Anisotropic opinion dynamics with an adaptive social rule

Juan Neirotti

Department of Mathematics, Aston University, The Aston Triangle, B4 7ET, Birmingham, United Kingdom

We consider an anisotropic opinion formation process where the set of rules B, that dictates what is the socially acceptable position, changes following the average voters' opinion. As in the case of a constant B, conservative (agreement with B) and liberal (agreement with neighbors) voters' attitudes are still represented by stable fixed points in the phase space of the system but with the difference that the conservative fixed point is stable for all possible values of the intervoter interaction. It has been also observed that, when the model is applied to sufficiently large populations, the time needed to consolidate a position in agreement with B is finite. We observed that there is also a range of values of the interaction where the two stable points coexist, opening the door for the modeling of bistability related phenomena, such as stochastic resonance and hysteresis.

I. INTRODUCTION

Opinions are highly dynamical mental representations of the individual's beliefs, resulting from inference processes frequently performed with insufficient information. They play a fundamental role in the individual's reaction to social situations that can trigger collective responses. In this article we analyze a model of anisotropic opinion formation in a community of interacting agents [1] where the social rule B, which is the model's source of anisotropy, slowly changes with the average opinion of the population. As is presented in Ref. [1], we assume that the agents are adaptive in the sense that they learn to produce an opinion from the available information, and that such a learning process is influenced by B and by neighboring agents. It has been observed that the dynamics that results from this process (and extended in Ref. [2]) presents two stable points in the phase space of the system; one, dubbed conservative, for low values of the interagent interaction, the other, dubbed liberal, for sufficiently high values of the interaction. These two points do not coexist if B is constant. We will demonstrate in the following that if B adapts following the average opinion of the population, the conservative fixed point becomes stable for all possible values of the interagent interaction.

Following Ref. [1] we will represent social issues $\mathbf{S} \in \{\pm 1\}^N$ by binary strings of length N. These issues are classified as socially acceptable (or not) if $\sigma_\mathbf{B}(\mathbf{S}) \equiv \mathrm{sgn}(\mathbf{B} \cdot \mathbf{S}) = 1\,(-1)$, where $\mathbf{B} \in \mathbb{R}^N$ is the (internal representation) vector associated with the perceptron B (for a clear definition of the perceptron network, see Ref. [3]), the function sign is defined as $\mathrm{sgn}(x) = 1$ if $x > 0$, -1 if $x < 0$, and 0 if $x = 0$ and the inner product is the usual one $\mathbf{B} \cdot \mathbf{S} \equiv \sum_{i=1}^N B_i S_i$. The agents, also represented by perceptrons, classify the social issues according to their own internal representations $\mathbf{J}_a \in \mathbb{R}^N$, according to $\sigma_{\mathbf{J}_a}(\mathbf{S}) \equiv \mathrm{sgn}(\mathbf{J}_a \cdot \mathbf{S})$, where $\sigma_{\mathbf{J}_a}(\mathbf{S})$ is the opinion of agent a on issue $\mathbf{S}$. In the current scenario we consider all vectors, i.e., $\mathbf{B}$ and all $\mathbf{J}_a$'s, to be plastic. Such plasticity is manifested through the way the agents' internal vectors get modified through the interaction with the society and their neighbors.

We represent the topology of the society by a directed graph $\boldsymbol{G} = \{\{a\}, \{g_{a,b}\}\}$ where $\{a\}$ is a set of vertexes associated with the social agents and $\{g_{a,b}\}$ is a set of strengths $g_{a,b}$ that represent the influence of agent b on agent a. The neighborhood of a is defined as $\mathbb{N}_a = \{c \in [M]: g_{a,c} > 0\}$. The total number of vertexes (or agents) is M, and the average neighborhood size is defined as

$$\nu \equiv \frac{1}{M} \sum_{a=1}^M |\mathbb{N}_a|, \tag{1}$$

where $|\mathbb{N}_a|$ is the cardinality of the set $\mathbb{N}_a$. We will consider graphs with $\nu \sim O(1)$ only. We say a bond (a, b) is *active* if $g_{a,b} > 0$ and *passive* otherwise.

