Do interactions among unequal agents undermine those of low status?

Guillaume Deffuant, Thibaut Roubin

To cite this version:

Guillaume Deffuant, Thibaut Roubin. Do interactions among unequal agents undermine those of low status?. Physica A: Statistical Mechanics and its Applications, 2022, 592, pp.126780. 10.1016/j.physa.2021.126780 . hal-04072464

HAL Id: hal-04072464
https://hal.inrae.fr/hal-04072464v1

Submitted on 18 Apr 2023

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Do interactions among unequal agents undermine those of low status?

Guillaume Deffuant [a,b,*], Thibaut Roubin [a]

[a] Université Clermont Auvergne, INRAE, UR LISC, Aubière, France
[b] Université Clermont Auvergne, LAPSCO, Clermont-Ferrand, France

ARTICLE INFO

Article history:
Received 8 September 2021
Received in revised form 15 November 2021
Available online 20 December 2021

Keywords:
Opinion dynamics
Positive bias
Negative bias
Gossip
Moment approximation

ABSTRACT

We consider a recent model in which agents hold opinions about each other and influence each other's opinions during random pair interactions. When the opinions are initially close, on the short term, all the opinions tend to increase over time. On the contrary, when the opinions are initially very unequal, the opinions about agents of high status increase, but the opinions about agents of low status tend to stagnate without gossip and to decrease with gossip. We derive a moment approximation of the average opinion changes that explains these observations.

1. Introduction

Most opinion dynamics models consider opinions[1] about objects, like commercial products, or about actions on the world, like political options [1–6], for a recent review see: [7]. It is generally assumed that when agents discuss about an object or an action, they influence each other's opinions. These interactions can therefore determine commercial or political successes or failures.

In these approaches, apart from a few exceptions [8,9], opinions about the agents themselves are not considered as deserving any specific attention. However, the opinions about agents determine the social network of positive or negative connections, hence in some respect the social structure. Moreover, it is generally recognised that this social structure has a strong influence on the agents' opinions. This suggests that opinions about agents do matter. Several opinion dynamics models include such a structure and in some cases it is evolving. This is for instance the case of some versions of the social impact model [10,11]. Moreover, other researches propose models of social structure dynamics, for instance hierarchies resulting from fights between primates [12]. In both cases, the social structure is generated by processes that are different from opinion dynamics.

This paper precisely focuses on the dynamics of opinions about agents as the source of social structure evolution. It builds on previous research [13,14] on models of agents that hold an opinion (a real number between -1 and $+1$) about all the others and themselves. The model dynamics repeats encounters of two randomly chosen agents influencing their self-opinions and their opinions about each other. Moreover, if gossip is activated, both agents influence their opinions about some other randomly chosen agents. The influence is attractive and the agents are more influenced by the ones that

[*] Corresponding author at: Université Clermont Auvergne, INRAE, UR LISC, Aubière, France.
E-mail address: guillaume.deffuant@inrae.fr (G. Deffuant).

[1] When using "opinions", we conform to the usage in the research community but we think that for our model, "attitudes" would be more appropriate.

https://doi.org/10.1016/j.physa.2021.126780
0378-4371/

they hold in high esteem. Importantly, agents do not have a transparent access to the opinions of others; they constantly make errors of interpretation which are modelled by a random noise.

A strong assumption of the approach is that the self-opinion of agent ego is shaped by its perception of the opinions of others about ego. Therefore, on average, ego's self-opinion measures how ego feels perceived by others. As stressed in [15], this is in line with the hypothesis considering self-esteem as a sociometer [16].

The approach also postulates a pivotal role of self-opinions in the influence function. Moreover, the initial version of the model [14] includes an additional dynamics, called vanity, in which the self-opinion plays a major role. When combining the attractive dynamics and vanity, a significant positive bias on self-opinions emerges: ego's self-opinion is significantly higher than the average of the opinions of others about ego (see details in [14]). Further investigations show that even without vanity, the model generates intriguing patterns and the self-opinions are also slightly higher than the opinions about the agent [13,17].

The main contribution of this paper is an analytical approximation of the average (first moment) evolution of the opinions in the model and the average evolution of their products (second moment). This moment approximation is inspired by a general approach already applied on other agent based models [18]. This moment approximation confirms the hypothesis (formulated in [13]) that the evolution of the opinions about an agent is determined by a combination of positive effects on the agent self-opinion and negative effects on the opinions of others about this agent. Moreover, the moment approximation explains why the opinions about the agents of low status tend to stagnate or to decrease (especially when there is gossip) while the opinions about agents of high status tend to increase. Finally, these results explain the patterns described in [13].

The following section describes the model and presents the patterns in more details. Section 3 is devoted to the moment approximation. Section 4 analyses the accuracy of the approximation and studies the effect of initial opinion inequalities in the case of a group of 10 agents. The last section is devoted to a discussion about the results and their possible connections with some research in social-psychology.

2. Model and patterns

This section first presents the model in details and the main patterns emerging from its dynamics that drew attention in previous research. Then, it recalls their hypothetical explanation, proposed in [13].

2.1. The model

The model is the same as in [13]. We present it here with slightly different notations. It includes N_a agents. Each agent $i \in \{1, \dots, N_a\}$ has an opinion a_{ij} about each agent $j \in \{1, \dots, N_a\}$ including themselves. The opinions are real values between -1 and +1. In [13], at the initialisation, all opinions are set to 0: agents have a neutral opinion about themselves and all the others at the beginning of the simulations. In this paper, we shall also consider specific initial values of the opinions expressing different levels of initial perceived inequalities.

Graphically, the agents' opinions can be represented as a matrix (see examples on Fig. 1) in which row i, with $1 \leq i \leq N_a$, represents the array of N_a opinions of agent i about the agents j. Column j, with $1 \leq j \leq N_a$, represents the opinions all agents i about j. Positive opinions are represented with red shades and negative opinions with blue shades. Lighter shades are used for opinions of weak intensity (close to 0).

The dynamics consists in repeating:

- choose randomly two distinct agents i and j;
- i and j interact: j influences i's opinions and i influences j's opinions.

In this interaction, $a_{ii}(t)$, i's self-opinion, is influenced by $a_{ji}(t)$, the opinion of j about i. As a result of this influence, $a_{ii}(t)$ gets closer to a noisy evaluation of $a_{ji}(t)$. The modification of $a_{ii}(t)$, denoted by $\Delta a_{ii}(t)$, is ruled by the following equation, in which $\theta_{ii}(t)$ designates a number that is uniformly drawn between $-\delta$ and δ (δ being a parameter of the model):

$$\Delta a_{ii}(t) = h_{ij}(t)(a_{ji}(t) - a_{ii}(t) + \theta_{ii}(t)), \tag{1}$$

Similarly, the change of opinion $a_{ji}(t)$, is:

$$\Delta a_{ji}(t) = h_{ji}(t)\left(a_{ii}(t) - a_{ji}(t) + \theta_{ji}(t)\right). \tag{2}$$

where $\theta_{ji}(t)$, is a uniformly drawn number between $-\delta$ and δ. The function of influence $h_{ij}(t)$ is given by Eq. (3), expressing that the more i perceives j as superior, the more j is influential on i.

$$h_{ij}(t) = H(a_{ii}(t) - a_{ij}(t)) = \frac{1}{1 + \exp\left(\frac{a_{ii}(t) - a_{ij}(t)}{\sigma}\right)}. \tag{3}$$

In this model, self-opinions measure how well agents think they are perceived by others, with a stronger weight attributed to agents perceived as superior. As stressed in the introduction, this is in line with the hypothesis considering self-opinion as a sociometer [16].

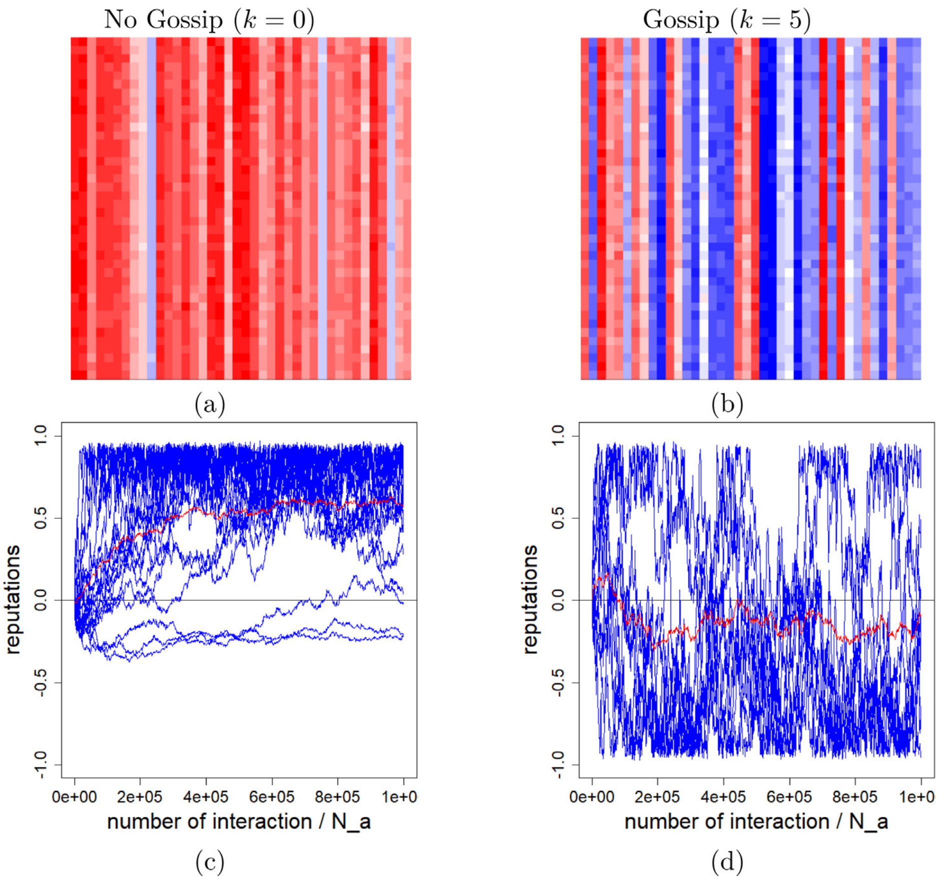

Fig. 1. Typical patterns, with $\delta = 0.1$ (noise), $\sigma = 0.3$ (influence function parameter) and $N_a = 40$ agents. Panels (a) and (b) show the matrix of opinions after 1 million $\times N_a$ pair interactions. Panels (c) and (d) show the evolution of the average opinion (in red) and the evolution of the agent reputations (in blue, the reputation of agent i being the average of the opinions about i).

When activating gossip, agents j and i influence their opinions about k agents g_p, $p \in \{1, \dots, k\}$ drawn at random such that $g_p \neq i$ and $g_p \neq j$. The changes of the opinion of i about agents g_p are:

$$\Delta a_{ig_p}(t) = h_{ij}(t)(a_{jg_p}(t) - a_{ig_p}(t) + \theta_{ig_p}(t)), \text{ for } p \in \{1, \dots, k\}, \tag{4}$$

where $\theta_{ig_p}(t)$ is a uniformly drawn number between $-\delta$ and δ. The changes of the opinion of j about these agents follow the same equations where j and i are inverted.

