Statistical biases on perception of self and others: From model to experiments

Guillaume Deffuant Thibaut Roubin

Inrae - Laboratoire d'Ingéniérie des Systèmes Complexes

October 2021

Positive bias on self-opinion because of self-enhancement

- Most of people tend to seek out and accept positive feedbacks about themselves and avoid or reject negative ones (Campbell et al., 1999).
- As a result, we tend to over-evaluate ourselves as attested by a lot of experiments (Dunning et al., 2004).

Positive bias on self-opinion without self-enhancement

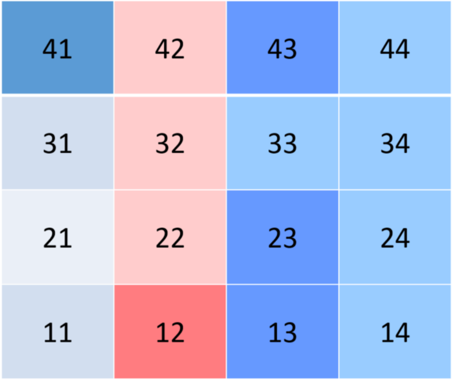
- A recent model of agents with opinions about each others suggests the existence of a positive bias on self-opinions without self-enhancement (Deffuant et al. 2018).
- The analysis shows the existence of a negative bias on the opinion about others.
- These biases have not been detected by social scientists.

Main points of this presentation

- A new analysis of the model, showing that the biases hit the agents differently according to their status to the detriment of low status agents?
- A first step of experimental work confirming the existence of the specific positive bias on self-opinion.

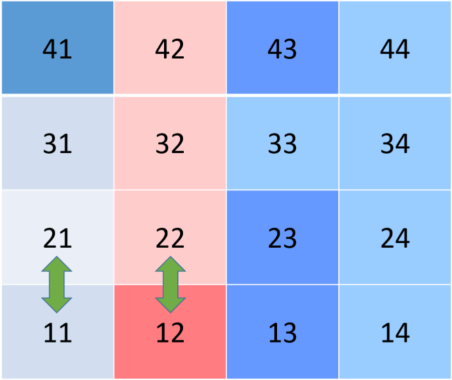
Simple model focusing on effect of interactions

- Each agent is defined by a self-opinion and an opinion about all the other agents



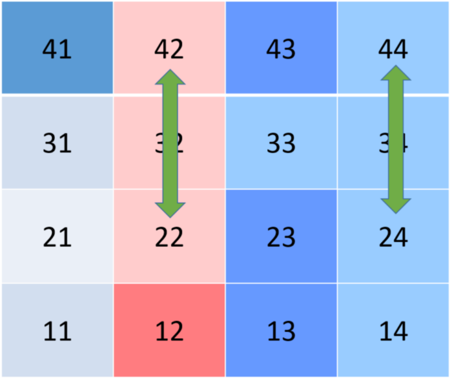
Pair interaction

- Agents 1 and 2 are chosen.
- Agent 2 opinions a_{22} and a_{21} influence agent 1 opinions a_{12} and a_{11}. And vice-versa.
- Agent 1 influences agent 2 similarly.



Pair interaction

- Agents 2 and 4 are chosen.
- Agent 4 opinions a_{44} and a_{42} influence agent 1 opinions a_{24} and a_{22}.
- Agent 2 influences agent 4 similarly.

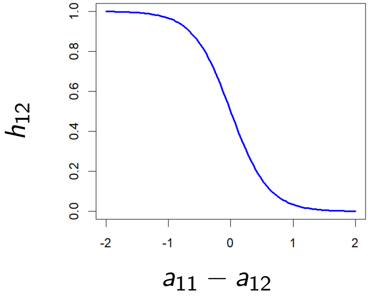


Noisy attraction

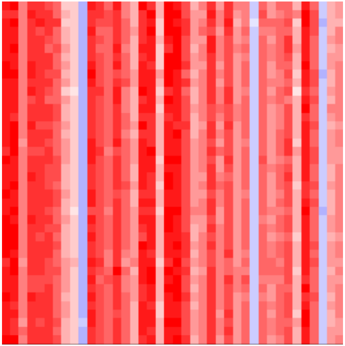
- Influence of a_{21} on a_{11}:

$$a_{11}(t+1) = a_{11}(t) + h_{12}(t)(a_{21}(t) + \theta - a_{11}(t))$$

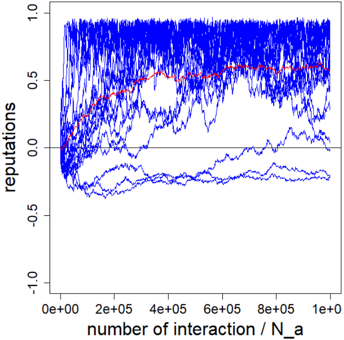
- θ is a uniform noise in $[-\delta, \delta]$
- $h_{12}(t) = \frac{1}{1+\exp\left(\frac{a_{11}(t)-a_{12}(t)}{\sigma}\right)}$ is the influence of 2 on 1.



Positive drift of opinions without gossip



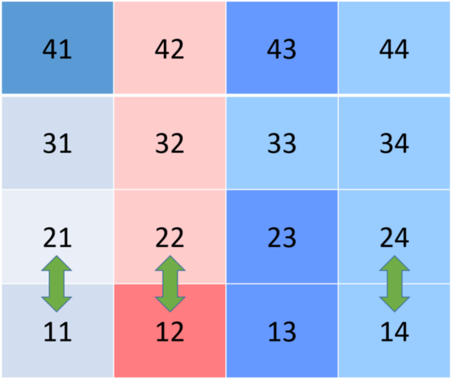
Matrix after 1 million iterations (starting with all opinions equal 0)



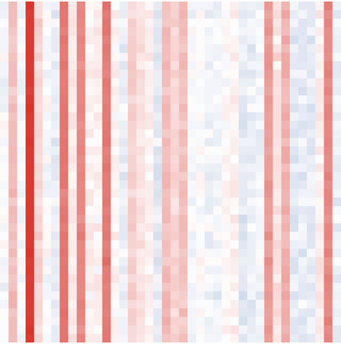
Red curve: average opinion. Blue curves: reputations (average columns).

Pair interaction with gossip

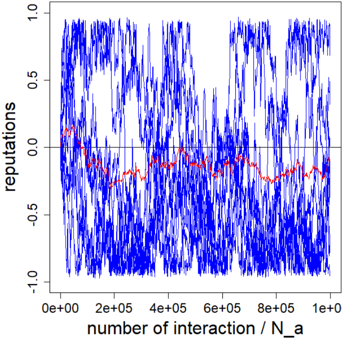
- If gossip is activated, agent 2's opinion a_{24} about randomly chosen agent 4 influences agent 1's opinion a_{14}.
- Agent 1 influences agent 2 similarly.



Negative drift of opinions with gossip



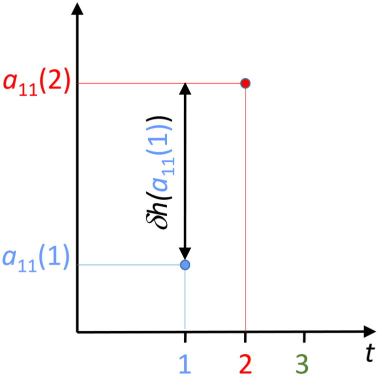
Matrix after 1 million iterations (starting with all opinions equal 0)



Red curve: average opinion.
Blue curves: reputations.

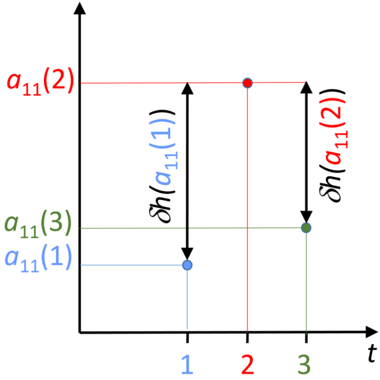
Positive bias on self-opinion: simplified case

- Assume $a_{21}(1) = a_{11}(1) + \delta$, then:
  $a_{11}(2) = a_{11}(1) + \delta h(a_{11}(1))$



Positive bias on self-opinion: simplified case

- Assume $a_{21}(2) = a_{11}(2) - \delta$,
- then:
  $a_{11}(3) = a_{11}(2) - \delta h(a_{11}(2))$
- Moreover: $a_{11}(3) - a_{11}(1) = \delta(h(a_{11}(1)) - h(a_{11}(2))$
- $h(a_{11}(1)) > h(a_{11}(2))$ as h is decreasing.



