A Logic for Social Influence through Communication

Zoé Christoff

Institute for Logic, Language and Computation
University of Amsterdam

11th European Workshop on Multi-Agent Systems (EUMAS)
Logical Aspects of Multi-Agent System (LAMAS)
Toulouse, December 13 2013

Outline

1) Seligman, Girard & Liu (2011, 2014)

- social network
- peer pressure effects, influence inbetween “friends”

Outline

1) Seligman, Girard & Liu (2011, 2014)

- social network
- peer pressure effects, influence inbetween "friends"

?
$+$

2) Baltag & Smets (2009, 2013)

- plausibility
- effects of group members sharing information with the rest of the group

Outline

1) Seligman, Girard & Liu (2011, 2014)

- social network
- peer pressure effects, influence inbetween "friends"

?

+

2) Baltag & Smets (2009, 2013)

- plausibility
- effects of group members sharing information with the rest of the group

3) Aim: a unified social network plausibility framework

- model social influence on beliefs through communication among agents in a social network
- define some particular communication protocols (in the new framework) inspired by 2) to represent some level of influence as defined in 1)

1) Social influence à la Girard, Liu & Seligman

The framework

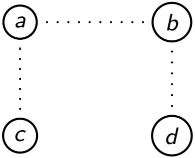
Static hybrid logic to represent who is friend with whom and who believes what
+ an (external) influence operator

The main ideas

- Agents are influenced by their friends and only by their friends.
- Simple "peer pressure principle": I tend to align with my friends.
- "Being influenced" is defined as "aligning my beliefs to the ones of my friends".
- No communication is (at least explicitly) involved. (transparency?)

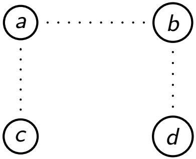
Friends network

Social network frame:



Friends network

Social network frame:



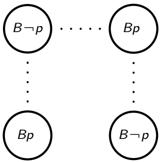
3 possible belief states (with respect to p)

- Bp
- $B\neg p$
- $Up := \neg Bp$ and $\neg B\neg p$

Belief revision induced by (direct) social influence

1) Strong influence

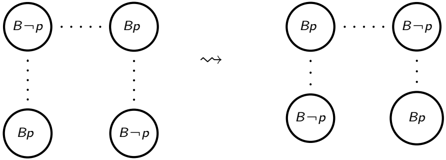
When all of my friends believe that p, I (successfully) *revise* with p. When all of my friends believe that $\neg p$, I (successfully) *revise* with $\neg p$.



Belief revision induced by (direct) social influence

1) Strong influence

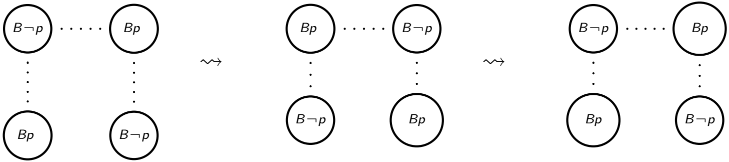
When all of my friends believe that p, I (successfully) *revise* with p. When all of my friends believe that $\neg p$, I (successfully) *revise* with $\neg p$.



Belief revision induced by (direct) social influence

1) Strong influence

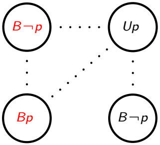
When all of my friends believe that p, I (successfully) *revise* with p. When all of my friends believe that $\neg p$, I (successfully) *revise* with $\neg p$.



Belief contraction induced by social influence

2) Weak influence

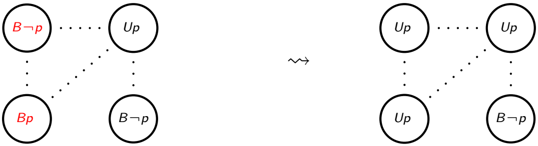
None of my friends supports my belief in p and some believe that $\neg p$.
I (successfully) *contract* it.
(And similarly for $\neg p$)



Belief contraction induced by social influence

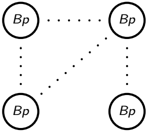
2) Weak influence

None of my friends supports my belief in p and some believe that $\neg p$.
I (successfully) *contract* it.
(And similarly for $\neg p$)



Stabilization

- Stable state: applying the social influence operator doesn't change the state of any agent.
- Stabilization: some configurations will reach a stable state after a finite number of applications of the influence operator (see example of weak influence above) and some won't (see example of strong influence).
- Sufficient condition for stability: all friends are in the same state.