Overall, after the encounter between i and j, the opinions about i change as follows:

$$a_{ii}(t+1) = a_{ii}(t) + \Delta a_{ii}(t), \tag{5}$$

$$a_{ji}(t+1) = a_{ji}(t) + \Delta a_{ji}(t). \tag{6}$$

The opinions about j change similarly (inverting j and i in the equations). If there is gossip ($k > 0$), k agents g_p are randomly chosen with $p \in \{1, \dots, k\}$, $g_p \neq i$ and $g_p \neq j$, and the opinions about g_p, for $p \in \{1, \dots, k\}$ change as follows:

$$a_{ig_p}(t+1) = a_{ig_p}(t) + \Delta a_{ig_p}(t), \tag{7}$$

$$a_{jg_p}(t+1) = a_{jg_p}(t) + \Delta a_{jg_p}(t). \tag{8}$$

The opinions are updated synchronously: at each encounter all the changes of opinions (e.g. Eqs. (1), (2) and (4)) are first computed and then the opinions are modified simultaneously (e.g. Eq. (5), (7)).

Overall, the following parameters tune the dynamics:

- σ defines the shape of the influence function h_{ij}; if σ is very small, the function is very tilted, meaning that agents are subject to high influence from the ones that they evaluate better than themselves and they almost completely disregard the opinions of the ones considered lower.

- δ represents the amplitude of the uniformly distributed errors that perturb the evaluation of others' expressed opinions. This noise stands for the inability of an agent to directly access the opinion of others. Without it, from all opinions at zero, there would be no opinion change at all.
- k is the number of agents subject of gossip in each pair interaction (hence if $k = 0$, there is no gossip).

2.2. The main patterns

Fig. 1 illustrates the main patterns of evolution of the opinions reported in [13].

Panels (a) and (c) illustrate the pattern obtained without gossip ($k = 0$). Panel (a) shows a typical opinion matrix after a large number of iterations. In each matrix column the opinions are close and the differences between the matrix columns are stronger than the differences of opinions within each column. This is explained by the attractive dynamics which tends to align the opinions about a given individual. Note that most of the columns are red, indicating that the opinions about most agents are positive. On panel (c) the red curve shows the evolution of the average opinion. The blue curves are the evolution of the agents' reputations (the average opinion about an agent). Starting from 0, the average opinion increases and then fluctuates around a significantly positive value (close to 0.5). As already noticed in [13], this pattern is surprising because, at a first glance, the equations do not privilege changing opinions upward and the noise is symmetric around 0.

Panels (b) and (d) illustrate the pattern taking place when gossip is activated (in this case, $k = 5$). The matrix of opinions after a large number of interactions (panel b) shows numerous blue columns. On panel (d), the evolution of the average opinion (red curve) remains negative with significant fluctuations while the reputations (blue curves) are more dispersed than without gossip, with a larger density in the low part of the opinion axis.

2.3. Hypothetical explanation of the patterns

In [13], the patterns are related to two biases which are observed in a simplified setting where only one opinion varies, between two interacting agents:

- when the self-opinion of ego varies, it is on average slightly higher than the opinion of alter about ego. There is a positive bias on the self-opinion.
- when the opinion of ego about alter varies, symmetrically, it is on average slightly lower than alter's self-opinion. There is a negative bias on the opinion about others.

In the following, paragraph 3.1 describes this setting in more details and derives mathematical expressions of the biases.

The authors of [13] hypothesise that similar biases are present when all opinions vary and more than two agents interact. Moreover, they suggest that the drift to positive or negative opinions is due to the dominance of one bias on the other:

- Without gossip, the positive bias on self-opinion dominates the negative bias on the opinions about others, which explains why the positive drift arises;
- Gossip increases the noise on the opinions about others, which increases the negative bias on opinion about others, leading to its possible domination over the positive bias on self-opinion.

This hypothesis is indirectly supported by experiments involving several agents but varying only the opinions about one specific agent. These experiments measure the average evolution of the opinions over a large number of simulations and their results are compatible with the hypothesis. However, these explanations remain very general and qualitative.

We now derive a moment approximation of the evolution of the opinions, in order to formally define the biases and to determine precisely their connection to the patterns.

3. Moment approximation

We first derive the moment approximation in the case (already presented in [13]) of only one opinion varying between two interacting agents. Then, we extend the approach to the general case of all varying opinions both with or without gossip. Finally, we introduce the equilibrium opinion that determines the effect of second order shared by all the opinions about an agent.

3.1. Single opinion varying between two interacting agents

We assume that only two agents, 1 and 2, interact and firstly only the self-opinion $a_{11}(t)$ is varying, starting from $a_{11}(0) = a$. The other opinions are fixed: for any value of t, $a_{12}(t) = b$ (the opinion of 1 about 2), $a_{21}(t) = a$ (the opinion of 2 about 1), $a_{22}(t) = b$ (self-opinion of 2). Moreover, for any value of i and j, we define $x_{ij}(t)$ as the opinion offset from $t = 0$:

$$x_{ij}(t) = a_{ij}(t) - a_{ij}(0). \tag{9}$$

At the first step, 1 perceives $a_{21}(1)$ as $a + \theta(1)$, $\theta(1)$ being drawn from the uniform distribution between $-\delta$ and δ. Applying the interaction rule:

$$x_{11}(1) = h_{12}(0)(a + \theta(1) - a) \tag{10}$$

$$= h\theta(1), \tag{11}$$

where $h = h_{12}(0) = H(a - b)$. For any expression y, let $\overline{y}$ be the average of expression y over all possible values of the noise. Then, $\overline{x_{11}}(1)$, the average of $x_{11}(1)$ over all possible draws of $\theta(1)$, is:

$$\overline{x_{11}}(1) = h\frac{\int_{-\delta}^{+\delta} \theta d\theta}{2\delta} = 0. \tag{12}$$

At the second step, applying the interaction rule again gives:

$$x_{11}(2) = x_{11}(1) + h_{12}(1)(a + \theta(2) - a - x_{11}(1)). \tag{13}$$

Assuming that $\theta(1)$ is small, approximating the influence function at the first order gives:

$$h_{12}(1) = H(a + h\theta(1) - b) \tag{14}$$

$$\approx h + h'h\theta(1), \tag{15}$$

where $h' = H'(a - b)$. We get:

$$x_{11}(2) \approx x_{11}(1) + (h + h'h\theta(1))(\theta(2) - h\theta(1)), \tag{16}$$

$$\approx (1 - h)h\theta(1) + h(\theta(2) - h\theta(1)) + h'h\theta(1)\theta(2) - h'h^2\theta^2(1). \tag{17}$$

Since $\overline{x_{11}}(1) = 0$, $\overline{\theta}(2) = 0$ and $\overline{\theta(1)\theta(2)} = 0$, we have:

$$\overline{x_{11}}(2) = -\frac{h'h^2 \int_{-\delta}^{+\delta} \theta^2 d\theta}{2\delta}, \tag{18}$$

$$= -\frac{h'h^2\delta^2}{3}, \tag{19}$$

Formula (19) applies to any function h. Therefore, at the second iteration, the average of $x_{11}(2)$ is positive as soon as function $H(a - b)$ is decreasing when $a - b$ is increasing. It is called the positive bias on self-opinions in [13].

With the choice of H specified by Eq. (3), we have $h' = \frac{-h(1-h)}{\sigma}$, hence:

$$\overline{x_{11}}(2) = \frac{h^3(1 - h)\delta^2}{3\sigma}. \tag{20}$$

For any number of iterations $t \geq 2$ it can be shown that:

$$\overline{x_{11}}(t) = \overline{x_{11}}(2)\left(\frac{1 - (1 - h)^{t-1}}{h} + \frac{(1 - h)^2}{h^2}\left(\frac{1 - (1 - h)^{2(t-2)}}{2 - h} - (1 - h)^{t-2}\left(1 - (1 - h)^{t-2}\right)\right)\right). \tag{21}$$

Therefore, for an infinite number of iterations:

$$\overline{x_{11}}(\infty) = \frac{-h'\delta^2}{3(2 - h)}, \tag{22}$$

$$= \frac{h(1 - h)\delta^2}{3\sigma(2 - h)}. \tag{23}$$

Assuming that: $H(b, a) = 1 - H(a, b) = 1 - h$, like for our choice of H (specified by Eq. (3)), it can easily be seen that $\overline{x_{21}}(2)$ is obtained by replacing h by $1 - h$ in the expression of $\overline{x_{11}}(2)$:

$$\overline{x_{21}}(2) = \frac{h'(1 - h)^2\delta^2}{3}. \tag{24}$$

If $H(a - b)$ is decreasing then h' is negative and $\overline{x_{21}}(2)$ is negative. With H defined by Eq. (3), we have:

$$\overline{x_{21}}(2) \approx -\frac{h(1 - h)^3\delta^2}{3\sigma}. \tag{25}$$

Finally, $\overline{x_{21}}(t)$ and $\overline{x_{21}}(\infty)$ are also obtained by replacing h by $1 - h$ in the expression of $\overline{x_{11}}(t)$ and $\overline{x_{11}}(\infty)$ (Eqs. (21) and (22)) and multiplying them by -1.

3.2. Evolution of average opinions without gossip

Now, we consider a set of N_a agents interacting as specified in Section 2.1. First, we average over the noise in the interactions defined by the sequence of randomly chosen couples $s_t = \{(i_1, j_1), \dots, (i_t, j_t)\}$. Then we average over all possible sequences s_t of randomly chosen couples.

For $(i, j) \in \{1, \dots, N_a\}^2$, let $a_{ij}(s_t)$ be the opinion of agent i about agent j after the sequence s_t of interactions and $x_{ij}(s_t)$ be the opinion offset from $t = 0$:

$$x_{ij}(s_t) = a_{ij}(s_t) - a_{ij}(0). \tag{26}$$

For any variable $y(s_t)$, let $\overline{y}(s_t)$ be the average of $y(s_t)$ over the noise during the interactions defined by the sequence of couples s_t, and let:

$$h_{ij}(s_t) = H(a_{ii}(s_t) - a_{ij}(s_t)); \tag{27}$$

$$\overline{h_{ij}}(s_t) = H(\overline{a_{ii}}(s_t) - \overline{a_{ij}}(s_t)); \tag{28}$$

$$\overline{h'_{ij}}(s_t) = H'(\overline{a_{ii}}(s_t) - \overline{a_{ij}}(s_t)). \tag{29}$$

We approximate $h_{ij}(s_t)$ at the first order around $\overline{h_{ij}}(s_t)$:

$$h_{ij}(s_t) \approx \overline{h_{ij}}(s_t) + \overline{h'_{ij}}(s_t)(x_{ii}(s_t) - x_{ij}(s_t) - \overline{z_{ij}}(s_t)), \tag{30}$$

where :

$$\overline{z_{ij}}(s_t) = \overline{x_{ii}}(s_t) - \overline{x_{ij}}(s_t). \tag{31}$$

For $(i, j) = (i_{t+1}, j_{t+1})$ or $(i, j) = (j_{t+1}, i_{t+1})$, applying the rule of opinion change, we get:

$$x_{ii}(s_{t+1}) = x_{ii}(s_t) + \left(\overline{h_{ij}}(s_t) + \overline{h'_{ij}}(s_t)\left(x_{ii}(s_t) - x_{ij}(s_t) - \overline{z_{ij}}(s_t)\right)\right)\left(x_{ji}(s_t) + \theta_{ii}(t) - x_{ii}(s_t)\right). \tag{32}$$

For sake of simplicity, we assume that all the opinions about any agent i are the same at $t = 0$: $a_{ii}(0) - a_{ji}(0) = 0$ for all couples $(i, j) \in \{1, \dots, N_a\}^2$. Indeed, when $a_{pi}(0) - a_{ji}(0) \neq 0$ the interactions tend rapidly to drive all the opinions about an agent to a very close value, hence this assumption is not restrictive.