A. Update algorithms

Assuming that the population of interacting agents receives information taken from the set $\mathbb{S} \equiv \{(\sigma_{B,n}, \sigma_{\mathbb{N}_a,n}, \mathbf{S}_n), n = 1, \dots, T\}$, where the issue $\mathbf{S}_n$ is presented at time n and then discarded, $\sigma_{B,n} = \mathrm{sgn}(\mathbf{B}_n \cdot \mathbf{S})$ and $\sigma_{\mathbb{N}_a,n} = \{\mathrm{sgn}(\mathbf{J}_{c,n} \cdot \mathbf{S}_n): c \in \mathbb{N}_a\}$ is the set formed by the opinions of the agent a's neighbors, the update equation for the internal representation of a is as follows:

$$\mathbf{J}_{a,n+1} = \mathbf{J}_{a,n} + \psi_{a,n} \frac{\sigma_{B,n} \mathbf{S}_n}{\sqrt{N}}, \tag{2}$$

where $\sigma_B \mathbf{S}/\sqrt{N}$ is the (unit length) Hebb vector [4], that indicates the direction of the socially acceptable position on $\mathbf{S}_n$ and $\psi_{a,n}$ is the learning amplitude, that regulates how the information is incorporated in the internal representation of a. The length of the opinion formation process T is considered to be proportional to the number of issues presented to the agents. Based on social corroboration experiments [5–7] and assuming that agent a is connected with the agents in $\mathbb{N}_a$, we propose $\psi_a \equiv f |\mathbf{J}_a|/\sqrt{N} \Psi_a$ where f is a units' constant, $|\mathbf{J}_a|/\sqrt{N} = \sqrt{\sum_{j=1}^N J_{a,j}^2/N}$ is a factor that has no impact on the learning efficiency of the algorithm [8], it has been only

considered for technical purposes, and

$$\Psi_a \equiv 1 - \Theta(-\sigma_B \sigma_a) \sum_{c \in \mathbb{N}_a} \frac{g_{a,c}}{f} \Theta(\sigma_a \sigma_c), \quad (3)$$

where $\Theta(x) = 1$ if $x > 0$ and 0 otherwise is the Heaviside step function. The update algorithm (2) has been introduced (and discussed) in Ref. [1]. Let us define the unit vectors $\boldsymbol{b} \equiv |\mathbf{B}|^{-1}\mathbf{B}$ in the direction of the internal representation of B, $\boldsymbol{j}_a \equiv |\mathbf{J}_a|^{-1}\mathbf{J}_a$ in the direction of the internal representation of agent a and $\boldsymbol{j}_{a,\perp} = [1 - (\boldsymbol{j}_a \cdot \boldsymbol{b})^2]^{-1/2}[\boldsymbol{j}_a - (\boldsymbol{j}_a \cdot \boldsymbol{b})\boldsymbol{b}]$ in the direction of the component of $\mathbf{J}_a$ perpendicular to $\mathbf{B}$. Given that an agent's opinion is obtained through information processing using the internal representation vector $\mathbf{J}_a$ and that any modification to the vector $\mathbf{B}$ in the direction of $\mathbf{B}$ does not produce any change in B's opinions, we will construct the update algorithm for B by considering the vector,

$$\mathbf{L} \equiv \frac{1}{M} \sum_{c=1}^{M} \boldsymbol{j}_{c,\perp}, \quad (4)$$

which is the arithmetic average over all the components of the internal representations $\mathbf{J}_c$ perpendicular to $\mathbf{B}$. Observe that $\mathbf{B} \cdot \mathbf{L} = 0$ and $|\mathbf{L} \cdot \mathbf{L}| \leqslant 1$. Then,

$$\mathbf{B}_{n+1} = \mathbf{B}_n + \frac{\lambda}{\sqrt{N}} f \mathbf{L}_n, \quad (5)$$

where $\lambda/\sqrt{N}$ is a suitable scale factor. Observe that if $\lambda \sim O(1)$ the updates of $\mathbf{B}$ at each time step are very small thus $\lambda/\sqrt{N}$ is a measure of the inverse social inertia (if the *mass* of B is infinite we would not expect any change at all, thus $\lambda = 0$). Observe also that $|\mathbf{B}_{n+1}|^2 = |\mathbf{B}_n|^2 + O(f^2 N^{-1})$, which implies (see below) that the length of vector $\mathbf{B}$ does not change with the update.

To help describe the state of the system, we define the variables,

$$\phi_a \equiv \sigma_B \boldsymbol{j}_a \cdot \mathbf{S}, \quad (6)$$

$$\beta \equiv \sigma_B \boldsymbol{b} \cdot \mathbf{S}, \quad (7)$$

and parameters,

$$R_a \equiv \boldsymbol{j}_a \cdot \boldsymbol{b}, \quad (8)$$

$$W_{a,b} \equiv \boldsymbol{j}_a \cdot \boldsymbol{j}_b, \quad (9)$$

$$Y_{a,b} \equiv \boldsymbol{j}_{a,\perp} \cdot \boldsymbol{j}_{b,\perp}. \quad (10)$$