2) Communication protocols à la Baltag & Smets

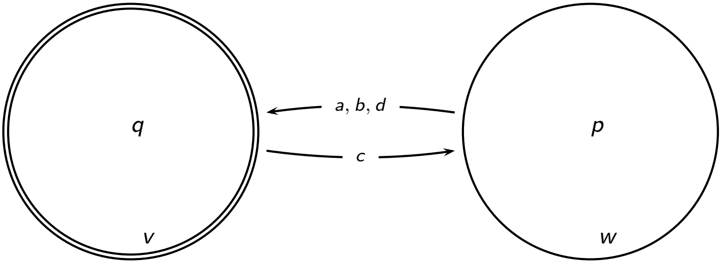
The framework

DEL type: plausibility modeling of (several) doxastic attitudes + communication events

The main ideas

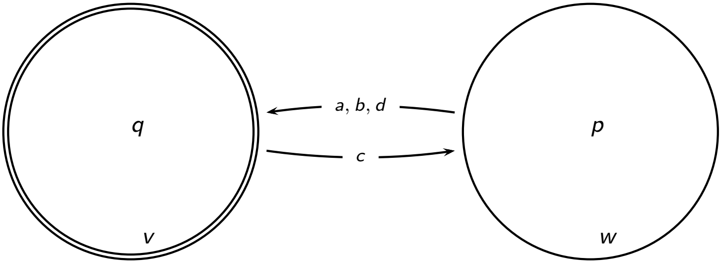
- Agents communicate via public announcements.
- Assuming that they trust each other enough, agents all revise their beliefs with each of the announced formula, sequentially.
- In this sense, each announcement influences everybody (else) into belief revision.

Plausibility model



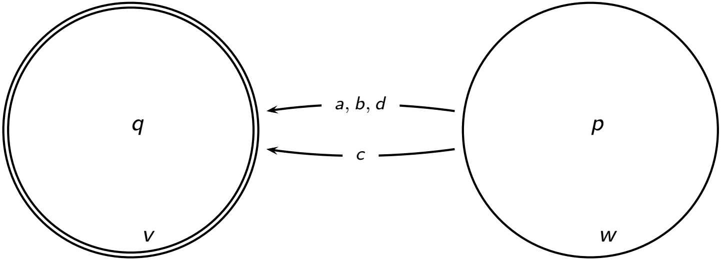
Plausibility model

q 



Plausibility model

q 



Reaching a stable state of agreement

How to communicate?

- Agents speak in turn (given expertise rank).
- An agent announces all and only (non-equivalent) sentences that she believes (exhaustivity + honesty).
- After a finite number of announcements (and corresponding revisions), everybody holds the same beliefs.
- This is a stable state: nothing which could be announced by any agent would change anything anymore.

Reaching a stable state of agreement

How to communicate?

- Agents speak in turn (given expertise rank).
- An agent announces all and only (non-equivalent) sentences that she believes (exhaustivity + honesty).
- After a finite number of announcements (and corresponding revisions), everybody holds the same beliefs.
- This is a stable state: nothing which could be announced by any agent would change anything anymore.

Lexicographic belief merge protocol

$$\rho_a := \prod\{\Uparrow \phi : \|\phi\| \subseteq S \text{ such that } \mathcal{M}, w \models B_a\phi\}$$

$$\rho_b := \prod\{\Uparrow \phi : \|\phi\| \subseteq S \text{ such that } \mathcal{M}_{[\rho_a]}, w \models B_b\phi\}$$

$$\text{etc for all } c \in \mathcal{A}$$

where $\prod$ is a sequential composition operator and $\mathcal{M}_{[\rho_a]}$ is the new model after joint revision with each formula announced by a.

Big picture

Common features

- Agents are influenced into revising their beliefs to make them closer to the ones of (some) others.
- A global agreement state is stable (both under honest communication and under social conformity pressure).

From 1)

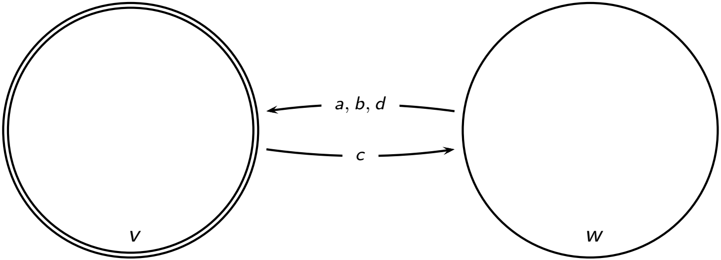
- **Social network**
- Synchronic
- **Over friends only**
- Equal power (among friends)
- Direct
- **Tools**: nominals, @, F