Generalisations

- The effect is the same with the sequence $a_{21}(1) = a_{11}(1) - \delta$, $a_{21}(2) = a_{11}(2) + \delta$
- The same mechanism is at work on the opinion about others except that h is growing when the opinion about the other increases, therefore the bias is negative.
- This analysis is not elaborate enough to determines which bias dominates during interactions and to explain the patterns.

Approximation of average opinions: Principle

- We consider opinion offsets: $x_{ij}(t) = a_{ij}(t) - a_{ij}(0)$
- We average the equations of opinion change such as:

$$x_{ii}(t+1) = x_{ii}(t) + h_{ij}(t)(x_{ji}(t) - x_{ii}(t) + \theta(t)),$$

  over the noise on the influence and over the randomness of the interacting pairs.
- We develop the influence function around its average:

$$h_{ij}(t) = \overline{h_{ij}}(t) + \overline{h'_{ij}}(t)(x_{ii}(t) - x_{ij}(t) - \overline{x_{ii}}(t) + \overline{x_{ji}}(t))$$

Evolution of average opinions without gossip

- Evolution of average self-opinion of i:

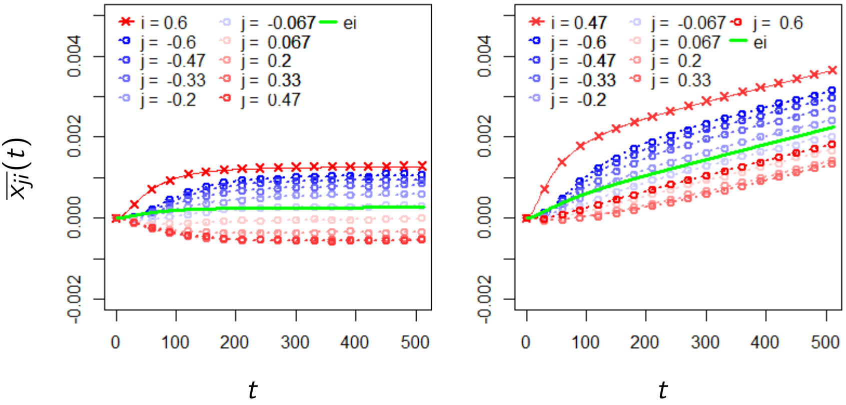
$$\overline{x_{ii}}(t+1) = \overline{x_{ii}}(t) + \frac{2}{N_c} \sum_{j \neq i} \Big(\widehat{h_{ij}}(t) \left(\overline{x_{ji}}(t) - \overline{x_{ii}}(t) \right) + \overline{h'_{ij}}(t) \left(\overline{x_{ii}(t).x_{ji}(t)} - \overline{x^2_{ii}}(t) \right) \Big)$$

- Evolution of average opinion of j about i:

$$\overline{x_{ji}}(t+1) = \overline{x_{ji}}(t) + \frac{2}{N_c} \Big(\widehat{h_{ji}}(t) \left(\overline{x_{ii}}(t) - \overline{x_{ji}}(t) \right) + \overline{h'_{ji}}(t) \left(\overline{x^2_{ji}}(t) - \overline{x_{ii}(t).x_{ji}(t)} \right) \Big).$$

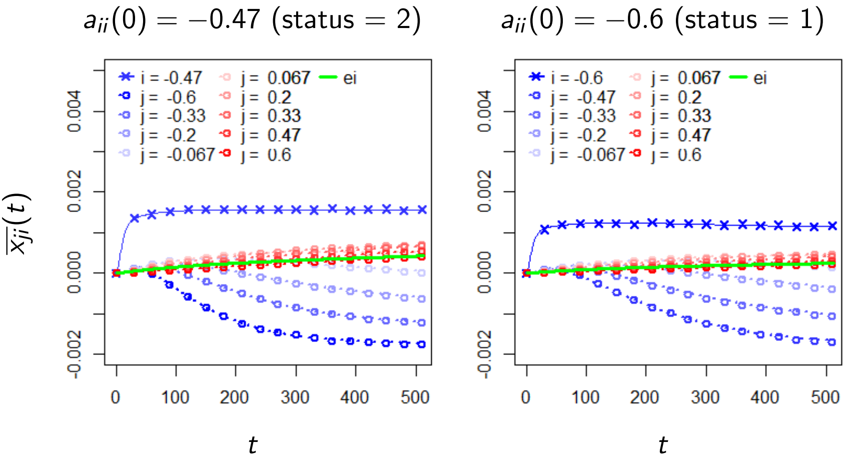
Example: 10 agents with initial opinions in [-0.6, 0.6]