The variables depend explicitly on the information $\{\sigma_B, \mathbf{S}\}$, whereas the parameters depend on the internal representations $\{\{\mathbf{J}_a\}, \mathbf{B}\}$ only. The variable $\beta \geqslant 0$ and the smaller the $\beta(\mathbf{S})$, the higher the likelihood of $\mathbf{S}$ to be a *socially neutral issue*. (An issue $\mathbf{S}_0$ is dubbed socially neutral if there is no socially accepted position about it, i.e., $\mathbf{B} \cdot \mathbf{S}_0 = 0$). The variable $\phi_a(\mathbf{S})$ indicates how much vector $\mathbf{J}_a$ has to be modified to agree with $\mathbf{B}$. If $\phi_a \gg 0$ the modification needed is negligible, if $\phi_a \simeq 0$ the modification needed is moderated, and if $\phi_a \ll 0$, the modification needed is substantial. The parameter R_a represents the level of agreement of agent a with the society B, $W_{a,b}$ represents the level of agreement between agents a and b, and the parameter $Y_{a,b}$ represents the level of agreement

between agents a and b on socially neutral issues. Given that $W_{a,b} = R_a R_b + Y_{a,b}\sqrt{(1 - R_a^2)(1 - R_b^2)}$ we only need to know $\{R_a\}$ and $\{Y_{a,b}\}$ to know the state of the society.

B. Update equations

Given a graph $\boldsymbol{G} = \{\{a\}, \{g_{a,b}\}\}$ with vertices $\{a\}$ and bonds $\{g_{a,b}\}$, the state of the society can be described by the sets of parameters $\{R_a\}$, defined on the vertices and $\{Y_{a,b}\}$, defined on the bonds of $\boldsymbol{G}$. The data accessible to agent a are $(\sigma_B, \phi_a, \phi_{\mathbb{N}_a}, \mathbf{S})$ where $\phi_{\mathbb{N}_a} \equiv \{\phi_c | c \in \mathbb{N}_a\}$. The length of such a training set is $T = \alpha_{\max}N$, which implies that $\alpha_{\max} = T/N$. For a given number $1 \leqslant n < N$ of examples presented to the perceptrons there is a $0 < \alpha < \alpha_{\max}$ such that $n = \alpha(n)N$. Observe that, given that the minimum increment in the number of examples presented is 1, $\Delta\alpha(n) \equiv \alpha(n + 1) - \alpha(n) = 1/N$. By defining $\Delta t \equiv f \Delta\alpha = f/N$ and by using the update rules (2) and (5), we have that the equation for the evolution of the parameters R_a and $Y_{a,b}$ is as follows:

$$\frac{\Delta R_a}{\Delta t} = \Psi_a(\beta - R_a \phi_a) + \lambda\sqrt{1 - R_a^2}\overline{Y}_a + O(\Delta t), \quad (11)$$

$$\frac{\Delta Y_{a,b}}{\Delta t} = \frac{\Psi_a}{\sqrt{1 - R_a^2}}\left[\frac{\phi_b - R_b \beta}{\sqrt{1 - R_b^2}} - Y_{a,b}\frac{\phi_a - R_a \beta}{\sqrt{1 - R_a^2}}\right]$$
$$- \frac{\lambda R_a}{\sqrt{1 - R_a^2}}(\overline{Y}_b - Y_{a,b}\overline{Y}_a) + \mathrm{IT}_{b,a} + O(\Delta t), \quad (12)$$

where $\overline{Y}_a \equiv M^{-1}\sum_c Y_{a,c}$ and $\mathrm{IT}_{b,a}$ represents a set of terms, identical to the previous ones in (12) with the indices a and b interchanged.

C. Large system size limit: Differential equations

For sufficiently large N and sufficiently small f, the divided difference equations (11) and (12) can be transformed into differential equations. It also occurs that the components of a social issue $\mathbf{S}$ are independent and identically distributed variables with $\mathcal{P}(\mathbf{S}) = \prod_{j=1}^{N}[\frac{1}{2}\delta_{S_j,1} + \frac{1}{2}\delta_{S_j,-1}]$ where $\delta_{S,X} = 1$ if $S = X$ and 0 otherwise is the Kronecker δ. This stochastic character is inherited by the variables β and $\{\phi_a\}$, whose joint probability in the large N limit can be estimated. In particular, for a society with only two voters we have that the joint probability is given by

$$\mathcal{P}(\beta, \phi_a, \phi_b) = \mathcal{N}\big(\beta | \Sigma_{a,b}\Lambda(\phi_a, \phi_b), \Sigma_{a,b}^2\big)\mathcal{N}$$
$$\times \big(\phi_b | W_{a,b}\phi_a, 1 - W_{a,b}^2\big)\mathcal{N}(\phi_a), \quad (13)$$

where $\mathcal{N}(x|\mu, \sigma^2) \equiv \exp[-(x - \mu)^2/2\sigma^2]/\sqrt{2\pi\sigma^2}$ is a Gaussian distribution on x, centered at μ, with variance $\sigma^2, \mathcal{N}(x) = \mathcal{N}(x|0, 1)$, and

$$\Sigma_{a,b}^2 \equiv \frac{(1 - R_a^2)(1 - R_b^2)(1 - Y_{a,b}^2)}{1 - W_{a,b}^2}, \quad (14)$$

$$\Lambda(\phi_a, \phi_b) \equiv \frac{(R_a - W_{a,b}\phi_b)\phi_a + (R_b - W_{a,b}R_a)\phi_b}{\sqrt{(1 - R_a^2)(1 - R_b^2)(1 - Y_{a,b}^2)(1 - W_{a,b}^2)}}. \quad (15)$$

