is Boolean*
- $\vdash \varphi \to [H{:}x{\uparrow}\mathbf{A}_{G:x}\bot]\varphi$ *if φ is Boolean*
- $\vdash \mathbf{A}_{G:x}\bot \to [H{:}x{\uparrow}\mathbf{A}_{G:x}\bot]\neg\mathbf{A}_{G:x}\bot$ *if $G \subseteq H$*

Comparison with AGM revision

$$[\alpha;\beta]\varphi \overset{def}{=} [\alpha][\beta]\varphi$$

$$[\psi!]\varphi \overset{def}{=} \psi \rightarrow \varphi$$

$$[\alpha \cup \beta]\varphi \overset{def}{=} [\alpha]\varphi \wedge [\beta]\varphi$$

$$[\text{if } \psi \text{ then } \alpha \text{ else } \beta]\varphi \overset{def}{=} [(\psi!;\alpha) \cup (\neg\psi!;\beta)]\varphi$$

$$[i{:}x \star \psi]\varphi \overset{def}{=} [\text{ if } \neg\mathbf{A}_{i:x}\neg\psi \text{ then } x!\psi \text{ else } i{:}x{\uparrow}\psi]\varphi$$

Comparison with AGM revision (cont.)

[Katsuno & Mendelzon, 1992]

R1. $\varphi \star \psi \models \psi$

R2. if $\varphi \not\models \psi$ then $\varphi \star \psi \equiv \varphi \wedge \psi$

R3. if $\varphi \star \psi \models \bot$ then $\psi \models \bot$

R4. if $\models \varphi \leftrightarrow \varphi'$ and $\models \psi \leftrightarrow \psi'$ then $\models \varphi \star \psi \leftrightarrow \varphi' \star \psi'$

Theorem

Let φ and ψ be Boolean then:

- $\vdash \mathbf{A}_{i:x}\varphi \to [i{:}x \star \psi]\mathbf{A}_{i:x}\psi$
- $\vdash (\mathbf{A}_{i:x}\varphi \wedge \widehat{\mathbf{A}}_{i:x}\psi) \to [i{:}x \star \psi]\mathbf{A}_{i:x}(\varphi \wedge \psi)$
- $\vdash [i{:}x \star \psi]\mathbf{A}_{i:x}\bot \to \mathbf{U}\neg\psi$
- *if* $\vdash \varphi \leftrightarrow \varphi'$ *and* $\vdash \psi \leftrightarrow \psi'$ *then*
  $\vdash (\mathbf{A}_{i:x}\varphi \to [i{:}x \star \psi]\mathbf{A}_{i:x}\chi) \leftrightarrow (\mathbf{A}_{i:x}\varphi' \to [i{:}x \star \psi']\mathbf{A}_{i:x}\chi)$

Making AL dynamic III: assignments on acceptance

AL with Announcements, Shifting and Assignments

AL^{+++} extends AL^{++} by $[p \overset{x}{\leadsto} \psi, x]\varphi$ and $[p \overset{x}{\leadsto} \psi, {\sim}x]\varphi$

$p \overset{x}{\leadsto} \psi$
'the truth value of ψ is assigned to p in the institution x'

$[p \overset{x}{\leadsto} \psi, x]\varphi$
'φ holds in the context of institution x, after assigning the truth value of ψ to p in x'

$[p \overset{x}{\leadsto} \psi, {\sim}x]\varphi$
'φ holds in the context of every institution *different* from x, after assigning the truth value of ψ to p in x'

$[p \overset{x}{\leadsto} \psi]\varphi \overset{def}{=} [p \overset{x}{\leadsto} \psi, x]\varphi \wedge [p \overset{x}{\leadsto} \psi, {\sim}x]\varphi$
'φ holds, after assigning the truth value of ψ to p in x'

Semantics of AL^{+++}

The satisfaction relations are as before plus:

$$\langle W, \mathcal{A}, \mathcal{V} \rangle, w \models [p \overset{x}{\rightsquigarrow} \psi, x]\varphi$$

iff

$$\langle W^{p \overset{x}{\rightsquigarrow} \psi}, \mathcal{A}^{p \overset{x}{\rightsquigarrow} \psi}, \mathcal{V}^{p \overset{x}{\rightsquigarrow} \psi} \rangle, w_x \models \varphi$$

$$\langle W, \mathcal{A}, \mathcal{V} \rangle, w \models [p \overset{x}{\rightsquigarrow} \psi, \sim x]\varphi$$

iff

$$\langle W^{p \overset{x}{\rightsquigarrow} \psi}, \mathcal{A}^{p \overset{x}{\rightsquigarrow} \psi}, \mathcal{V}^{p \overset{x}{\rightsquigarrow} \psi} \rangle, w_{\sim x} \models \varphi$$