From 2) 

- **Plausibility**
- Sequential
- Over everybody
- Ranking
- **Via communication**
- **Tools:** $B, \uparrow, \Uparrow$

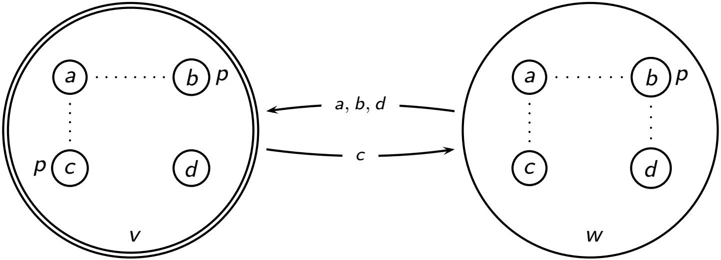
3) A social network plausibility framework

plausibility model:



3) A social network plausibility framework  + 

Social network plausibility model:



Social network plausibility model

$$\mathcal{M} = (S, \mathcal{A}, \leq_{a\in\mathcal{A}}, \|\cdot\|, s_0, \asymp_{s\in S})$$

- S is a (finite) set of possible states.
- $\mathcal{A}$ is a (finite) set of agents.
- $\leq_a \subseteq S \times S$ is a locally connected preorder, interpreted as the subjective plausibility relation of agent a, for each $a \in \mathcal{A}$
- $s_0 \in S$ is a designated state, interpreted as the actual state
- $\asymp_s \subseteq A \times A$ is an irreflexive and symmetric relation, interpreted as friendship, for each state $s \in S$
- $\|\cdot\| : \Phi \cup N \to \mathcal{P}(S \times \mathcal{A})$ is a valuation, assigning:
  - a set $\|p\| \subseteq S \times \mathcal{A}$ to every element p of some given set Φ of "atomic propositions"
  - a set $\|n\| = S \times \{a\}$ for some $a \in \mathcal{A}$ to every element n of some given set N of "nominals".

Syntax

$$\phi := p \mid n \mid \neg\phi \mid \phi \wedge \phi \mid F\phi \mid @n\,\phi \mid B\phi$$

where p belongs to a set of atomic propositions Φ and n to a set of nominals N.

Inheritated indexicality

Formulas evaluated both at a state $w \in S$ and at an agent $a \in A$.

- p : "I am blonde."
- BFp: "I believe that all my friends are blonde."
- FBp: "All of my friends believe that they are blonde".

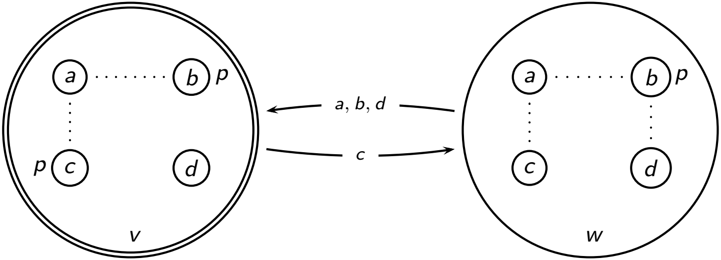
Semantic clauses

- $\mathcal{M}, w, a \vDash p$ iff $\langle w, a \rangle \in \|p\|$
- $\mathcal{M}, w, a \vDash n$ iff $\langle w, a \rangle \in \|n\|$ iff $a = \underline{n}$
- $\mathcal{M}, w, a \vDash \neg\phi$ iff $\mathcal{M}, w, a \nvDash \phi$
- $\mathcal{M}, w, a \vDash \phi \wedge \psi$ iff $\mathcal{M}, w, a \vDash \phi$ and $M, w, a \vDash \psi$
- $\mathcal{M}, w, a \vDash F\phi$ iff $\mathcal{M}, w, b \vDash \phi$ for all b such that $a \asymp b$
- $\mathcal{M}, w, a \vDash @b\,\phi$ iff $\mathcal{M}, w, \underline{b} \vDash \phi$
- $\mathcal{M}, w, a \vDash B\phi$ iff $\mathcal{M}, v, a \vDash \phi$ for all $v \in S$ such that $v \in best_a w(a)$

notation:

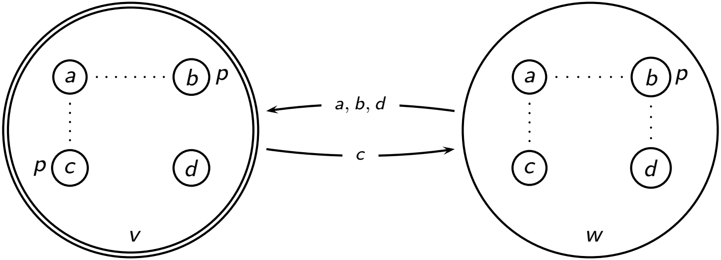
- $\underline{n}$ the unique agent at which the nominal n holds
- $s(a)$ the comparability class of state s relative to agent a: $t \in s(a)$ iff $s \leq_a t$ or $t \leq_a s$
- $best_a s(a)$ the most plausible states in $s(a)$ according to a: $best_a s(a) := \{s \in s(a) : t \leq_a s$ for all $t \in s(a)\}$

Example



- $M, v, \underline{c} \vDash p$
- $M, v, \underline{a} \vDash Fp$
- $M, v, \underline{a} \vDash \langle F \rangle b$
- $M, w, \underline{d} \vDash FBp$
- $M, w, \underline{a} \vDash BFp$

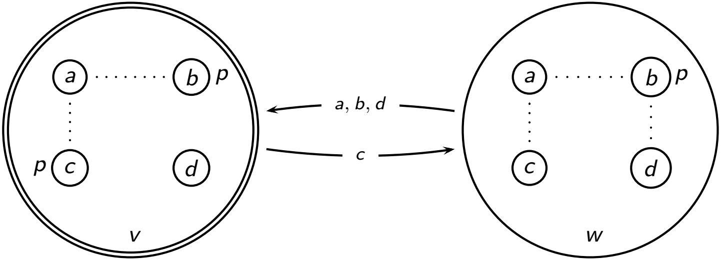
Example



- $M, v, \underline{c} \vDash p$
- $M, v, \underline{a} \vDash Fp$
- $M, v, \underline{a} \vDash \langle F \rangle b$
- $M, w, \underline{d} \vDash FBp$
- $M, w, \underline{a} \vDash BFp$
- $M, w, \underline{c} \vDash B@b\langle F \rangle d$

Example

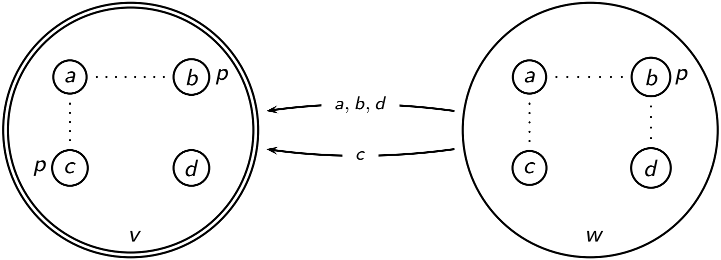
$\neg @b\langle F\rangle d$



- $M, v, \underline{c} \vDash p$
- $M, v, \underline{a} \vDash Fp$
- $M, v, \underline{a} \vDash \langle F\rangle b$
- $M, w, \underline{d} \vDash FBp$
- $M, w, \underline{a} \vDash BFp$
- $M, w, \underline{c} \vDash B@b\langle F\rangle d$

Example

$\neg @b\langle F\rangle d$ 



- $M, v, \underline{c} \models p$
- $M, v, \underline{a} \models Fp$
- $M, v, \underline{a} \models \langle F\rangle b$

- $M, w, \underline{d} \models FBp$
- $M, w, \underline{a} \models BFp$
- $M, w, \underline{c} \models B\neg @b\langle F\rangle d$

Influence dynamics

Simplifying assumptions

- agents speak in turn (rank)
- only friends communicate
- agents revise with (all) sentences announced (trust)

Revision operator

Joint radical upgrade $\Uparrow \phi$

- "Promote" all the $||\phi||$-worlds so that they become more plausible than all $\neg||\phi||$-worlds (in the same information cell), keeping everything else the same:

Revision operator

Joint radical upgrade $\Uparrow \phi$

- "Promote" all the $\|\phi\|$-worlds so that they become more plausible than all $\neg\|\phi\|$-worlds (in the same information cell), keeping everything else the same:
- $\Uparrow \phi$ is a model transformer which takes as input any model $\mathcal{M}= (S, \mathcal{A}, \leq_{a\in\mathcal{A}}, \|\cdot\|, s_0, \asymp_{s\in S})$ and outputs a new model $\mathcal{M}'=(S, \mathcal{A}, \leq'_{a\in\mathcal{A}}, \|\cdot\|, s_0, \asymp_{s\in S})$ such that:

  $s \leq'_a t$ iff either $(s, t \notin \|\phi\|$ and $s \leq_a t)$ or $(s, t \in \|\phi\|$ and $s \leq_a t)$ or $(t \in s(a)$ and $s \notin \|\phi\|$ and $t \in \|\phi\|)$.