The full derivation of these expressions is presented in the appendix of Ref. [1]. For larger communities the joint probability is much harder to obtain, but it can be estimated by considering an independent bond approximation (also presented in Ref. [1]). Finally we have to consider the distribution of the parameters $\{R_a\}$ and $\{Y_{a,b}\}$. Following Ref. [9], these parameters can be proved to be self-averaging in the large N limit, i.e., $\lim_{N\to\infty}\langle R_a \rangle = R_a$ and $\lim_{N\to\infty}\langle Y_{a,b}\rangle = Y_{a,b}$, where $\langle x \rangle = \int dx\, \mathcal{P}(x)x$. This is a consequence from the fact that for most cases the distribution of the parameters, inferred from the distribution of the issues **S**, is such that the variances $\sigma_R^2 \equiv \langle (R_a - \langle R_a \rangle)^2 \rangle$ and $\sigma_Y^2 \equiv \langle (Y_{a,b} - \langle Y_{a,b}\rangle)^2\rangle$ satisfy $\lim_{N\to\infty}\sigma_R^2 = \lim_{N\to\infty}\sigma_Y^2 = 0$.

Dimer

In the large N limit and for a population of only two voters, we obtain the following equations:

$$\dot{R}_a = \frac{2 - \eta_{a,b}}{2}\left(1 - R_a^2\right) + \frac{\eta_{a,b}}{2}\left[\left(1 - R_a^2\right)\frac{\arccos(Y_{a,b})}{\pi}\right.$$
$$\left. + \rho_{a,b}(R_b - W_{a,b}R_a)\right] + \lambda\sqrt{1 - R_a^2}\frac{1 + Y_{a,b}}{2}, \quad (16)$$

$$\dot{R}_b = \frac{2 - \eta_{b,a}}{2}\left(1 - R_b^2\right) + \frac{\eta_{b,a}}{2}\left[\left(1 - R_b^2\right)\frac{\arccos(Y_{a,b})}{\pi}\right.$$
$$\left. + \rho_{b,a}(R_a - W_{a,b}R_b)\right] + \lambda\sqrt{1 - R_b^2}\frac{1 + Y_{a,b}}{2}, \quad (17)$$

$$\dot{Y}_{a,b} = \frac{1 - Y_{a,b}^2}{2}\left[\frac{2\sqrt{1 - R_b^2}\,\eta_{a,b}\rho_{a,b} - \lambda R_a}{\sqrt{1 - R_a^2}} + \mathrm{IT}_{b,a}\right], \quad (18)$$

where $\eta_{c,d} \equiv \lim_{f\to 0} f^{-1}g_{c,d}$ and

$$\rho_{a,b} \equiv \frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{R_a - W_{a,b}R_b}{\sqrt{\left(1 - R_a^2\right)\left(1 - R_b^2\right)\left(1 - Y_{a,b}^2\right)}}\right). \quad (19)$$

Supposing without loss of generality that $\eta_{a,b} < \eta_{b,a}$, we found that there are only two stable points in the phase space of the system defined over the domain $\mathcal{S} \equiv \{(R_a, R_b, Y_{a,b}): R_a, R_b, Y_{a,b} \in [-1,1]\}$, namely, $\boldsymbol{C}^\star = (1, 1, -1)$, the conservative point, and $\boldsymbol{L}^\star = (-R_r, -R_v, 1)$, the liberal point, where

$$R_r \equiv \sqrt{1 - \left(\frac{2\lambda}{\eta_{a,b} - 2}\right)^2}, \quad (20)$$

$$R_v \equiv \frac{A\sqrt{A^2 + C^2 - 4\lambda^2} - 2\lambda C}{A^2 + C^2}, \quad (21)$$

and

$$A \equiv \eta_{b,a}(1 + R_r) - 2, \quad (22)$$

$$C \equiv \frac{2\eta_{b,a}\lambda}{\eta_{a,b} - 2}. \quad (23)$$

For values of $\eta_{a,b} \in [0, 2(\lambda + 1)]$ and $\eta_{b,a} > \eta_{a,b}$ we have that $\boldsymbol{C}^\star$ is the only stable point and for $2(1 + \lambda) < \eta_{a,b} < \eta_{b,a}$ both

