Logical dynamics of belief change in the community

Patrick Girard

Department of Philosophy, University of Auckland, New Zealand

Joint work with Fenrong Liu and Jeremy Seligman
December 14, 2012

Outline

The Intersection

The Intersection

With a local guide; an expert.

Belief revision.

The Intersection

With a tour guide; a leader.

Belief aggregation with deference.

The Intersection

With a friend; just a dude.

Belief aggregation with majority.

Doxastic influence

Consider influence regarding a single proposition p. If I do not believe p and most of my friends believe p, how should I respond?

1. Ignore their opinions and remain doxastically unperturbed.
2. Change my beliefs under peef influence:
   - **Revise** my beliefs with p:

$$[Rp]Bp$$

   - **Contract** my beliefs with $\neg p$:

$$[C\neg p]\neg B\neg p$$

Assume belief change is successful.

Two kinds of doxastic influence

1. Strong influence that leads to revision: Sp.
2. Weak influence that leads to contraction: Wp.

Social influence operator regarding p (Ip):

$$\texttt{if } \mathsf{S}p \texttt{ then } Rp \texttt{ else if } \mathsf{W}p \texttt{ then } C\neg p\texttt{;}$$
$$\texttt{if } \mathsf{S}\neg p \texttt{ then } R\neg p \texttt{ else if } \mathsf{W}\neg p \texttt{ then } Cp$$

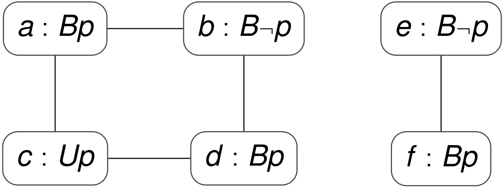
Distribution of doxastic states

Three possible doxastic states: belief (Bp), disbelief ($B\neg p$) and no belief (Up) defined as $Up = (\neg Bp \wedge \neg B\neg p)$.

Framework: 'logic in the community,' ([Seligman et al., 2011]):

- Friendship: symmetric and irreflexive relation.
- Social network: A set of agents related by friendship.
- Community: A subset of agents that are connected by friendship.

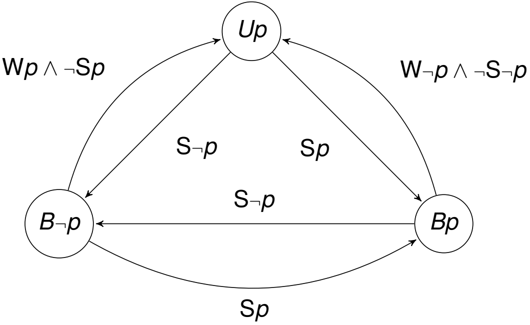
Distribution of doxastic states: Example



A network of six agents, two communities. Links between nodes indicate friendship.

- FBp: 'all my friends believe p'
- $\langle F \rangle Bp$: 'some of my friends believe p'.

Finite state of automaton



Threshhold influence

- If x% of my friends believe p, then I am strongly influenced to believe p.
  - Assume $x = 100$: I am **strongly influenced** to believe p iff **ALL** (at least one) of my friends believe p .
- If at least one of my friends believes p and x% of my friends disbelieve $\neg p$ (i.e., Bp or Ip), then I am weakly influenced to contract my belief in $\neg p$.
  - Assume $x = 0$: I am **weakly influenced** to contract my belief in $\neg p$ iff **NONE** of my friends believe $\neg p$.

Strong and weak influence captured with the following axioms:

$$S\varphi \leftrightarrow (FB\varphi \wedge \langle F \rangle B\varphi)$$
$$W\varphi \leftrightarrow (F\neg B\neg\varphi \wedge \langle F \rangle B\varphi)$$

Threshhold influence: Example

Example 1:

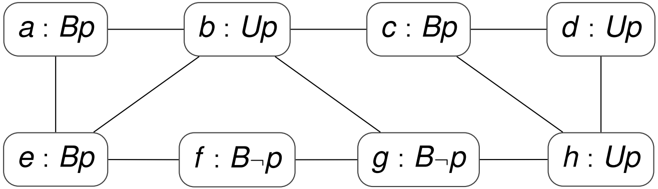
$$\begin{tikzcd}
a : Bp \arrow[r, dash] \arrow[d, dash] & b : B\neg p \arrow[d, dash] & & a : Up \arrow[r, dash] \arrow[d, dash] & b : Bp \arrow[d, dash] \\
c : Up \arrow[r, dash] & d : Bp \arrow[r, phantom, "\overset{Ip}{\rightsquigarrow}"] & {} & c : Bp \arrow[r, dash] & d : Up
\end{tikzcd}$$

Stability and flux

A community is *stable* if the operator *Ip* has no effect on the doxastic states of any agent in the community.