$a_{ii}(0) = 0.6$ (status = 10) $a_{ii}(0) = 0.47$ (status = 9)



Lines: moment approximation, points: average of 10 M simulations

Example: 10 agents with initial opinions in [-0.6, 0.6]



Lines: moment approximation, points: average of 10 M simulations

First order equilibrium opinion

- The equilibrium opinion $e_i(t)$ is weighted average of the opinions about i:

$$e_i(t) = \frac{1}{1 + S_i(t)} \left(\overline{x_{ii}}(t) + \sum_{j \neq i} \frac{\widehat{h_{ij}}(t)}{\widehat{h_{ji}}(t)} \overline{x_{ji}}(t) \right),$$

  with $S_i(t) = \sum_{j \neq i} \frac{\widehat{h_{ij}}(t)}{\widehat{h_{ji}}(t)}$.

- The trajectory of the equilibrium opinion $e_i(t)$ reflects the trajectories of the opinions about i.

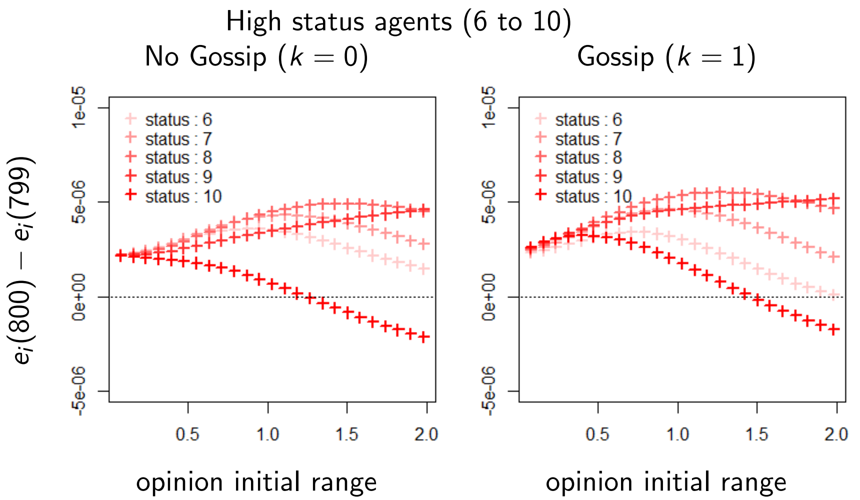
Evolution

- The evolution of $e_i(t)$ includes only second order terms:

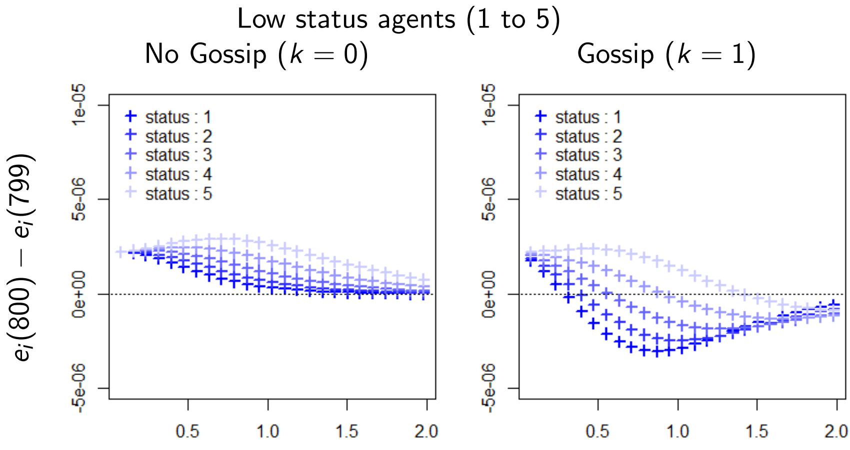
$$e_i(t+1) = e_i(t) + \frac{2}{N_c(1+S_i(t))} \sum_{i \neq j} \overline{h'_{ji}}(t) \left(\overline{x_{ii}(t).x_{ji}(t)} - \overline{x^2_{ii}}(t) + \frac{\widehat{h_{ij}(t)}}{\widehat{h_{ji}(t)}} \left(\overline{x^2_{ji}}(t) - \overline{x_{ii}(t).x_{ji}(t)} \right) \right).$$