Moreover, it can easily be checked that the average product of opinions about two different agents is always zero, hence $\overline{x_{ii}(s_t).x_{jj}(s_t)} = 0$, $\overline{x_{ij}(s_t).x_{ii}(s_t)} = 0$ and $\overline{x_{ij}(s_t).x_{ji}(s_t)} = 0$. Therefore, since $\overline{\theta_{ii}}(t) = 0$, averaging equation (32) and neglecting the terms of order higher than 2 yields:

$$\overline{x_{ii}}(s_{t+1}) = \overline{x_{ii}}(s_t) + \widehat{h_{ij}}(s_t)\left(\overline{x_{ji}}(s_t) - \overline{x_{ii}}(s_t)\right) + \overline{h'_{ij}}(s_t)\left(\overline{x_{ji}(s_t).x_{ii}(s_t)} - \overline{x^2_{ii}}(s_t)\right), \tag{33}$$

with $\widehat{h_{ij}}(s_t) = \overline{h_{ij}}(s_t) - \overline{h'_{ij}}(s_t)\overline{z_{ij}}(s_t)$.

Applying the same approach to $x_{ji}(s_{t+1})$, we get:

$$\overline{x_{ji}}(s_{t+1}) = \overline{x_{ji}}(s_t) + \widehat{h_{ji}}(s_t)\left(\overline{x_{ii}}(s_t) - \overline{x_{ji}}(s_t)\right) + \overline{h'_{ji}}(s_t)\left(\overline{x^2_{ji}}(s_t) - \overline{x_{ji}(s_t).x_{ii}(s_t)}\right). \tag{34}$$

For $(i, j) = (i_{t+1}, j_{t+1})$ or $(j, i) = (i_{t+1}, j_{t+1})$, we can similarly derive the value of the second moment $\overline{x^2_{ii}}(s_{t+1})$. Neglecting the terms of degree higher than 2, we get:

$$\overline{x^2_{ii}}(s_{t+1}) = \overline{x^2_{ii}}(s_t) + 2\widehat{h_{ij}}(s_t)\left(\overline{x_{ii}(s_t)x_{ji}(s_t)} - \overline{x^2_{ii}}(s_t)\right) + \widehat{h_{ij}}(s_t)^2\left(\overline{x^2_{ji}}(s_t) + \overline{x^2_{ii}}(s_t) - 2\overline{x_{ii}(s_t)x_{ji}(s_t)}\right) + \overline{h_{ij}}^2(s_t)\frac{\delta^2}{3}. \tag{35}$$

Similarly, for $\overline{x^2_{ji}}(s_{t+1})$:

$$\overline{x^2_{ji}}(s_{t+1}) = \overline{x^2_{ji}}(s_t) + 2\widehat{h_{ji}}(s_t)\left(\overline{x_{ii}(s_t)x_{ji}(s_t)} - \overline{x^2_{ji}}(s_t)\right) + \widehat{h_{ji}}(s_t)^2\left(\overline{x^2_{ji}}(s_t) + \overline{x^2_{ii}}(s_t) - 2\overline{x_{ii}(s_t)x_{ji}(s_t)}\right) + \overline{h_{ji}}^2(s_t)\frac{\delta^2}{3}. \tag{36}$$

In both cases, the last term of the equation uses the result obtained in Section 3.1, for any interaction noise $\theta(t)$:

$$\overline{\theta^2}(t) = \frac{\delta^2}{3}. \tag{37}$$

Using a similar approach, we compute the expressions of $\overline{x_{ii}(s_{t+1}).x_{ji}(s_{t+1})}$ and $\overline{x_{pi}(s_{t+1}).x_{ji}(s_{t+1})}$, for $(i, j, p) \in \{1, \dots, N_a\}^2$ (see Appendix A.1).

Now, we average the previous equations over all possible sequences of interactions s_t. For any expression $\overline{y}(s_t)$, let $\overline{y}(t)$ be the average of $\overline{y}(s_t)$ over all interaction sequences s_t. Drawing couple (i, j) or couple (j, i) at t has the probability $\frac{2}{N_a(N_a-1)}$, hence averaging equation (33) over all possible sequences s_t yields:

$$\overline{x_{ii}}(t+1) = \overline{x_{ii}}(t) + \frac{2}{N_c}\sum_{j \neq i}\left(\widehat{h_{ij}}(t)\left(\overline{x_{ji}}(t) - \overline{x_{ii}}(t)\right) + \overline{h'_{ij}}(t)\left(\overline{x_{ii}(t).x_{ji}(t)} - \overline{x^2_{ii}}(t)\right)\right), \tag{38}$$

with:

$$N_c = N_a(N_a - 1), \tag{39}$$

$$\widehat{h_{ij}}(t) = \overline{h_{ij}}(t) - \overline{h'_{ij}}(t)\overline{z_{ij}}(t), \tag{40}$$

$$\overline{z_{ij}}(t) = \overline{x_{ii}}(t) - \overline{x_{ij}}(t), \tag{41}$$

$$\overline{h_{ij}}(t) = H(\overline{a_{ii}}(t) - \overline{a_{ij}}(t)), \tag{42}$$

$$\overline{h'_{ij}}(t) = H'(\overline{a_{ii}}(t) - \overline{a_{ij}}(t)). \tag{43}$$

Similarly, averaging equation (34) over all possible sequences s_t, yields:

$$\overline{x_{ji}}(t+1) = \overline{x_{ji}}(t) + \frac{2}{N_c}\left(\widehat{h_{ji}}(t)\left(\overline{x_{ii}}(t) - \overline{x_{ji}}(t)\right) + \overline{h'_{ji}}(t)\left(\overline{x^2_{ji}}(t) - \overline{x_{ii}(t).x_{ji}(t)}\right)\right). \tag{44}$$

Moreover, we derive the equations of the second moments $\overline{x^2_{ii}}(t+1)$, $\overline{x^2_{ij}}(t+1)$, $\overline{x_{ii}(t+1).x_{ji}(t+1)}$ and $\overline{x_{pi}(t+1).x_{ji}(t+1)}$ for $(i,j,p) \in \{1,\dots,N_a\}^2$ (see, Appendix A.2). Then, with the initial values of these variables at $t = 0$, we can compute the values of $\overline{x_{ii}}(t+1)$ and $\overline{x_{ij}}(t+1)$ for $(i,j) \in \{1,\dots,N_a\}^2$ at any time step t by induction.

For $N_a = 2$, we could derive simple direct expressions of $\overline{x_{ii}}(t)$ and $\overline{x_{ji}}(t)$ (not reported in this paper) but for $N_a > 2$ we only get the values by iterating the formulas until reaching t.

3.3. Evolution of the average opinions about an agent when gossip is activated

Now, in the sequence defining the interactions, to each pair (i_t, j_t) we add a set $(g_{1_t},\dots,g_{k_t})$ of k elements of $\{1,\dots,N_a\}$ distinct from i_t and j_t, about which j_t and i_t gossip.

- If $(i,j) = (i_{t+1}, j_{t+1})$ or $(j,i) = (i_{t+1}, j_{t+1})$, the equations of $\overline{x_{ii}}(s_{t+1})$ and $\overline{x_{ji}}(s_{t+1})$ are the same as in the previous paragraph.
- Moreover, for any $g \in \{g_{1_{t+1}},\dots,g_{k_{t+1}}\}$, we have:

$$\overline{x_{ig}}(s_{t+1}) = \overline{x_{ig}}(s_t) + \widehat{h_{ij}}(s_t)\left(\overline{x_{jg}}(s_t) - \overline{x_{ig}}(s_t)\right). \tag{45}$$

The equations specifying $\overline{x^2_{ii}}(s_{t+1})$, $\overline{x^2_{ji}}(s_{t+1})$ and the other second order moments are specified in Appendix A.4.

Now, we derive the expression of the evolution of the opinion offsets averaged over the noise and the sequences of interactions. The expression of $\overline{x_{ii}}(t+1)$ is the same as without gossip (Eq. (38)).

The expression of $\overline{x_{ji}}(t+1)$ includes an additional sum representing the average effect of agents gossipping with j about i.

$$\overline{x_{ji}}(t+1) = \overline{x_{ji}}(t) + \frac{2}{N_c}\left(\widehat{h_{ji}}(t)\left(\overline{x_{ii}}(t) - \overline{x_{ji}}(t)\right) + \overline{h'_{ji}}(t)\left(\overline{x^2_{ji}}(t) - \overline{x_{ii}(t).x_{ji}(t)}\right)\right) + \frac{2k}{N_T}\sum_{p\notin\{i,j\}}\widehat{h_{jp}}(t)\left(\overline{x_{pi}}(t) - \overline{x_{ji}}(t)\right), \tag{46}$$

where $N_T = N_a(N_a-1)(N_a-2)$. The equations of $\overline{x^2_{ii}}(t+1)$, $\overline{x^2_{ij}}(t+1)$, $\overline{x_{ii}(t+1).x_{ji}(t+1)}$, $\overline{x_{ji}(t+1).x_{pi}(t+1)}$ for $(i,j,p) \in \{1,\dots,N_a\}^2$ are specified in Appendix A.4. Again, using the values of the terms at $t = 0$, we can compute the values of $\overline{x_{ii}}(t)$ and $\overline{x_{ij}}(t)$ at any time step t by induction.

The expressions of the positive bias on self-opinion and negative biases on the opinions about i are the same as when there is gossip. Introducing the equilibrium opinion helps to understand how the biases are combined in the interactions and the impact of gossip.

3.4. Interpretation of the equations and first order equilibrium opinion

With or without gossip, at $t = 2$, the expressions of $\overline{x_{ii}}(2)$ and $\overline{x_{ji}}(2)$ are:

$$\overline{x_{ii}}(2) = -\frac{4}{N_c^2}\left(\sum_{j\neq i} h'_{ij}(0)\right)\left(\sum_{j\neq i} h^2_{ij}(0)\right)\frac{\delta^2}{3}, \tag{47}$$

$$\overline{x_{ji}}(2) = \frac{4}{N_c^2}h'_{ij}(0)\left(1 - h_{ij}(0)\right)^2\frac{\delta^2}{3}, \text{ for } j \neq i. \tag{48}$$

Like in the simplified case of only one varying opinion presented in Section 3.1, $\overline{x_{ii}}(2)$ is positive and $\overline{x_{ji}}(2)$ is negative (because $h'_{ij}(0)$ is negative). Therefore, there is also a positive bias on the self-opinions and negative bias on the opinions about other, at step 2, whatever the number of interacting agents. Note that, because of the sums in the expression of $\overline{x_{ii}}(2)$, the positive bias is higher than the negative bias in absolute value, and this difference increases with the number of agents. Finally, both $\overline{x_{ii}}(2)$ and $\overline{x_{ji}}(2)$ are multiplied by $\frac{1}{N_c^2}$ indicating that the effect of the biases in one interaction decreases very strongly (in about $\frac{1}{N_a^4}$) when N_a increases.