$\boldsymbol{C}^\star$ and $\boldsymbol{L}^\star$ are stable. To demonstrate this statement we first observe that for $|Y_{a,b}| \simeq 1$ we can approximate

$$\rho_{a,b} \simeq \Theta\left(Y_{a,b}R_b\sqrt{1 - R_a^2} - R_a\sqrt{1 - R_b^2}\right), \quad (24)$$

$$\rho_{b,a} \simeq \Theta\left(Y_{a,b}R_a\sqrt{1 - R_b^2} - R_b\sqrt{1 - R_a^2}\right). \quad (25)$$

Close to the conservative point we can write $R_a = \cos\alpha$, $R_b = \cos\beta$, and $Y_{a,b} = \cos(\pi - y)$ for sufficiently small $\alpha, \beta, y > 0$. (Given that $\boldsymbol{C}^\star$ is a point in the border of the domain $\mathcal{S}$, $\alpha, \beta, y > 0$.) In such a case $Y_{a,b}R_a\sqrt{1 - R_b^2} - R_b\sqrt{1 - R_a^2} \simeq Y_{a,b}R_b\sqrt{1 - R_a^2} - R_a\sqrt{1 - R_b^2} \simeq -\alpha - \beta$ and $\rho_{a,b} \simeq \rho_{b,a} \simeq 0$. Therefore the system formed by Eqs. (16)–(18) can be expressed as

$$\dot{\alpha} \simeq -\alpha, \quad (26)$$

$$\dot{\beta} \simeq -\beta, \quad (27)$$

$$\dot{y} \simeq -\lambda\left(\frac{1}{\alpha} + \frac{1}{\beta}\right)\frac{y^2}{2}. \quad (28)$$

Thus α, β, and y decay exponentially.

If $2(1 + \lambda) < \eta_{a,b} < \eta_{b,a}$ the previous stability analysis is also valid, and $\boldsymbol{C}^\star$ is still a stable point of the system. For values of R_a and R_b sufficiently close to $-R_r$ and $-R_v$, we can propose $Y_{a,b} = \cos y$, $R_a = \cos(\pi - \theta_r + \alpha)$, and $R_b = \cos(\pi - \theta_v + \beta)$, where $\theta_r = \arccos(R_r)$, $\theta_v = \arccos(R_v)$ and with α, β, y sufficiently small. Observe that $\eta_{a,b} < \eta_{b,a}$ then $R_a > R_b$, $R_v > R_r$, and $\theta_r > \theta_v$ (see the Appendix). Then, $Y_{a,b}R_a\sqrt{1 - R_b^2} - R_b\sqrt{1 - R_a^2} \simeq -R_v\sqrt{1 - R_r^2} + R_r\sqrt{1 - R_v^2} = \sin(\theta_v - \theta_r)$ and $\rho_{a,b} \simeq 0$. Similarly, $\rho_{b,a} \simeq 1$, thus,

$$\dot{\alpha} \simeq -\frac{\eta_{a,b} - 2}{2}R_r\alpha, \quad (29)$$

$$\dot{\beta} \simeq \frac{\eta_{b,a}}{2}\cos(\theta_r - \theta_v)\alpha$$
$$- \left[\frac{\eta_{b,a} - 2}{2}R_v + \frac{\eta_{b,a}}{2}\cos(\theta_r - \theta_v)\right]\beta, \quad (30)$$

$$\dot{y} \simeq -\left[2\sqrt{\frac{1 - R_r^2}{1 - R_v^2}}\eta_{b,a} + \lambda\left(\frac{R_v}{\sqrt{1 - R_v^2}} + \frac{R_r}{\sqrt{1 - R_r^2}}\right)\right]\frac{y^2}{2}. \quad (31)$$

The coefficient on the right hand side of (29) is positive if $\eta_{a,b} > 2$. R_r is well defined if $\eta_{a,b} > 2(1 + \lambda)$ thus $\alpha \to 0$ exponentially. The factor in the second term on the right hand side of (30) is positive, then β also decays exponentially. Finally the right hand side of (31) is negative thus $Y_{a,b} \to 1$ from below.