Belief merge

Baltag & Smets' lexicographic belief merge protocol

$$\rho_a := \prod \{\Uparrow \phi : \|\phi\| \subseteq S \text{ such that } \mathcal{M}, w \models B_a\phi\}$$

$$\rho_b := \prod \{\Uparrow \phi : \|\phi\| \subseteq S \text{ such that } \mathcal{M}_{[\rho_a]}, w \models B_b\phi\}$$

$$\text{etc for all } c \in \mathcal{A}$$

where $\prod$ is a sequential composition operator and $\mathcal{M}_{[\rho_a]}$ is the new model after joint revision with each formula announced by a.

Belief merge

Indexical lexicographic belief merge protocol

$$\rho_a := \prod\{\Uparrow @_a\phi : \|\phi\| \subseteq S \times \mathcal{A} \text{ such that } \mathcal{M}, w, a \models B\phi\}$$

$$\rho_b := \prod\{\Uparrow @_b\phi : \|\phi\| \subseteq S \times \mathcal{A} \text{ such that } \mathcal{M}_{[\rho_a]}, w, b \models B\phi\}$$

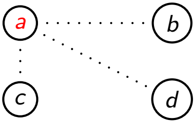
$$\text{etc for all } c \in \mathcal{A}$$

where $\prod$ is a sequential composition operator and $\mathcal{M}_{[\rho_a]}$ is the new model after joint revision with each formula announced by a.

A central friend

Assumptions

- a is other agents' only friend.
- a speaks first.

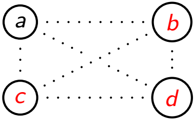


One-to-others unilateral strong influence protocol

One step version of the indexical lexicographic belief merge protocol:

$$\rho_a := \prod\{\Uparrow @_a\phi : \|\phi\| \subseteq S \times \mathcal{A} \text{ such that } \mathcal{M}, w, \underline{a} \models B\phi\}$$

Everybody is friends with everybody else



Assumption

- Connectedness

Others-to-one unilateral strong influence protocol

$$\rho_b := \prod\{\Uparrow @_b B\phi : \|\phi\| \subseteq S \times \mathcal{A} \text{ such that } \mathcal{M}, w, \underline{b} \models B\phi\}$$

$$\rho_c := \prod\{\Uparrow @_c B\phi : \|\phi\| \subseteq S \times \mathcal{A} \text{ such that } \mathcal{M}, w, \underline{c} \models B\phi\}$$

$$\text{etc, for all } d \in \mathcal{A} \text{ such that } \mathcal{M}, w, d \models \langle F \rangle a$$

$$\rho_a := \prod\{\Uparrow @_a \phi \text{ iff } \mathcal{M}_{[\rho_b;\rho_c,\ldots]}, w, \underline{a} \models BFB\phi\}$$

where $\mathcal{M}_{[\rho_b;\rho_c,\ldots]}$ is the model resulting from the successive revisions (by all friends) with each of the formulas announced by each of them.

Summary

- Social network plausibility framework with communication events
- Indexical protocol to merge beliefs
- Unilateral strong influence *one-to-all-the-others* protocol
- Unilateral strong influence *all-the-others-to-one* protocol

To do next

- Private (and synchronic?) communication: *friends to friends* influence (level of privacy to determine)
- Different doxastic attitudes (conditional belief, strong belief, safe belief) + different levels of trust (dynamic attitudes) corresponding to different types of revision (minimal revision, update).
- Consider how to merge (as quickly as possible) knowledge and/or belief within a social network.

Thank you 

zoe.christoff@gmail.com

References

Baltag, A. and Smets, S. (2009).
Protocols for belief merge: Reaching agreement via communication.
volume 494 of *CEUR Workshop Proceedings*, pages 129–141.

Baltag, A. and Smets, S. (2013).
Protocols for belief merge: Reaching agreement via communication.
Logic Journal of the IGPL, 21(3):468–487.

Liu, F., Seligman, J., and Girard, P.
Logical dynamics of belief change in the community.
Synthese.
Special Issue on Social Epistemology, C. Proietti and F. Zenker, editors, to appear.

Seligman, J., Liu, F., and Girard, P. (2011).
Logic in the community.
In Banerjee, M. and Seth, A., editors, *Logic and Its Applications*, volume 6521 of *Lecture Notes in Computer Science*, pages 178–188. Springer Berlin Heidelberg.