More generally, for $t > 2$, let us consider the term of second order in Eq. (38) (same equation with or without gossip):

$$\frac{2}{N_c}\sum_{j\neq i}\overline{h'_{ij}}(t)\left(\overline{x_{ii}(t).x_{ji}(t)}-\overline{x_{ii}^2}(t)\right). \tag{49}$$

This term is positive as the derivative is assumed strictly negative. Therefore, the effect of this term is to increase the self-opinions and it can be seen as a positive bias on self-opinions. Similarly, the term of second order in Eqs. (44) and (46):

$$\frac{2}{N_c}\overline{h'_{ij}}(t)\left(\overline{x_{ji}^2}(t)-\overline{x_{ii}(t).x_{ji}(t)}\right), \tag{50}$$

is negative and tends to decrease the opinion offset x_{ji}. It can be interpreted as a negative bias on opinions about others. Then, consider the terms of first order in Eqs. (38) and (44) respectively:

$$\frac{2}{N_c}\sum_{j\neq i}\widehat{h_{ij}}(t)\left(\overline{x_{ji}}(t)-\overline{x_{ii}}(t)\right), \tag{51}$$

$$\frac{2}{N_c}\widehat{h_{ji}}(t)\left(\overline{x_{ii}}(t)-\overline{x_{ji}}(t)\right),\ \text{for } j\neq i. \tag{52}$$

The effect of these terms is that opinions about i attract each other, which keeps them close to each other. Therefore the positive and negative biases are combined into a common trend shared by all opinions about i.

This common trend can be expressed by the first order equilibrium opinion offset $e_i(t)$ of agent i, or equilibrium opinion for short, which is defined as follows:

$$e_i(t)=\frac{1}{1+S_i(t)}\left(\overline{x_{ii}}(t)+\sum_{j\neq i}\frac{\widehat{h_{ij}}(t)}{\widehat{h_{ji}}(t)}\overline{x_{ji}}(t)\right), \tag{53}$$

with:

$$S_i(t)=\sum_{j\neq i}\frac{\widehat{h_{ij}}(t)}{\widehat{h_{ji}}(t)}. \tag{54}$$

Indeed, when there is no gossip, applying Eqs. (38) and (44) yields:

$$e_i(t+1)=e_i(t)+\frac{2}{N_c(1+S_i(t))}\sum_{i\neq j}\overline{h'_{ji}}(t)\left(\overline{x_{ii}(t).x_{ji}(t)}-\overline{x_{ii}^2}(t)+\frac{\widehat{h_{ij}}(t)}{\widehat{h_{ji}}(t)}\left(\overline{x_{ji}^2}(t)-\overline{x_{ii}(t).x_{ji}(t)}\right)\right). \tag{55}$$

At any time step t, $e_i(t)$ is the value that would be reached by all the opinions about i if the $h_{ij}(t)$ were frozen. More precisely, imagining that from a given time t_0, for all $t > t_0$ and for all $(i,j) \in \{1,\dots,N_a\}$, $h_{ij}(t) = h_{ij}(t_0)$, then $\overline{x_{ji}}(t)$ for all j would converge to $e_i(t_0)$ and remain at this value. Therefore, the term of second order in Eq. (55) determines the second order effect applied to an opinion which is at the equilibrium of the first order effects. In the long run, this trend is common to all opinions about i, as the opinions about i reach their equilibrium distances from each other (see trajectory examples on Fig. 3).

The trend is thus expressed as a weighted sum of the positive bias on the self-opinion and the negative biases on the opinions about i. The negative biases are multiplied by the factor $\frac{\widehat{h_{ij}(t)}}{\widehat{h_{ji}(t)}}$, which is smaller than 1 when $\overline{a_{ii}} > \overline{a_{ij}}$ and $\overline{a_{jj}} < \overline{a_{ji}}$ and higher than 1 when $\overline{a_{ii}} < \overline{a_{ij}}$ and $\overline{a_{jj}} > \overline{a_{ji}}$. Therefore, when the agents are in a consensual hierarchy, these factors are low for agents of i of high status and high for agents i of low status. Hence, the opinions about the agents of low status grow less (or even can decrease) than the opinions about the agents of high status.

When gossip is activated, the equation of $e_i(t+1)$ remains the same except that a term of first order coming from gossip is added. However, simulations show that the effect of this term is negligible. Therefore, like in the case without gossip, $e_i(t)$ provides the common trend of the evolution of the opinions about i.

However, the additional term accounting for gossip modifies the negative bias on the opinion about others. Indeed, we have:

$$\overline{x_{ji}^2}(t+1)=\overline{x_{ji}^2}(t)+\frac{2}{N_c}\left(\overline{G_{ji}^2}(t)-\overline{x_{ji}^2}(t)+\overline{h_{ji}}^2(t)\frac{\delta^2}{3}\right)+\frac{2k}{N_T}\sum_{p\notin\{j,i\}}\left(\overline{J_{jpi}^2}(t)-\overline{x_{ji}^2}(t)+\overline{h_{jp}}^2(t)\frac{\delta^2}{3}\right), \tag{56}$$

with:

$$\overline{G_{ji}}(t)=\overline{x_{ji}(t)}+\widehat{h_{ji}}(t)\left(\overline{x_{ii}}(t)-\overline{x_{ji}}(t)\right), \tag{57}$$

$$\overline{J_{jpi}}(t)=\overline{x_{ji}}(t)+\widehat{h_{ji}}(t)\left(\overline{x_{pi}}(t)-\overline{x_{ji}}(t)\right). \tag{58}$$

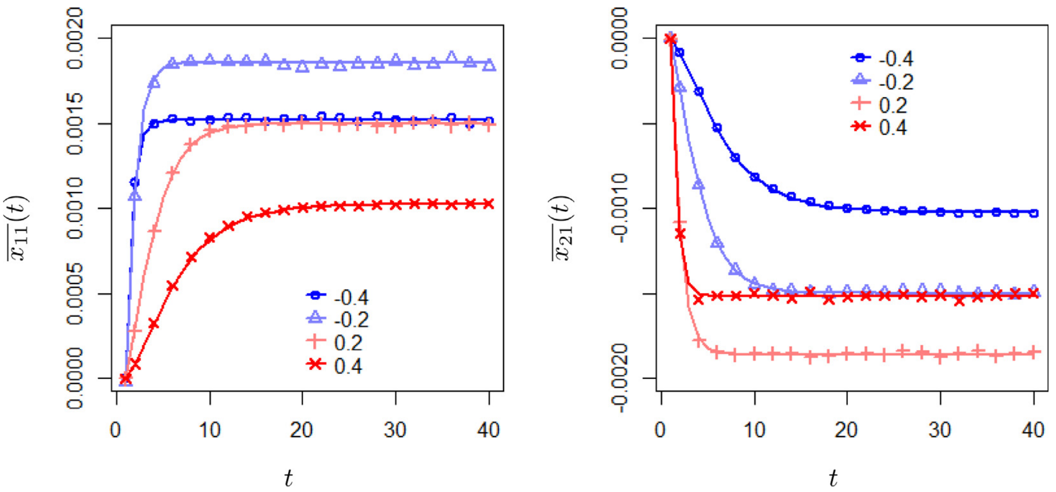

Fig. 2. Biases when only one opinion is varying. Left panel: $\overline{x_{11}}(t)$, right panel: $\overline{x_{21}}(t)$, for $a_{12}(t) = b = 0$. The different colours represent values of $a_{11}(0)$ (left panel) or $a_{21}(0)$ (right panel) −0.4, −0.2, 0.2, and 0.4, as specified in the legend. t is the number of encounters. Each point is the average of the biases over 10 million simulations and the lines are computed with Eq. (21) for the positive bias and its equivalent for the negative bias. Influence parameter $\sigma = 0.3$. Noise parameter $\delta = 0.1$.

This additional sum increases $\overline{x_{ji}^2}(t+1)$, which increases the negative bias on x_{ji} in the following time steps. In particular, at time $t = 2$, $\overline{x_{ii}}(2)$ has the same expression as in Eq. (47), and we have, for $i \neq j$:

$$\overline{x_{ji}}(2) = \frac{4}{N_c^2} h'_{ij}(0)(1 - h_{ij}(0))^2 \frac{\delta^2}{3} + \frac{4}{N_c N_T} h'_{ij}(0) \left(\sum_{p \notin \{i,j\}} h_{jp}^2(0) \right) \frac{\delta^2}{3}, \tag{59}$$

When the agents are in a consensual hierarchy, the additional negative bias is stronger for i of low status, because $\overline{h_{jp}}(t)$ is higher for j of low status and $\overline{h'_{ij}}(t)$ is higher when the statuses of i and j are close. This explains a stronger negative effect of gossip on agents of low status.

4. Numerical experiments

In a first set of experiments, we check the accuracy of the moment approximation. In the second set of experiments, using the moment approximation, we investigate the effect of inequalities on the evolution of the opinions.

4.1. Accuracy of the moment approximation

4.1.1. Examples when only one opinion varies between two interacting agents

We first check the accuracy of the approximation in the simplified setting of Section 3.1 where only x_{11} or x_{21} is varying. Fig. 2 shows the value of $\overline{x_{11}}(t)$ and $\overline{x_{21}}(t)$ from the theoretical formulas (Eq. (21) and its transformation for the negative bias) and from the average of 10 million repetitions of the simulation during 40 encounters. The value of $a_{12}(t) = b = 0$ is fixed and the curves corresponding to four different values of $a_{11}(0) = a_{21}(0) = a$ are shown in different colours in the graphs (see legend). The approximation appears very accurate.

4.1.2. Examples of trajectories of opinions about an agent for 10 interacting agents, without gossip

Fig. 3 shows examples of the trajectories of $\overline{x_{ji}}(t)$ for $j \in \{1, \ldots, N_a\}$, for a given agent i. The title above each panel specifies the value of $a_{ii}(0)$. The lines (solid for the self-opinions $\overline{x_{ii}}(t)$, dashed for the opinions $\overline{x_{ji}}(t)$ with $j \neq i$) are obtained by the moment approximation while the points are the average values of 10 million simulations. The accuracy seems quite satisfactory.

The trajectory of the equilibrium opinion, computed with the moment approximation (see Section 3.4), is represented in green. In the top panels, i is of high status (high values of $a_{ii}(0)$) and from $t = 300$, all the trajectories grow with a very similar slope. This is not the case for the bottom panels where i is of low status (low values of $a_{ii}(0)$). Indeed, the trajectories are increasing for j of high status (red shades) while they are decreasing for j of low status (blue shades). The trajectory of the equilibrium opinion (in green) tends to be closer to the trajectories of $\overline{x_{ji}}(t)$ for j of high status (red shades). For larger values of t however (not visible on the graph), all the trajectories become progressively almost parallel. Moreover, in all cases in our simulations, for all $j \neq i$ and for all t, we have: $\overline{x_{ii}}(t) > \overline{x_{ji}}(t)$.