These results can be condensed into an effective model defined by the following equations:

$$\dot{R}_a = 1 - R_a^2 + \frac{1 + Y_{a,b}^{\mathrm{eff}}}{2}\left[-\frac{\eta_{a,b}}{2}\left(1 - R_a^2\right)\right.$$
$$\left. + \left(\frac{\eta_{a,b}}{2}\Theta(R_b - R_a)\sin(\theta_a - \theta_b) + \lambda\right)\sqrt{1 - R_a^2}\right], \quad (32)$$
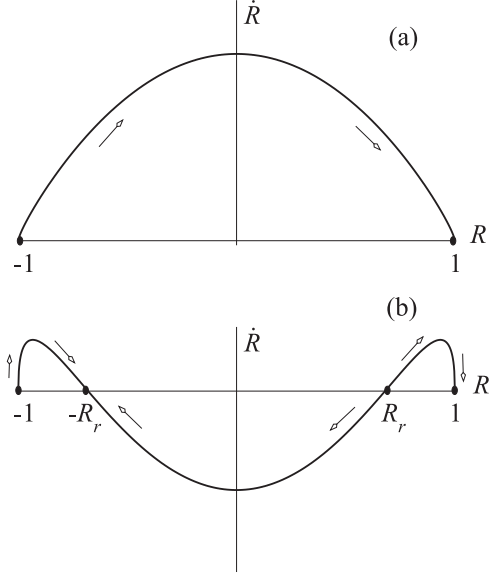
FIG. 1. $\dot{R}$ as a function of R when (a) $2\lambda > \eta - 2$ and (b) $2\lambda < \eta - 2$. In the first case, there is only on stable fixed point ($R = 1$), in the second case, there are two coexisting stable fixed points $R = -R_r$ [Eq. (20)] and $R = 1$.

$$\dot{R}_b = 1 - R_b^2 + \frac{1 + Y_{a,b}^{\text{eff}}}{2}\left[-\frac{\eta_{b,a}}{2}\left(1 - R_b^2\right)\right.$$
$$\left. + \left(\frac{\eta_{b,a}}{2}\Theta(R_a - R_b)\sin(\theta_b - \theta_a) + \lambda\right)\sqrt{1 - R_b^2}\right],$$
$$(33)$$

where $Y_{a,b}^{\text{eff}}$ is completely determined by the values of the social strengths and the initial conditions. Let us define $\eta_{\min} \equiv \{\eta_{a,b}, \eta_{b,a}\}$ and $R_{\min} = R_a$ if $\eta_{\min} = \eta_{a,b}$ and R_b otherwise. Then $Y_{a,b}^{\text{eff}} = 1$ if $\eta_{\min} > 2(1 + \lambda)$ and $R_{\min}(0) < R_r$. Observe that the presence of this effective bond $Y_{a,b}^{\text{eff}}$ is a direct consequence of B adapting to the average position of the dimer. The stability of this effective system is illustrated in Fig. 1. If $\eta_{\min} < 2(1 + \lambda)$, then there is only one stable point [Fig. 1(a)], whereas if $\eta_{\min} > 2(1 + \lambda)$, there are two stable points [Fig. 1(b)].

Following the definitions introduced in Ref. [1], we will dub an agent *a consensual* if it is within the basin of attraction of $C^\star$ and *polarizing* if it is not. In terms of the social strengths, an agent has an attitude that is *conservative* if $\eta_{a,b} < 2(1 + \lambda)$ and *liberal* otherwise.

D. Beyond the dimer

Following Ref. [1], we obtain for the vertices and bond variables of a given graph the following set of equations:

$$\dot{R}_a = \left(1 - \sum_{c \in \mathbb{N}_a}\frac{\eta_{a,c}}{2}\right)\left(1 - R_a^2\right) + \sum_{c \in \mathbb{N}_a}\frac{\eta_{a,c}}{2}\left[\left(1 - R_a^2\right)\frac{\varphi_{a,c}}{\pi}\right.$$
$$\left. + \rho_{a,c}(R_c - W_{a,c}R_a)\right] + \lambda\sqrt{1 - R_a^2}\overline{Y}_a, \quad (34)$$

$$\dot{Y}_{a,b} = \left(1 - Y_{a,b}^2\right)\sqrt{\frac{1 - R_b^2}{1 - R_a^2}}\eta_{a,b}\rho_{a,b}$$
$$- \frac{\lambda R_a}{\sqrt{1 - R_a^2}}(\overline{Y}_b - Y_{a,b}\overline{Y}_a) + \text{IT}_{b,a}. \quad (35)$$

With the insight gained from the dimer case, we assume that there exists at least two fixed points, one conservative, present for all values of the social strengths if $\lambda > 0$ and characterized by a positive value of the magnetization,

$$\mu \equiv \lim_{t \to \infty}\frac{1}{M}\sum_a R_a(t), \quad (36)$$

and one liberal, present for sufficiently large values of the social strengths and characterized by a negative value of the magnetization.