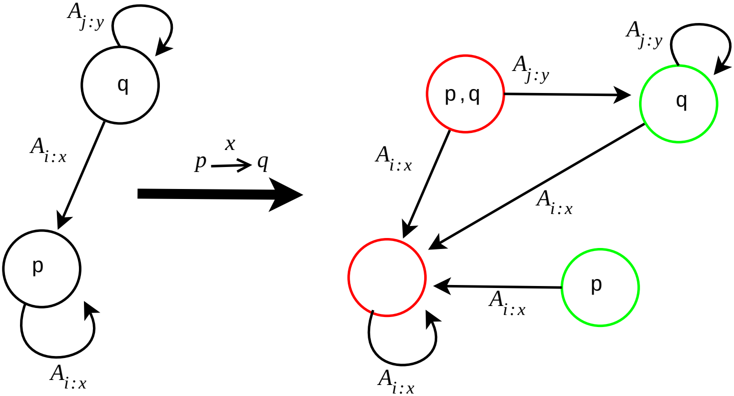
Semantics of AL^{+++} (cont.)

$$
\begin{aligned}
W^{p \overset{x}{\rightsquigarrow} \psi} =& \{w_x | w \in W\} \cup \\
& \{w_{\sim x} | w \in W\}; \\
\mathcal{A}_{G,x}^{p \overset{x}{\rightsquigarrow} \psi} =& \{(w_x, v_x) | v, w \in W \text{ and } (w, v) \in \mathcal{A}_{G,x}\} \cup \\
& \{(w_{\sim x}, v_x) | v, w \in W \text{ and } (w, v) \in \mathcal{A}_{G,x}\}; \\
\text{If } y \neq x \text{ then, } \mathcal{A}_{G,y}^{p \overset{x}{\rightsquigarrow} \psi} =& \{(w_x, v_{\sim x}) | v, w \in W \text{ and } (w, v) \in \mathcal{A}_{G,y}\} \cup \\
& \{(w_{\sim x}, v_{\sim x}) | v, w \in W \text{ and } (w, v) \in \mathcal{A}_{G,y}\}; \\
\mathcal{V}^{p \overset{x}{\rightsquigarrow} \psi}(p) =& \{w_x | w \in W \text{ and } M, w \models \psi\} \cup \\
& \{w_{\sim x} | w \in W \text{ and } M, w \models p\}.
\end{aligned}
$$

Theorem

If $\mathcal{M}$ is an acceptance model then $\mathcal{M}^{p \overset{x}{\rightsquigarrow} \psi}$ is an acceptance model.

Example



Axiomatization of AL^{+++}

All axioms and inference rules of logic AL^{++}.

(R.14) $[p \overset{x}{\rightsquigarrow} \psi, x]p \leftrightarrow \psi$

(R.15) $[p \overset{x}{\rightsquigarrow} \psi, x]q \leftrightarrow q$ if $q \neq p$

(R.16) $[p \overset{x}{\rightsquigarrow} \psi, {\sim}x]q \leftrightarrow q$

(R.17) $[p \overset{x}{\rightsquigarrow} \psi, x]\neg\varphi \leftrightarrow \neg[p \overset{x}{\rightsquigarrow} \psi, x]\varphi$

(R.18) $[p \overset{x}{\rightsquigarrow} \psi, {\sim}x]\neg\varphi \leftrightarrow \neg[p \overset{x}{\rightsquigarrow} \psi, {\sim}x]\varphi$

(R.19) $[p \overset{x}{\rightsquigarrow} \psi, x](\varphi_1 \wedge \varphi_2) \leftrightarrow ([p \overset{x}{\rightsquigarrow} \psi, x]\varphi_1 \wedge [p \overset{x}{\rightsquigarrow} \psi, x]\varphi_2)$

(R.20) $[p \overset{x}{\rightsquigarrow} \psi, {\sim}x](\varphi_1 \wedge \varphi_2) \leftrightarrow$
$([p \overset{x}{\rightsquigarrow} \psi, {\sim}x]\varphi_1 \wedge [p \overset{x}{\rightsquigarrow} \psi, {\sim}x]\varphi_2)$

(R.21) $[p \overset{x}{\rightsquigarrow} \psi, x]\mathbf{A}_{H:x}\varphi \leftrightarrow \mathbf{A}_{H:x}[p \overset{x}{\rightsquigarrow} \psi, x]\varphi$

(R.22) $[p \overset{x}{\rightsquigarrow} \psi, {\sim}x]\mathbf{A}_{H:x}\varphi \leftrightarrow \mathbf{A}_{H:x}[p \overset{x}{\rightsquigarrow} \psi, x]\varphi$

(R.23) $[p \overset{x}{\rightsquigarrow} \psi, x]\mathbf{A}_{H:y}\varphi \leftrightarrow \mathbf{A}_{H:y}[p \overset{x}{\rightsquigarrow} \psi, {\sim}x]\varphi$ if $y \neq x$

(R.24) $[p \overset{x}{\rightsquigarrow} \psi, {\sim}x]\mathbf{A}_{H:y}\varphi \leftrightarrow \mathbf{A}_{H:y}[p \overset{x}{\rightsquigarrow} \psi, {\sim}x]\varphi$ if $y \neq x$

Theorem

This is a sound and complete axiomatization w.r.t. the class of acceptance models.