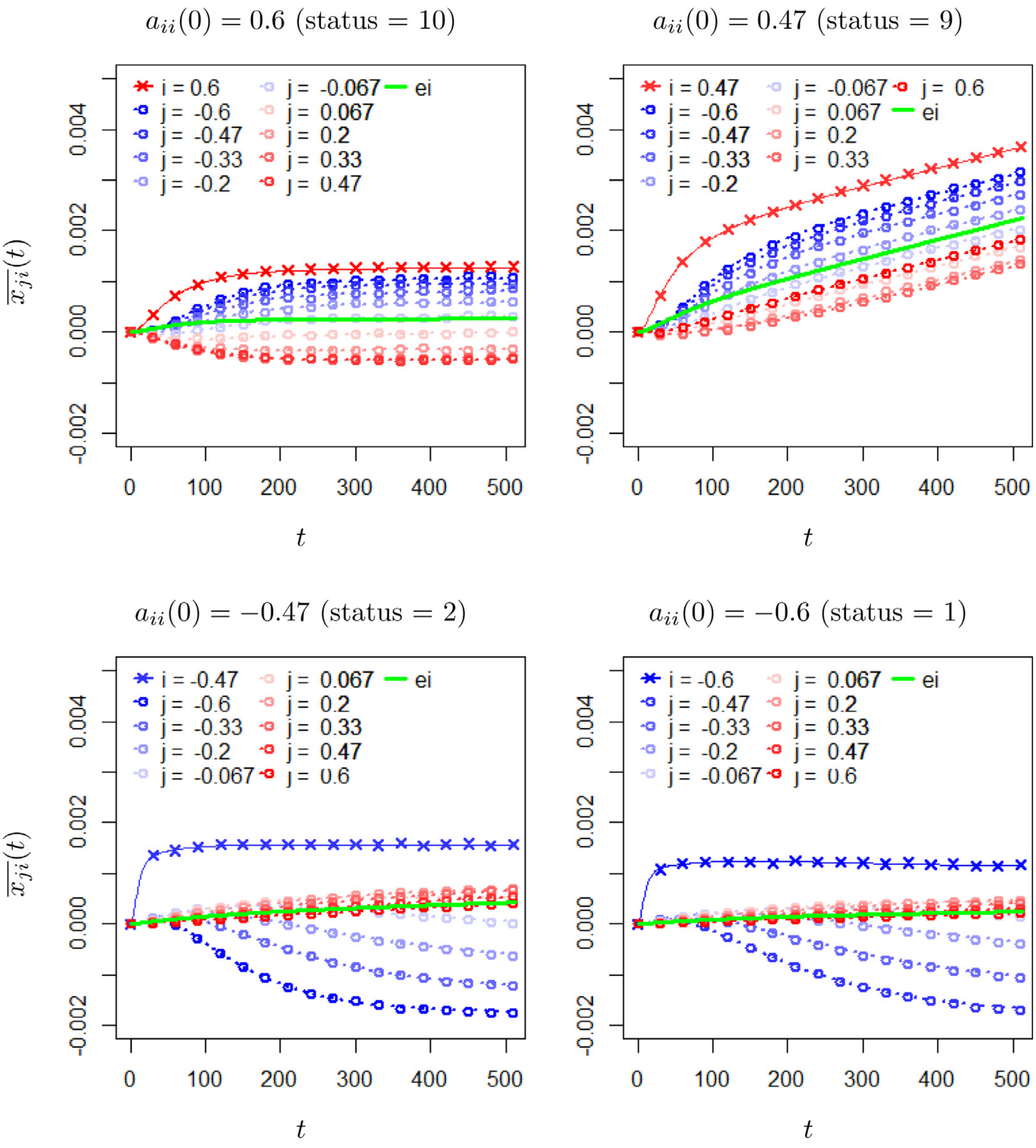

Fig. 3. Examples of evolution of average opinion offsets $\overline{x_{ji}}(t)$ for 10 agents without gossip, with $a_{ii}(0) \in [-0.6, 0.6]$. The lines are obtained with the moment approximation and the points by averaging the results of 10 million simulations. The legend indicates the colour corresponding to the rank of the initial self-opinion $a_{ii}(0)$. The equilibrium opinion $e_i(t)$ is represented in green. Noise parameter $\delta = 0.1$. Influence function parameter $\sigma = 0.3$. Number of gossip $k = 0$.

4.1.3. RRMSE of the moment approximation for different number of agents

In order to evaluate more quantitatively the accuracy of the moment approximation, we compute the root of the relative mean squared error (RRMSE)[2] between the moment approximation and the average results over 10 million simulations. The initial opinions are all such that for all (i, j), $x_{ji}(0) = x_{ii}(0)$, and $x_{ii}(0)$ are regularly distributed on the interval $[-0.3, 0.3]$. For any t and any couple (i, j), let $\overline{\overline{x_{ji}}}(t)$ be the average of $x_{ji}(t)$ over 10 million simulations. Keeping the notation $\overline{x_{ji}}(t)$ for the moment approximation, $\mathcal{E}\left(\overline{x_{ji}}(1, \ldots, T)\right)$, the RRMSE of $\overline{x_{ji}}(t)$ for $t \in [1, T]$ is:

$$\mathcal{E}\left(\overline{x_{ji}}(1, \ldots, T)\right) = \frac{\sqrt{T\left(\sum_{t=1}^{T}\left(\overline{x_{ji}}(t) - \overline{\overline{x_{ji}}}(t)\right)^2\right)}}{\sum_{t=1}^{T}|\overline{\overline{x_{ji}}}(t)|}. \tag{60}$$

We define the RRMSE for the second moment variables similarly.

[2] The RRMSE is the root mean squared error divided by the mean of absolute value of the target values.

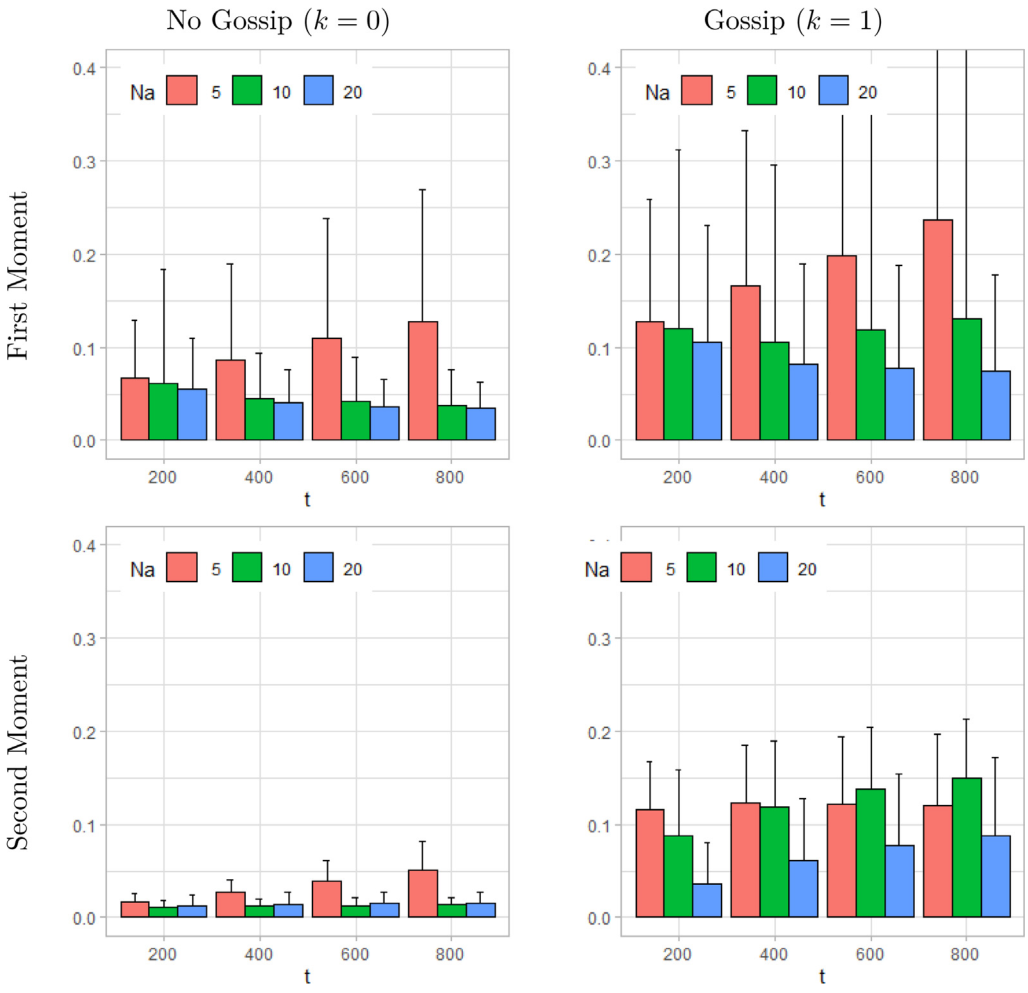

Fig. 4. Average RRMSE of the approximation (defined by Eq. (60)). Top panels: average for all first moment variables ($\overline{x_{ii}}(t)$ and $\overline{x_{ji}}(t)$). Bottom panels average for all second moment variables ($\overline{x_{ii}^2}(t)$, $\overline{x_{ji}^2}(t)$ and $\overline{x_{ii}(t)x_{ji}(t)}$) for a number of agents $N_a \in \{5, 10, 20\}$. The RRMSE is computed on the interval $[1, t]$, t being defined on the horizontal axis. The error bars show the standard deviations in the considered set of variables (N_a^2 variables for the first moment, N_a^3 variables for the second moment). Noise parameter $\delta = 0.1$. Influence function parameter $\sigma = 0.3$.

Fig. 4 shows the average RRMSE computed for the first moment and the second moment variables, for a number of agents $N_a \in \{5, 10, 20\}$. The RRMSE is on average lower than 10% for the dynamics without gossip and $N_a > 5$. With gossip, the RRMSE is higher, but it is lower than 15% when $N_a > 5$. It can be expected that the approximation, neglecting the terms of degree higher than 2, is more accurate while the values of x_{ji} and x_{ji}^2 remain small. When the number of agents increases, we have seen that $\overline{x_{ji}}(t)$ is multiplied by a factor of order $\frac{1}{Na^4}$, therefore it can be expected that the error gets smaller for the same values of t, when N_a increases. Similarly, gossip increases $x_{ji}^2(t)$, which is expected to decrease the approximation accuracy at t.

4.2. Effect of initial inequalities on the evolution of the opinions

In the following experiments, we illustrate the effect of initial inequalities on the evolution of opinions in the short term (a few hundreds of interactions) on the case of 10 agents. Indeed, this number of agents is low enough for readable exhaustive representations and high enough for a reasonably accurate moment approximation (RRMSE < 15% on average). The initial opinions are the same in each column of the opinion matrix, and the initial self-opinions (that equal the initial reputations) are regularly distributed in an interval $[-w, w]$. Increasing w corresponds to increasing inequalities. The status of the agent of the highest initial self-opinion is 10 (highest status) and the status of the agent of initial lowest self-opinion is 1 (lowest status).