If we impose the homogeneity condition,

$$\overline{Y}_a = \overline{Y}_b = \overline{Y} \equiv \frac{1}{M^2}\sum_{a,b}Y_{a,b}, \quad (37)$$

which we expect to be valid for sufficiently large societies (i.e., $1 \ll M$) with low connectivity (i.e., if the size of a typical neighborhood is a quantity of order 1) and considering that most of the terms in (37) come from passive bonds ($\eta_{a,b} = \eta_{b,a} = 0$),

$$\overline{Y} = \frac{1}{M}\sum_a\left\{\frac{1}{M}\sum_{b \in \mathbb{N}_a}Y_{a,b} + \frac{1}{M}\sum_{b \notin \mathbb{N}_a}Y_{a,b}\right\}$$
$$= \frac{1}{M^2}\sum_a\sum_{b \notin \mathbb{N}_a}Y_{a,b} + O(M^{-1}),$$

we conclude that

$$\dot{\overline{Y}} = -2\lambda\overline{K}\,\overline{Y}(1 - \overline{Y}) + O(M^{-1}), \quad (38)$$

where

$$\overline{K} \equiv \frac{1}{M}\sum_a\frac{R_a}{\sqrt{1 - R_a^2}}. \quad (39)$$

Observe that there are two possible stable values for $\overline{Y}$, satisfying $\overline{Y}^\star = \Theta(-\overline{K})$. We expect $\overline{K} > 0(<0)$ if there is a sufficient number of consensual (polarizing) agents with $R_a \to 1(-1)$. This estimate implies that, if $\overline{K} > 0$, we are in a conservative phase where the average position within a neighborhood of voters differs to the average position of a different neighborhood on neutral social issues. Collectively, this adds up to a $\overline{Y} = 0$. If $\overline{K} < 0$ we are in a liberal phase where different neighborhoods agree on socially neutral issues, producing an effective $\overline{Y} = 1$. In these circumstances, Eq. (35) for active bonds is transformed into

$$\dot{Y}_{a,b} \simeq \left[\sqrt{\frac{1 - R_b^2}{1 - R_a^2}}\eta_{a,b}\rho_{a,b}(1 + Y_{a,b})\right.$$
$$\left. - \lambda\frac{R_a}{\sqrt{1 - R_a^2}}\Theta(-\overline{K}) + \text{IT}_{b,a}\right](1 - Y_{a,b}). \quad (40)$$

It is straightforward to conclude that, if $\overline{K} > 0$ or, if $\overline{K} \ll 0$ (where most of the agents are polarizing, i.e., $R_c < 0$)

$Y_{a,b} \to 1$. Equation (40) and the correspondent (34) become very difficult to analyze for intermediate values of $\overline{K} \in (-K_{\min}, 0)$, which occur for values of the social strengths such that $\nu\overline{\eta} \simeq 2(1 + \lambda)$, where ν is the average neighborhood size (1) and $\overline{\eta} = \sum_{\eta_{a,b} > 0} \eta_{a,b} / \sum_{\eta_{a,b} > 0} 1$ is the average social strength considering only active bonds. Preliminary numerical experiments suggest that for these values of $\overline{\eta}$

the final states of the system are extremely sensitive to the initial conditions. Such an analysis will be left for a future work.

By using that for the set of values of the social strengths for which the system is tractable the values of $Y_{a,b}$ correspondent to active bonds converge to one, we propose the following equation for the set $\{R_a\}$:

$$\dot{R}_a = \left(1 - \sum_{c \in \mathbb{N}_a} \frac{\eta_{a,c}}{2}\right)(1 - R_a^2) + \left[\sum_{c \in \mathbb{N}_a} \frac{\eta_{a,c}}{2} \Theta(R_c - R_a)\sin(\theta_a - \theta_c) + \lambda\right]\sqrt{1 - R_a^2}. \tag{41}$$

Let us define the effective social strength on a by its neighbors by the expression,

$$H_a \equiv \sum_{c \in \mathbb{N}_a} \eta_{a,c}. \tag{42}$$

Let us assume that if $H_a < H_b$ and after a sufficiently long time $0 \ll t_0$ $R_a \gtrsim R_b$. With this assumption in mind and without loss of generality, let us rearrange the labels in such a way that $H_1 < H_2 < \cdots < H_M$. Then we have that for a time $t > t_0$ there are a number of agents with equations of the form

$$\dot{R}_n = \left(1 - \frac{H_n}{2}\right)(1 - R_n^2) + \lambda\sqrt{1 - R_n^2}, \tag{43}$$

where all agents (n), whose equations of motion do not depend on the variables associated with other agents, are characterized by having constants $H_n < H_{n_i}$ where $n_i \in \mathbb{N}_n$, so n has a more conservative attitude than its neighbors. We will call such agents the *nucleating centers*. The dynamic behavior of the nucleating centers depends on the value of the constant H_n and [if $H_n > 2(1 + \lambda)$] of the value of R_n at time t_0 where the order $R_n > R_{n_1}$ for all $n_1 \in \mathbb{N}_n$ is established. To support this claim, let us integrate (43),

$$t - t_0 = -2 \int_{\theta_0}^{\theta} \frac{dx}{2\lambda + (2 - H_n)\sin x}. \tag{44}$$