- With or without gossip, the weight of the negative bias $\frac{\widehat{h_{ij}}(t)}{\widehat{h_{ji}}(t)}$ is larger when i is of low status;
- With gossip, the negative bias $\overline{h'_{ji}}(t) \left(\overline{x^2_{ji}}(t) - \overline{x_{ii}(t).x_{ji}(t)} \right)$ is larger when i is of low status;

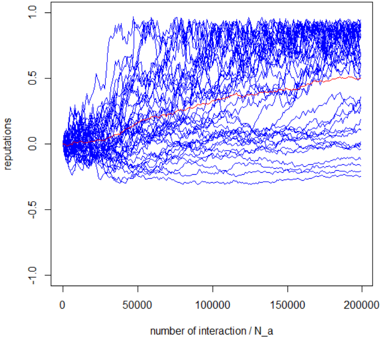
Slope of $e_i(t)$ at $t = 800$ when inequalities vary



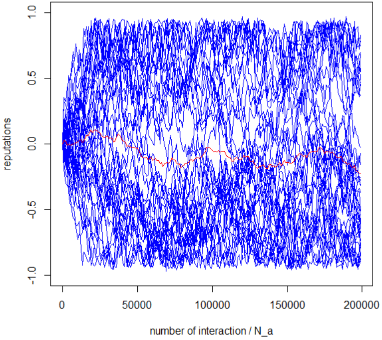
Slope of $e_i(t)$ at $t = 800$ when inequalities vary



Explanation of the patterns



no gossip ($k = 0$)



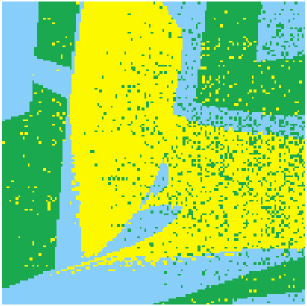
Gossip ($k = 5$).

The experiment

- 1500 participants recruited by a specialised firm
- Online questionnaire:
    - We request the participant to complete a specific task
    - The participants receive a series of 4 evaluations about their performance (2 are $a_{ii} + \delta$ (positive evaluations) and 2 are $a_{ii} - \delta$ (negative evaluations)
    - After each evaluation, the participants express their self-evaluation.

The task

- 3 pictures are displayed for 5 seconds, and for each, we ask "What percentage of green do you see in the image ?" to the participant



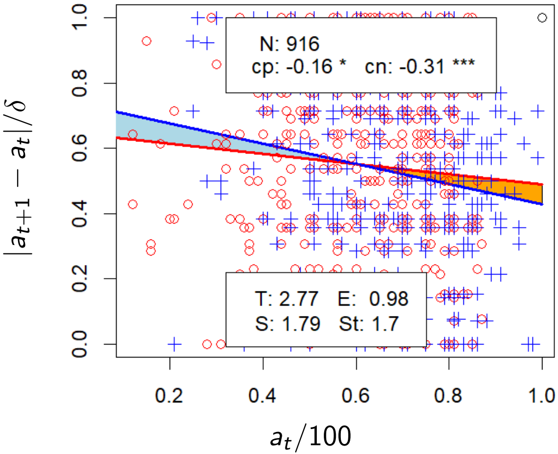
- The task is : unusual hence the participants have no idea of their likely performance at it, so they can believe fake evaluations.

Anchor and first self-evaluation

- For each participants, we collect 4 triples: (a_t, f_t, a_{t+1}), with $f_t = a_t \pm \delta$.
- By hypothesis $a_{t+1} - a_t = \pm h(a_t)\delta$
- We perform two regressions $a_{t+1} - a_t$ by a_t:
  - for $f_t = a_t + \delta$, providing an approximation of $h_+(a_t)$
  - for $f_t = a_t - \delta$, providing an approximation of $h_-(a_t)$

Results for High trust, high anchor

- The influence is decreasing.
- Bias because from decreasing influence $S = 1.7$ (bigger than the bis from self-enhancement $E = 0.98$).
- As a result, the total bias is 2.77 (percentages of δ)



Conclusion

- Main results:
  - The moment approximation explains how the biases interact and explains the patterns
  - The experiment confirms the existence of the bias on self-opinion from decreasing influence.
- Perspectives:
  - Extending the model to larger populations and introducing other processes (vanity, group identity) and networks.
  - Performing lab experiments checking the existence of the negative bias on opinions about others.