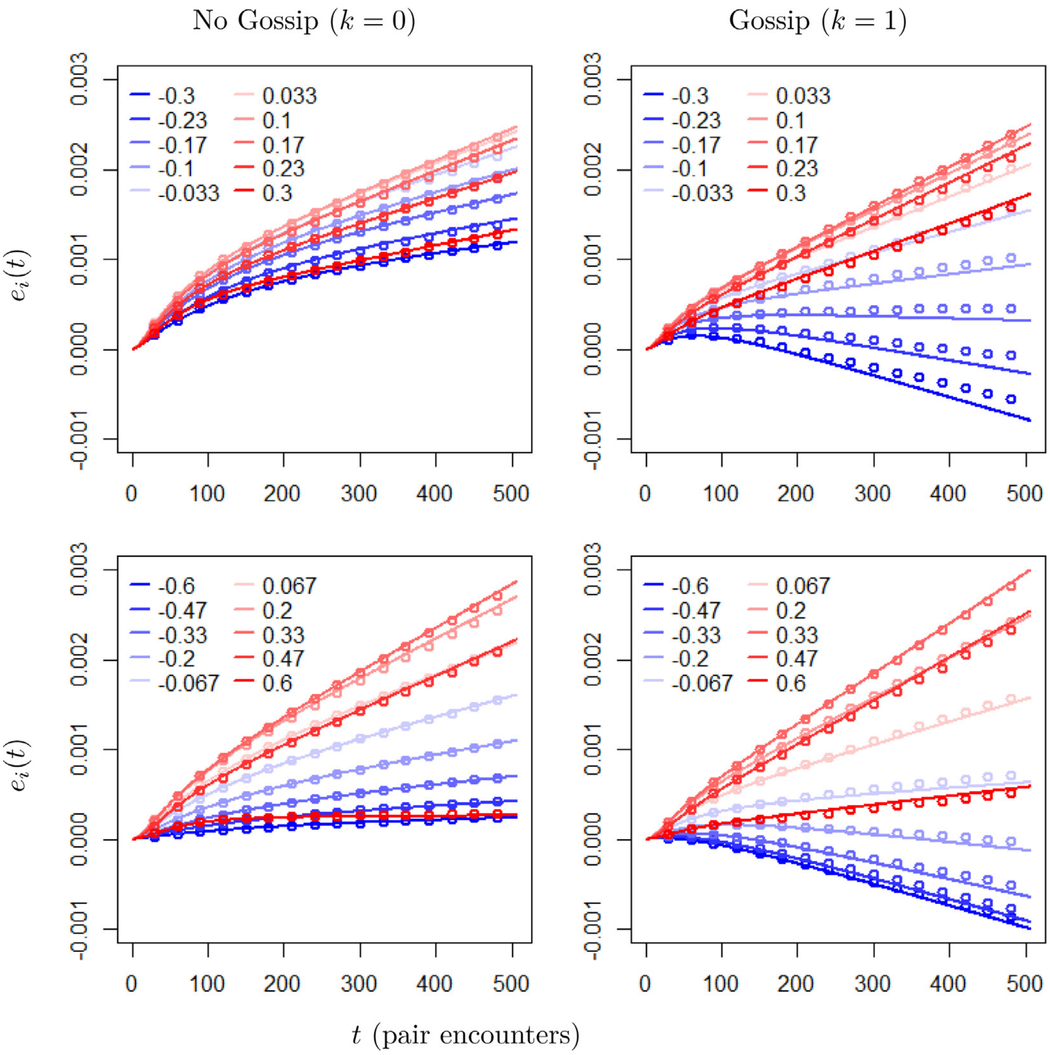

Fig. 5. Evolution of equilibrium opinions ($\overline{r}_i(t)$) for 10 agents with $a_{ii}(0) \in [-0.3, 0.3]$ (top panels) or $a_{ii}(0) \in [-0.6, 0.6]$ (bottom panels). The lines are obtained with the moment approximation and the points by averaging the results of 10 million simulations. The legend provides the colour corresponding to the initial self-opinion $a_{ii}(0)$. Noise parameter $\delta = 0.1$. Influence function parameter $\sigma = 0.3$.

4.2.1. Comparing trajectories of equilibrium opinion for two different inequality widths

Fig. 5 represents the average evolution of equilibrium opinions $e_i(t)$ for $i \in \{1, \dots, 10\}$ during 500 pair encounters when the starting self-opinions are uniformly distributed in $[-0.3, 0.3]$ or in $[-0.6, 0.6]$ and with or without gossip. The main features shown by this figure are the following:

- In the top left panel, with small inequalities and without gossip, all equilibrium opinions are increasing and remain close to each other.
- In the top right panel, with small inequalities and with gossip, the two lowest equilibrium opinions are slightly decreasing. The trajectories of highest status agents are similar with or without gossip.
- In the left bottom panel, with large inequalities and without gossip, the trajectories for agents of high status (in shades of red) increase like when inequalities are low, except for the agent of top status which increases more slowly. However, for the agents of lower status (in shades of blue), the trajectories increase significantly less than when inequalities are low;
- In the bottom right panel, with large inequalities and with gossip, the trajectories for agents of high status are similar to the ones without gossip. However, for the four agents of lowest status, the trajectories are significantly different; they are decreasing (in blue).

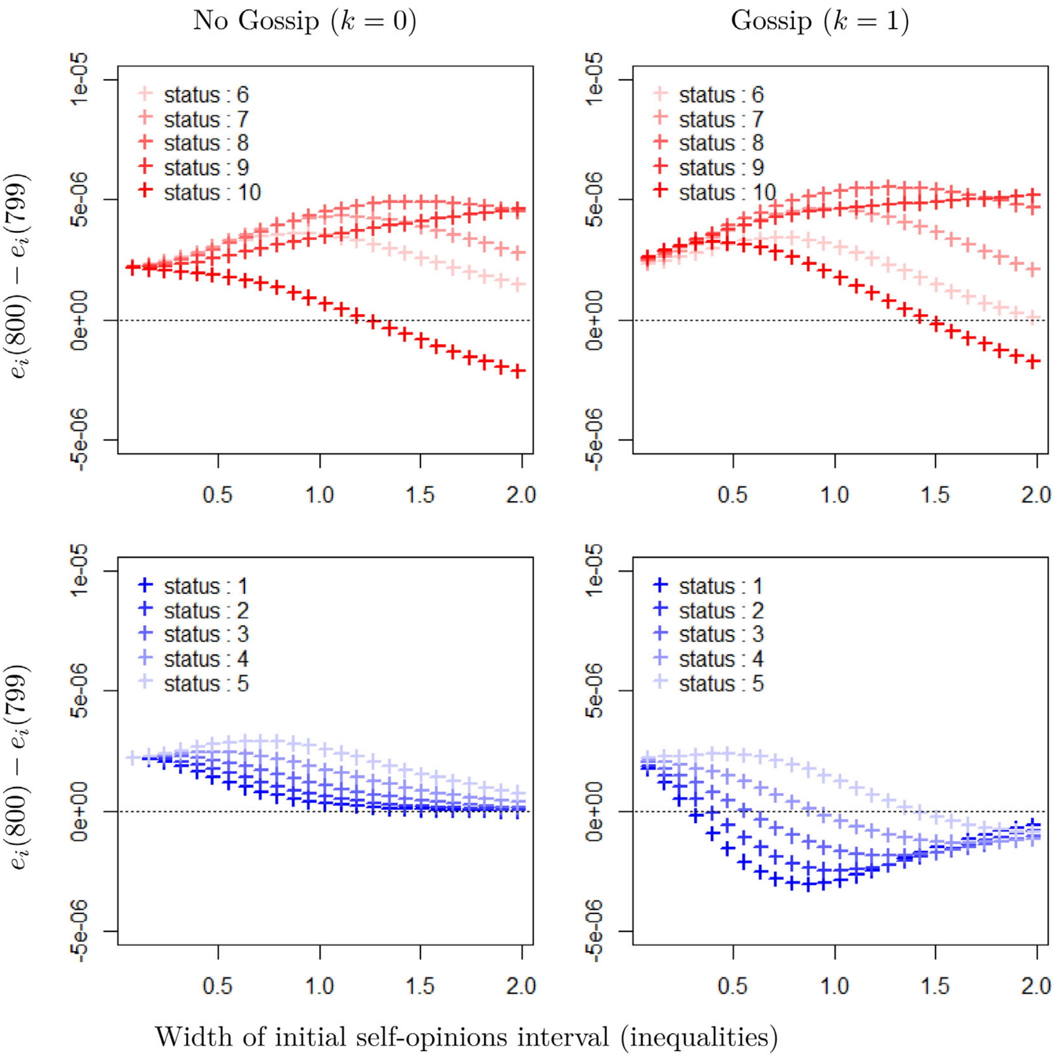

Fig. 6. Slope of equilibrium opinion trajectory at $t = 800$ ($e_i(800) - e_i(799)$) computed by the moment approximation, for 25 different initial ranges of opinions (horizontal axis). The colour of the points codes for the status of the agent; the top panels represent the high statuses and the bottom panels the low statuses. On the left panels, there is no gossip, on the right panels there is ($k = 1$). Noise parameter $\delta = 0.1$. Influence function parameter $\sigma = 0.3$.

4.2.2. Slope of equilibrium opinion trajectory at after 800 encounters when inequalities vary

In each panel of Fig. 6, the *x*-axis represents the width of the initial opinion intervals varying from $[-0.06, 0.06]$ to $[-0.9, 0.9]$, the curves represent $e_i(800) - e_i(799)$, the slope of the trajectory of the equilibrium opinion at $t = 800$, computed with the moment approximation. The colour of the curve codes for the status of agent *i*. In general, this slope is close to the slopes of the opinions about agent *i*.

This figure shows that:

- For agents *i* of high status (in red, top panels):
  - The left and right panels are similar, except for the agent of status 6 for which the equilibrium opinion shows a significantly lower slope with gossip, when inequalities increase;
  - The slope for the agent of the highest status decreases when the inequalities are above a threshold, while the slopes for all the other agents of high status remain positive, both with and without gossip;
  - The agents of the highest statuses (7 and above) have a slightly higher slope when there is gossip:
- For agents *i* of low status (in blue, bottom panels):
  - When there is no gossip (left panel), all the slopes remain positive, but they all tend to 0 when the inequalities increase;

- When there is gossip (right panel), the slopes become negative as the inequalities increase. The slope for the agent of the lowest status becomes negative first, then the slope for the agent for second lowest status becomes negative and so on until the slopes of all the agents of low status become negative.

5. Discussion

5.1. Relevance of the moment approximation

The moment approximation appears reliable while the number of agents is higher or equal to 10 and the number of encounters remains below 1000. Moreover, it provides explanations of the model behaviour.

- After a few hundred interactions, the opinions about an agent tend to evolve in parallel, and their evolution is driven by a second order effect, which is a weighted sum of a positive bias on self-opinions and negative biases on the opinions about others;
- The weights on the negative biases are low for agents of high status and high for the agents of low status and these differences increase when the inequalities increase. This explains why the opinions about the low status agents tend to stagnate or decrease (when there is gossip);
- When there is gossip, an additional term increases the negative bias on the opinion about others. Moreover, this added term is stronger when i and j are of low status and quite small for agents i of high status, especially when the initial inequalities are high. This explains the observed higher negative effect of gossip on the agents of low status.