Example: obligation promulgation and derogation

Definition
Promulgation. $+_x(\varphi) \stackrel{def}{=} \mathrm{V} \stackrel{x}{\rightsquigarrow} (\mathrm{V} \vee \neg\varphi)$

Definition
Derogation. $-_x(\varphi) \stackrel{def}{=} \mathrm{V} \stackrel{x}{\rightsquigarrow} (\varphi \wedge \mathrm{V})$

Example
$\Rightarrow$ After promulgating in x the prohibition to smoke in public spaces ($\neg SP$), (in x) it is forbidden to smoke in public spaces:

$$[+_x(\neg SP)]\mathbf{O}_x\neg SP.$$

Example
$\Rightarrow$ If smoking in public spaces is conceivable by institution x then, after derogating in x the prohibition to smoke in public spaces, (in x) it is permitted to smoke in public spaces:

$$\langle x\rangle SP \rightarrow [-_x(\neg SP)]\mathbf{P}_x SP.$$

Some properties

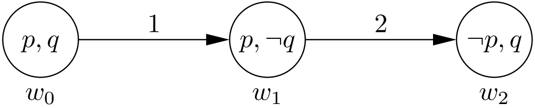
Theorem

Let φ be Boolean then:

- $\vdash [+_x(\varphi)]\mathbf{O}_x\varphi$
- $\vdash \mathbf{O}_y\varphi \rightarrow [+_x(\psi)]\mathbf{O}_y\varphi$
- $\vdash \mathbf{P}_y\varphi \rightarrow [+_x(\psi)]\mathbf{P}_y\varphi$ *if* $y \neq x$
- $\vdash \langle x\rangle\neg\varphi \rightarrow [-_x(\varphi)]\mathbf{P}_x\neg\varphi$
- $\vdash \mathbf{P}_y\varphi \rightarrow [-_x(\psi)]\mathbf{P}_y\varphi$
- $\vdash \mathbf{O}_y\varphi \rightarrow [-_x(\psi)]\mathbf{O}_y\varphi$ *if* $y \neq x$

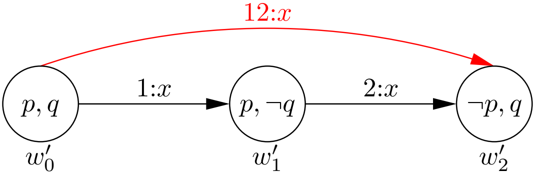
Common Belief vs. Collective Acceptance

Let $\mathcal{M}$ be the epistemic model:



$\mathcal{M}, w_0 \not\models \mathbf{CB}_{12}[!q]p$ and $\mathcal{M}, w_0 \models [!q]\mathbf{CB}_{12}p$

Group acceptances are not computed from the individual acceptances. The corresponding acceptance model $\mathcal{M}'$ has more arrows. In particular:



$\mathcal{M}', w'_0 \not\models \mathbf{A}_{12:x}[x!q]p$ and $\mathcal{M}', w'_0 \not\models [x!q]\mathbf{A}_{12:x}p$

Extending Acceptance logic with beliefs

Acceptance Logic with beliefs

Language of AL + B (Acceptance Logic with beliefs):

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

where G 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 Euclidean) accessibility relation $\mathcal{B}_i \subseteq W \times W$ for every $i \in AGT$.

Interaction principles between acceptance and belief

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

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

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

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

Discussion

$\Rightarrow$ Acceptances are public

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

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

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

Discussion (cont.)

$\Rightarrow$ Acceptance and belief might be incompatible: some agents can privately disbelieve something they accept while functioning as members of a given institution

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):

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

should be satisfiable.

Discussion (cont.)

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

In certain situations the principle

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

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:

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

Suppose WHO members express unanimous opinions on the issue:

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

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

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

Discussion (cont.)

The formula

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

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

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

Two children 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:

$\vdash \mathbf{B}_i(\mathbf{B}_i\varphi \to \varphi)$

Proof:

1. $\vdash \neg\mathbf{B}_i\varphi \to \mathbf{B}_i\neg\mathbf{B}_i\varphi$ 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. $\vdash \mathbf{B}_i(\mathbf{B}_i\varphi \to \varphi)$ 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 $\mathbf{B}_{i_1}\mathbf{A}_{i_1:court}innocent \rightarrow \mathbf{B}_{i_1}innocent$.

Thank you!