Unanimity within the community is sufficient for stability but not necessary:

Example 2:



Stability and flux: one more example

Configurations that never become stable will be said to be *in flux*:

Example 3:

$$\boxed{\begin{array}{c} a : Bp \\ | \\ b : B\neg p \end{array}} \;\overset{Ip}{\rightsquigarrow}\; \boxed{\begin{array}{c} a : B\neg p \\ | \\ b : Bp \end{array}} \;\overset{Ip}{\rightsquigarrow}\; \boxed{\begin{array}{c} a : Bp \\ | \\ b : B\neg p \end{array}} \;\overset{Ip}{\rightsquigarrow}\; \ldots$$

Characterizing stability

$$\neg(B\neg p \wedge \mathrm{W}p) \ \wedge\ \neg(Up \wedge \mathrm{S}p) \ \wedge\ \neg(Up \wedge \mathrm{S}\neg p) \ \wedge\ \neg(Bp \wedge \mathrm{W}\neg p)$$

Under the assumption of threshold influence, it is equivalent to

$$\begin{aligned}&\neg(B\neg p \wedge F\neg B\neg p \wedge \langle F\rangle Bp) \wedge\\ &\neg(\neg Bp \wedge \neg B\neg p \wedge FBp \wedge \langle F\rangle Bp) \wedge\\ &\neg((\neg Bp \wedge \neg B\neg p \wedge FB\neg p \wedge \langle F\rangle B\neg p) \wedge\\ &\neg(Bp \wedge F\neg Bp \wedge \langle F\rangle B\neg p)\end{aligned}$$

A community is *stable* when every agent in the community satisfies this condition.

Characterizing flux

[Zhen and Seligman, 2011] contains a theorem that shows how to characterizing flux for the preference dynamics.

Conjecture: A community (of at least two agents) is in flux iff every agent in the community satisfies the condition

$$(FBp \wedge FFB\neg p) \;\vee\; (FB\neg p \wedge FFBp)$$

In particular, if there is any agent in the community in state *Up*, then the community will eventually become stable – if not stable already.

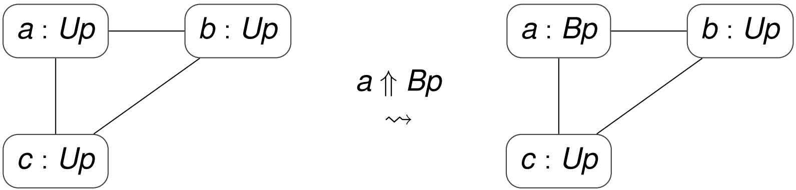
Dynamics I: Private belief change in a community

Agents may change their minds for many reasons other than the influence of their friends. This raises the question of if and how such changes are propagated to other members of the community.

- A very coherent community may resist all changes, ensuring that any agent who changes their mind unilaterally will soon be brought back into conformity.
- A less coherent community may be highly affected by the change, going into flux or even following the agents who changed their mind into a new stable configuration.

Isolated private belief change

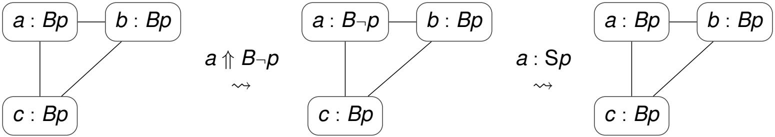
Example 4:



More active resistence

Unanimous belief within a community can be strong enough to resist private belief changes even further:

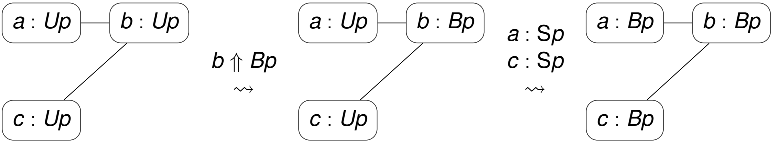
Example 5:



location is critical

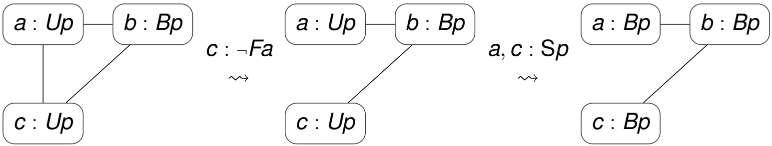
For a community of undecided agents to be influenced by a private belief change, the location of the agent who comes to believe p is critical.

Example 6:



Dynamics II: Gaining and losing friends

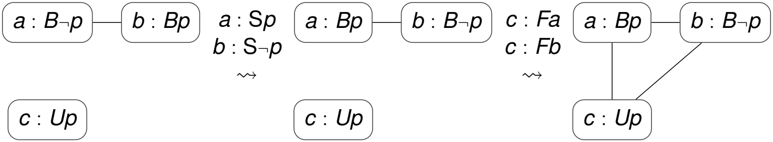
Example 7:



We start with a stable distribution of opinions among three mutual friends, with only one believer. One friendship is broken, putting the mutual friend into a position of greater influence.