However, we could not explain mathematically that the average self-opinion of agent is systematicall higher than all the average opinions about this agent, which expresses a more standard definition of positive bias.

In addition to the explanations that mathematical expressions can bring, the moment approximation provides a means to explore the average behaviour of the agent based model, without running millions of simulations. Such an exploration for 10 agents, when varying the width of the interval of initial self-opinions, reveals the following features (see Fig. 6):

- The opinions about the agents of high status (except the top status) tend to grow in roughly the same way, with or without gossip;
- Without gossip, the opinions about agents of low status tend to grow but this growth progressively decreases and even becomes close to zero when the initial inequalities increase. With gossip, these opinions grow only when the inequalities are very low and then the opinions about low status agents start decreasing as the inequalities increase.

The same explorations conducted with 20 and 40 agents yield similar results.

These observations provide some explanations to the patterns recalled in Section 2.2. Indeed, initially, in these patterns, all the opinions are the same, therefore, both with and without gossip, all the average opinions tend to grow together in a first period of a few thousand steps. However, because of the noise, more or less dispersion of the opinions takes place, introducing inequalities between agents:

- Without gossip, since all opinions tend to grow at a similar pace, the opinion inequalities remain moderate for a while and all opinions grow on average. When the inequalities reach a threshold though, the opinions about agents of low status grow more and more slowly or stagnate, while the opinions about agents of high status fluctuate when reaching the opinion limit at +1. Overall, the distribution of opinions is therefore significantly positive on average;
- When there is gossip, because the opinions about agents of low status grow much more slowly than the opinions about agents of high status, the inequalities of opinions increase more rapidly and easily reach a level in which the opinions about the lowest status agent starts decreasing, which further increases the opinion inequalities, and the opinions about other agents of low status start decreasing, which further increases the opinion inequalities. Ultimately, when the inequalities are maximum, the opinions about a majority of agents tend to decrease. This explains why the overall distribution of opinions becomes negative on average.

We checked the validity of these explanations by introducing a process that limits the inequalities of the opinions by regularly driving them slightly towards their average. More precisely, every N_a interactions, all opinions are modified, using parameter λ, as follows:

$$a_{ij}(t+1) = (1-\lambda)a_{ij}(t) + \lambda.\bar{a}, \forall i, j \in N_a, \tag{61}$$

with $\bar{a}$ being the average opinion. With this modified dynamics, the opinions grow and stabilise to a high average positive value, even when parameter λ is small (0.0001) and when there is gossip.

5.2. Connections with the literature in social-psychology

We now discuss how the model behaviour relates to some researches in social-psychology. We first consider the biases and then we discuss the effect of inequalities on the evolution of opinions.

Social-psychology robustly established that people tend to overestimate themselves (for instance overoptimism or overconfidence in judgement and predictions or the ability to complete a task or about forecasting events in general, see [19] for a review). This tendency is often called positivity bias. A widely accepted explanation relates the positivity bias to the well established tendency of most people to self-enhancement or self-protection. People tend to seek out and accept positive feed-backs and to avoid or reject negative ones [20]. Indeed, when we receive a negative feedback, we often tend to decrease our evaluation of its source and thus we decrease its importance (e.g. [21]). As a result, on average, negative feed-backs tend to have a lower impact than positive ones on self-evaluation, which leads to self-overestimation [22]. This process presents strong similarities with the positive bias observed in the Leviathan model, when vanity is active, because vanity decreases the evaluation of the source of negative feed-backs (see details in [14]).

However, the positive bias studied in this paper is generated by the model without vanity and suggests the existence of another mechanism. Indeed, this positive bias cannot be attributed to any self-enhancement. It is a statistical effect of the noise combined with the decreasing influence function. As far as we know, this specific bias has not been observed by social-psychologists.

Considering the negative bias now, the literature reports some negative tendencies in judging others. People show a negative bias on the opinion about others in some specific contexts, for instance when requested to express their opinion in front of an audience of higher status (see for instance [23]). Also, when judging moral qualities of others, we tend to put a higher weight on the negative features than on the positive ones (while this is the opposite when judging the abilities) [24,25]. However, the contexts of these observations are difficult to relate to our model.

Therefore, it seems that, if they do exist in human interactions, the biases observed in our model have been overlooked by social-psychologists. This would not be surprising, since these biases are small (of second order). Moreover, though our simulations suggest that their long term effect is potentially huge, it is impossible to relate these effects to their causes without the type of analysis that we carried out.

Now, we consider possible connections between the literature in social-psychology and the effect of inequalities observed in the model.

A study involving a cohort of 3058 adolescents in Denmark, followed from ages 15 to 21, shows that the self-esteem of the adolescents from the richest tertile grows significantly more than the self-esteem of the poorest tertile [26]. This result seems in line with our model patterns. However, the mechanisms involved are probably quite different. Indeed, the impact of inequalities on self-esteem is generally related to personal or group self-deprivation or feeling of injustice [27,28]. These feelings are absent from the model. Again, the model shows a statistical phenomenon of second order, while the reaction to self-deprivation, that could probably be also modelled using the vanity process of the Leviathan model, is certainly of first order. Hence, again, it seems very likely that the research in social psychology missed the phenomenon suggested by our model.

Since we cannot rely on existing literature to get adequate experimental data challenging the model, a solution is to get this data by running specifically designed experiments. The following directions can be envisaged:

- The model suggests the existence of a positive bias on self-opinions and a negative bias on opinions about others, without self-enhancement or self-protection (i.e. with symmetric reactions of same intensity to negative and positive feed-backs of same intensity), as soon as the influence function is decreasing when the self-opinion is increasing. The main feature to check is thus this property of the influence function, because its presence mathematically implies the biases. It seems possible to design an experiment achieving this;
- The model suggests that the interactions in a group with wide perceived inequalities tend to widen these perceived inequalities by decreasing the opinions about the agents of low status (especially when there is gossip) and by increasing the opinions about the agents high status. However, in a group with small perceived inequalities, the interactions tend to increase the opinions about all the agents, even if there is gossip. Therefore, in this perspective, introducing mechanisms that limit the perceived inequalities (like the mechanism described by Eq. (61)) in a group should be beneficial to the opinions about all its members. Experiments checking these predictions could consider a participant interacting with a set of experimenters and controlling the respective statuses of the participant and of the experimenters.

Overall, this work identifies second order effects of interactions that seem impossible to observe without suspecting their existence. However, it seems possible to design specifically targeted experiments that would detect them.

Additional information

The code of the model is available at: https://www.comses.net/codebases/12d44111-5823-4773-ad59-754ebacb33a1/releases/1.0.0/.

CRediT authorship contribution statement

Guillaume Deffuant: Designed several versions of the moment approximation of the model, Implemented and tested them in R, Wrote the text of the paper, Arranged the figures. **Thibaut Roubin:** Checked and corrected the formulas of the moment approximation, Developed faster versions of the model in C++, Performed systematic simulations of the approximated and agent models, Computed the differences between the average agent model and its approximation, Provided most materials for the figures of the paper.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Thibaut Roubin reports fnancial support was provided by French Natonal Research Agency.

Acknowledgement

This work has been partly supported by the Agence Nationale de la Recherche through the ToRealSim project within the ORA program and the FuturICT2.0 projet within the Flagera program. We are grateful to Sylvie Huet for her comments on a preliminary version of the paper and to the whole ToRealSim project team for helpful discussions.

Appendix

A.1. Equations of second moments for a given sequence of interactions, without gossip

Let $\widehat{h_{ij}}(s_t) = \overline{h_{ij}}(s_t) - \overline{h'_{ij}}(s_t)\overline{z_{ij}}(s_t)$. For $(i,j) = (i_{t+1}, j_{t+1})$, or $(i,j) = (j_{t+1}, i_{t+1})$:

$$\overline{x_{ii}}(s_{t+1}) = \overline{x_{ii}}(s_t) + \widehat{h_{ij}}(s_t)\left(\overline{x_{ji}}(s_t) - x_{ii}(s_t)\right) - \overline{h'_{ij}}(s_t)\left(\overline{x^2_{ii}}(s_t) - \overline{x_{ii}(t)x_{ji}(s_t)}\right), \tag{62}$$

and:

$$\overline{x_{ji}}(s_{t+1}) = \overline{x_{ji}}(s_t) + \widehat{h_{ji}}(s_t)\left(\overline{x_{ii}}(s_t) - \overline{x_{ji}}(s_t)\right) + \overline{h'_{ji}}(s_t)\left(\overline{x^2_{ji}}(s_t) - \overline{x_{ii}(t)x_{ji}(s_t)}\right). \tag{63}$$

Let:

$$\overline{F_{ij}}(s_t) = \overline{x_{ii}}(s_t) + \widehat{h_{ij}}(s_t)\left(\overline{x_{ji}}(s_t) - \overline{x_{ii}}(s_t)\right). \tag{64}$$

$$\overline{G_{ji}}(s_t) = \overline{x_{ji}(s_t)} + \widehat{h_{ji}}(s_t)\left(\overline{x_{ii}}(s_t) - \overline{x_{ji}}(s_t)\right). \tag{65}$$

Neglecting the terms of order higher than 2, we get:

$$\overline{x^2_{ii}}(s_{t+1}) = \overline{F^2_{ij}}(s_t) + \overline{h_{ij}}^2(s_t)\frac{\delta^2}{3}, \tag{66}$$

with:

$$\overline{F^2_{ij}}(s_t) = (1 - \widehat{h_{ij}}(s_t))^2\overline{x^2_{ii}}(s_t) + \widehat{h_{ij}}^2(s_t)\overline{x^2_{ji}}(s_t) + 2(1 - \widehat{h_{ij}}(s_t))\widehat{h_{ij}}(s_t)\overline{x_{ii}(s_t)x_{ji}(s_t)}, \tag{67}$$

and:

$$\overline{x^2_{ji}}(s_{t+1}) = \overline{G^2_{ji}}(s_t) + \overline{h_{ji}}^2(s_t)\frac{\delta^2}{3}, \tag{68}$$

with:

$$\overline{G^2_{ji}}(s_t) = (1 - \widehat{h_{ji}}(s_t))^2\overline{x^2_{ji}}(s_t) + \widehat{h_{ji}}^2(s_t)\overline{x^2_{ii}}(s_t) + 2(1 - \widehat{h_{ji}}(s_t))\widehat{h_{ji}}(s_t)\overline{x_{ii}(s_t)x_{ji}(s_t)}. \tag{69}$$

Similarly, we get:

$$\overline{x_{ii}(s_{t+1}).x_{ji}(s_{t+1})} = \overline{F_{ij}(s_t)G_{ji}(s_t)}. \tag{70}$$

For $p \neq i$ and $p \neq j$:

$$\overline{x_{ii}(s_{t+1}).x_{pi}(s_{t+1})} = \overline{F_{ij}(s_t)x_{pi}(s_t)}. \tag{71}$$

$$\overline{x_{ji}(s_{t+1}).x_{pi}(s_{t+1})} = \overline{G_{ji}(s_t)x_{pi}(s_t)}. \tag{72}$$