New friend joining in

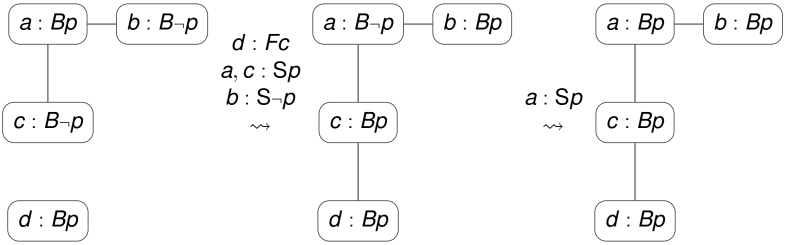
Example 8:



The oscillating pair of friends at the top is calmed when an indifferent agent joins their circle. In one more step, they will all become undecided.

A different newcomer

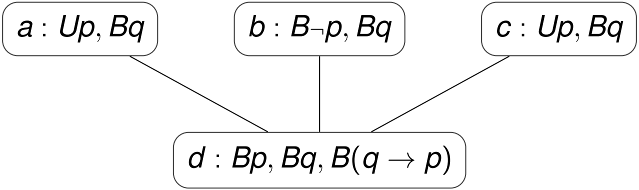
Example 9:



Plausibility influence

We have ignored the interdependence of a person's beliefs:

Example 10:



How is *d* to change her beliefs?

Four plausibility states

Given a fixed domain W of possible outcomes, consider each agent's judgements regarding the relative plausibility of elements of W. For u and v in W, we write $u \leq_a v$ to mean that a judges v to be at least as plausible as u.
There are now four relevant possible states:

- agent a may find v strictly more plausible than u ($u \leq_a v$ and $v \not\leq_a u$)
- vice versa,
- may regard them as equally plausible ($u \leq_a v$ and $v \leq_a u$)
- have no view at all ($u \not\leq_a v$ and $v \not\leq_a u$).

Plausibility influence: the simplest case

If they all take v to be at least as plausible as u, so does she, and if they all take v not to be at least as plausible as u, nor does she. We will call this *plausibility influence*($\mathcal{I}$). The action of plausibility influence is characterised by:

$$
\begin{tikzcd}
u \leq v \arrow[rr, bend left, "F(u \not\leq v)"] & & u \not\leq v \arrow[ll, bend left, "F(u \leq v)"]
\end{tikzcd}
$$

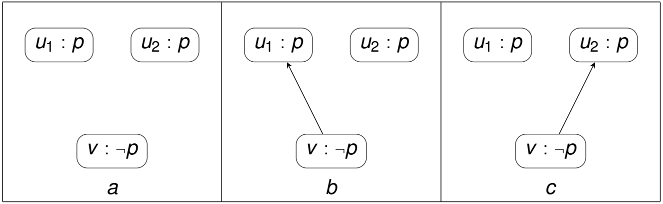
Two more automata

Label R (for 'right') v is strictly more plausible than u, L (for 'left') u is strictly more plausible than v, I (for 'impartial') u and v are equally plausible, and O (for 'no opinion').
In these terms $u \leq v$ is $(I \vee R)$, $u \not\leq v$ is $(O \vee L)$, $v \leq u$ is $(I \vee L)$, and $v \not\leq u$ is $(O \vee R)$. The dynamics of the two parts of the comparison can be represented as the two automata:

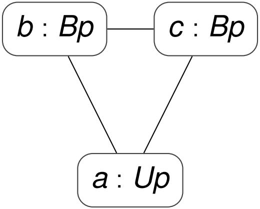
$$\begin{tikzcd} I \vee R \arrow[r, bend left, "F(O \vee L)"] & O \vee L \arrow[l, bend left, "F(I \vee R)"] \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} I \vee L \arrow[r, bend left, "F(O \vee R)"] & O \vee R \arrow[l, bend left, "F(I \vee L)"] \end{tikzcd}$$

Example

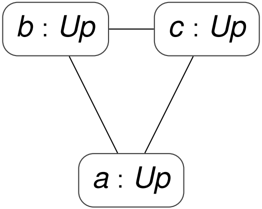
There are three possible outcomes, u_1, u_2 and v with p true at each of u_1 and u_2 but not at v. agent b regards u_1 as strictly more plausible than v, agent c regards u_2 as strictly more plausible than v. agent a makes no judgements at all.



Agents b and c agree that u_1 and u_2 are the only maximally plausible outcomes, and so believe that p. Their friend a also allows that outcome v is maximally plausible and so is undecided about p. Thus we have the following configuration:



After plausibility influence, agent b drops her judgement that u_1 is more plausible than v because it is not supported by either friend. Agent c drops her judgement that u_2 is more plausible than v, making all three agents converge to a's initial view.



We can interpret this as capturing, to some extent, the influence of reasons rather than mere beliefs.

Other interesting issues

- Ranked friends
- Reliability
- Belief aggregation
- A right dynamic logic language

Thank you for your attention!

Email: p.girard@auckland.ac.nz

Seligman, J., Liu, F., and Girard, P. (2011).
Logic in the community.
In Banerjee, M. and Seth, A., editors, *Logic and Its Applications - 4th Indian Conference, ICLA 2011, Delhi, India, January 5-11, 2011. Proceedings*, volume 6521 of *Lecture Notes in Computer Science*, pages 178–188.

Zhen, L. and Seligman, J. (2011).
A logical model of the dynamics of peer pressure.
Electronic Notes in Theoretical Computer Science, 278:275–288.