A.2. Equations of second moments for all sequences of interactions, without gossip

For any $i \in \{1, \dots, N_a\}$:

$$\overline{x^2_{ii}}(t+1) = \frac{N_a - 2}{N_a}\overline{x^2_{ii}}(t) + \frac{2}{N_c}\sum_{j \neq i}\left(\overline{F^2_{ij}}(t) + \overline{h_{ij}}(t)^2\frac{\delta^2}{3}\right). \tag{73}$$

For $j \neq i$:

$$\overline{x^2_{ji}}(t+1) = \frac{N_c - 2}{N_c}\overline{x^2_{ji}}(t) + \frac{2}{N_c}\left(\overline{G^2_{ji}}(t) + \overline{h_{ji}}^2\frac{\delta^2}{3}\right). \tag{74}$$

Moreover:

$$\overline{x_{ii}(t+1).x_{ji}(t+1)} = \frac{N_a-2}{N_a}\overline{x_{ii}(t).x_{ji}(t)} + \frac{2}{N_c}\left(\overline{F_{ij}(t)G_{ji}(t)} + \sum_{p\notin\{i,j\}}\overline{F_{ip}(t)x_{ji}(t)}\right). \tag{75}$$

For $(i,j,p) \in \{1,\dots,N_a\}^3$, $i \neq j$, $j \neq p$, $i \neq p$:

$$\overline{x_{ji}(t+1).x_{pi}(t+1)} = \frac{N_c-4}{N_c}\overline{x_{ji}(t).x_{pi}(t)} + \frac{2}{N_c}\left(\overline{x_{ji}(t)G_{pi}(t)} + \overline{x_{pi}(t)G_{ji}(t)}\right), \tag{76}$$

with, for instance:

$$\overline{F^2_{ij}}(t) = (1-\widehat{h_{ij}}(t))^2\overline{x^2_{ii}}(t) + \widehat{h_{ij}}^2(t)\overline{x^2_{ji}}(t) + 2(1-\widehat{h_{ij}}(t))\widehat{h_{ij}}(t)\overline{x_{ii}(t)x_{ji}(t)}. \tag{77}$$

Starting from the values at $t = 0$, applying these equations, we can compute $\overline{x_{ii}}(t)$ and $\overline{x_{ij}}(t)$.

A.3. Equations of second moments for a given sequence, with gossip

For $(i,j) = (i_{t+1}, j_{t+1})$, all the products are the same as the ones in the case without gossip. Moreover, for $g \in \{g_{1_t},\dots,g_{k_t}\}$, let:

$$\overline{J_{ijg}}(s_t) = \overline{x_{ig}}(s_t) + \widehat{h_{ij}}(s_t)\left(\overline{x_{jg}}(s_t) - \overline{x_{ig}}(s_t)\right). \tag{78}$$

We have:

$$\overline{x^2_{ig}}(s_{t+1}) = \overline{J^2_{ijg}}(s_t) + \overline{h_{ij}}^2(t)\frac{\delta^2}{3}. \tag{79}$$

Moreover:

$$\overline{x_{ig}(s_{t+1})x_{jg}(s_{t+1})} = \overline{J_{ijg}(s_t)J_{jig}(s_t)}. \tag{80}$$

A.4. Equations of second moments for all sequences of interactions, with gossip

For $(i,j) \in \{1,\dots,N_a\}^2$, like without gossip we have :

$$\overline{x^2_{ii}}(t+1) = \frac{N_a-2}{N_a}\overline{x^2_{ii}}(t) + \frac{2}{N_c}\sum_{j\neq i}\left(\overline{F^2_{ij}}(t) + \overline{h_{ij}}^2(t)\frac{\delta^2}{3}\right). \tag{81}$$

However, $\overline{x^2_{ji}}(t+1)$ is different with gossip:

$$\overline{x^2_{ji}}(t+1) = P_1.\overline{x^2_{ji}}(t) + \frac{2}{N_c}\left(\overline{G^2_{ji}}(t) + \overline{h_{ji}}^2(t)\frac{\delta^2}{3}\right) + \frac{2k}{N_T}\sum_{p\notin\{j,i\}}\left(\overline{J^2_{jpi}}(t) + \overline{h_{jp}}^2(t)\frac{\delta^2}{3}\right), \tag{82}$$

with:

$$N_T = N_a(N_a-1)(N_a-2), \tag{83}$$

$$P_1 = 1 - \frac{2}{N_c} - \frac{2k}{N_c}. \tag{84}$$

The average products $\overline{x_{ii}(t+1).x_{ji}(t+1)}$ and $\overline{x_{ji}(t+1).x_{pi}(t+1)}$ are also different with gossip:

$$\overline{x_{ii}(t+1).x_{ji}(t+1)} = P_2.\overline{x_{ii}(t).x_{ji}(t)} + \frac{2}{N_c}\left(\overline{F_{ij}(t)G_{ji}(t)} + \sum_{p\notin\{i,j\}}\overline{F_{ip}(t)x_{ji}(t)}\right) + \frac{2k}{N_T}\sum_{p\notin\{j,i\}}\overline{J_{jpi}(t)x_{ii}(t)}, \tag{85}$$

with:

$$P_2 = 1 - \frac{2}{N_a} - \frac{2k}{N_c}. \tag{86}$$

Moreover, for $p \neq i$ and $j \neq i$ and $p \neq j$:

$$\begin{aligned}\overline{x_{ji}(t+1).x_{pi}(t+1)} = {} & P_3.\overline{x_{ji}(t).x_{pi}(t)} + \frac{2}{N_c}\left(\overline{x_{pi}(t)G_{ji}(t)} + \overline{x_{ji}(t)G_{pi}(t)}\right) + \frac{2k}{N_T}\overline{J_{jpi}(t)J_{pji}(t)} \\ & + \frac{2k}{N_T}\sum_{q\notin\{j,i,p\}}\left(\overline{J_{jqi}(t)x_{pi}(t)} + \overline{J_{pqi}(t)x_{ji}(t)}\right),\end{aligned} \tag{87}$$

with:

$$P_3 = 1 - \frac{4}{N_c} - \frac{2k}{N_T} - \frac{4k(N_a - 3)}{N_t}. \quad (88)$$

References

[1] J. French, A formal theory of social power, Psychol. Rev. 63 (3) (1956) 181–194.
[2] S. Galam, Minority opinion spreading in random geometry, Eur. Phys. J. B 25 (4) (2002) 403–406.
[3] G. Deffuant, D. Neau, F. Amblard, G. Weisbuch, Mixing beliefs among interacting agents, Adv. Complex Syst. 3 (2000) 87–98.
[4] G. Deffuant, Comparing extremism propagation patterns in continuous opinion models, J. Artif. Soc. Soc. Simul. 9 (3) (2006).
[5] R. Hegselmann, U. Krause, Opinion dynamics and bounded confidence: Models, analysis and simulation, J. Artif. Soc. Soc. Simul. 5 (3) (2002).
[6] A. Flache, M. Macy, Small worlds and cultural polarization, J. Math. Sociol. 35 (1–3) (2011) 146–176.
[7] A. Flache, M. Maes, T. Feliciani, E. Chattoe-Brown, G. Deffuant, S. Huet, J. Lorenz, Models of social influence: Towards the next frontiers, J. Artif. Soc. Soc. Simul. 20 (4) (2017).
[8] F. Bagnoli, T. Carletti, D. Fanelli, A. Guarino, A. Guazzini, Dynamical affinity in opinion dynamics modeling, Phys. Rev. E 76 (2007) 66105–66108.
[9] T. Carletti, D. Fanelli, R. Simone, Emerging structures in social networks guided by opinions, Adv. Complex Syst. 14 (01) (2011) 13–30.
[10] B. Latané, The psychology of social impact, Am. Psychol. 36 (1981) 343–356.
[11] J. Holyst, K. Kacperski, F. Schweitzer, Social impact models of opinion dynamics, Annual Review of Computational Physics IX (2001) 253–273.
[12] E. Bonabeau, G. Theraulaz, J.-L. Deneubourg, Phase diagram of a model of self-organizing hierarchies, Physica A 217 (3) (1995) 373–392.
[13] G. Deffuant, I. Bertazzi, S. Huet, The dark side of gossips: Hints from a simple opinions dynamics model, Adv. Complex Syst. 21 (2018) 1–20.
[14] G. Deffuant, T. Carletti, S. Huet, The leviathan model: Absolute dominance, generalised distrust and other patterns emerging from combining vanity with opinion propagation, J. Artif. Soc. Soc. Simul. 16 (23) (2013).
[15] S. Huet, F. Gargiulo, F. Pratto, Can gender inequality be created without intergroup discrimination, PLoS One 15 (12) (2020).
[16] M.R. Leary, E. Tambor, S. Terdal, D. Downs, Self-esteem as an interpersonal monitor: The sociometer hypothesis, J. Personal. Soc. Psychol. 16 (11) (2005) 76–111.
[17] S. Huet, G. Deffuant, The Leviathan Model Without Gossips and Vanity: The Richness of Influence Based on Perceived Hierarchy, Springer, 2017, pp. 149–162.
[18] R. Law, U. Dieckmann, Moment approximations of individual-based models, in: U. Dieckmann (Ed.), The Geometry of Ecological Interactions: Simplifying Spatial Complexity, Cambridge University Press, 2000, pp. 252–269.
[19] D. Dunning, C. Heath, S. Jerry, Flawed self-assessment. Implications for health, education and the workplace, Psychol. Sci. Public Interest 21 (2004) 69–106.
[20] C. Sedikides, M. Strube, Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better, Adv. Exp. Soc. Psychol. 9 (1997) 209–269.
[21] W.K. Campbell, C. Sedikides, Self-threat magnifies the self-serving bias: A meta-analytic integration, Rev. General Psychol. 3 (1999) 23–43.
[22] R.L. Moreland, P.D. Sweeney, Self-expectancies and relations to evaluations of personal performance, Personality 52 (1984) 156–176.
[23] T. Amabile, A. Glazebrook, A negativity bias in interpersonal evaluation, J. Exp. Soc. Psychol. 18 (1981) 1–22.
[24] C. Martijn, R. Spears, J.V.D. Pligt, E. Jakobs, Negativity and positivity effects in person perception and inference: Ability versus morality, European J. Soc. Psychol. 22 (5) (1992) 453–463.
[25] J. Skowornski, D. Carlston, Caught in the act: When impressions based on highly diagnostic behaviours are resistant to contradiction, European J. Soc. Psychol. 22 (5) (1992) 435–452.
[26] C. Hansen, The impact of economic inequalities on self-esteem and depression: a longitudinal study from Denmark, European J. Public Health (ISSN: 1101-1262) 24 (suppl_2) (2014) cku166-052.
[27] I. Walker, Effects of personal and group relative deprivation on personal and collective self-esteem, Group Process. Intergroup Relat. 2 (4) (1999) 365–380.
[28] J. Crocker, H. Blanton, Social inequality and self-esteem: The moderating effects of social comparison,legitimacy, and contingencies of self-esteem, in: T. Tyler, R. Kramer, O. John (Eds.), The Psychology of the Social Self, Lawrence Erlbaum Associate, 1999, pp. 171–191.