If $H < 2(1 + \lambda)$ there is only one fixed point, $R^\star = 1$, reached from any initial condition $\theta_0 \in (0, \pi)$. There are three possible situations in this phase,

$$t - t_0 = \begin{cases} \frac{4}{\sqrt{(2-H_n)^2 - 4\lambda^2}}\left[\text{arctanh}\left(\frac{\sqrt{(2-H_n)^2 - 4\lambda^2}}{2 - H_n}\right) - \text{arctanh}\left(\frac{\sqrt{(2-H_n)^2 - 4\lambda^2}}{2\lambda \tan(\theta_0/2) + 2 - H_n}\right)\right], & H_n < 2(1 - \lambda), \\ \frac{2}{\lambda}\frac{\tan(\theta_0/2)}{\tan(\theta_0/2) + 1}, & H_n = 2(1 - \lambda), \\ \frac{4}{\sqrt{4\lambda^2 - (2-H_n)^2}}\left[\arctan\left(\frac{2\lambda \tan(\theta_0/2) + 2 - H_n}{\sqrt{4\lambda^2 - (2-H_n)^2}}\right) - \arctan\left(\frac{2 - H_n}{\sqrt{4\lambda^2 - (2-H_n)^2}}\right)\right], & 2(1 - \lambda) < H_n < 2(1 + \lambda), \end{cases} \tag{45}$$

but in all cases $t - t_0 < \infty$. At the critical point we have that $H_n = 2(1 + \lambda)$ in which case we have that the stable points are $R_n^\star = \cos(\theta^\star)$ with $\theta^\star = 0, \pi$ and basins of attraction $(0, \pi/2)$ and $(\pi/2, \pi)$, respectively,

$$t - t_0 = \begin{cases} \frac{2}{\lambda}\frac{\tan(\theta_0/2)}{\tan(\theta_0/2) + 1}, & \theta_0 \in \left(0, \frac{\pi}{2}\right), \\ \lim_{\theta \to 0}\frac{1}{\theta}, & \theta_0 \in \left(\frac{\pi}{2}, \pi\right). \end{cases} \tag{46}$$

Finally, for $H_n > 2(1 + \lambda)$ we have that the two stable points are $R_n^\star = \cos(\theta^\star)$ with $\theta^\star = 0, \pi - \theta_r$, where

$$\theta_r \equiv \arccos\left(\sqrt{1 - \left(\frac{2\lambda}{H_n - 2}\right)^2}\right), \tag{47}$$

and basins of attraction $(0, \theta_r)$ and (θ_r, π), respectively,

$$t - t_0 = \begin{cases} \frac{4}{\sqrt{(H_n-2)^2 - 4\lambda^2}}\left[\text{arctanh}\left(\frac{\sqrt{(H_n-2)^2 - 4\lambda^2}}{H_n - 2 - 2\lambda \tan(\theta_0/2)}\right) - \text{arctanh}\left(\frac{\sqrt{(H_n-2)^2 - 4\lambda^2}}{H_n - 2}\right)\right], & \theta_0 \in (0, \theta_r), \\ \lim_{\theta \to 0}\ln\left(\frac{1}{\theta}\right), & \theta_0 \in (\theta_r, \pi). \end{cases} \tag{48}$$

Those we observe for all initial conditions belonging to the basin of attraction of the conservative point $R_n^\star = 1$ the duration of the opinion formation process is finite. The time the process lasts to reach the liberal position $R_n^\star = \cos(\pi - \theta_r)$ when $H_n > 2(1 + \lambda)$ and $\theta_0 \in (\theta_r, \pi)$ is infinite.

Suppose agent (o) is a nucleating center, suppose also that the initial condition $R_o(t_0) \in (\cos \theta_r, 1)$ (conservative basin of attraction) and the current time t is sufficiently larger than t_0 to ensure convergence to $R_o^\star = 1$. The element of $o_1 \in \mathbb{N}_o$ with the lowest effective social strength H_{o_1} is ruled by the equation,

$$\dot{R}_{o_1} = \left(1 - \frac{H_{o_1} - \eta_{o_1,o}}{2}\right)(1 - R_{o_1}^2) + \lambda\sqrt{1 - R_{o_1}^2}, \quad (49)$$

which is identical to (43) and can be analyzed in the same way. Also observe that the presence of a consensual agent increases the chances of producing new nucleating centers due to the reduction of the effective social strength of its neighbors as is observed in the first term of the right hand side of (49). This nucleating effect has been observed previously, close to the critical value of the social strength, i.e., $H \sim 2(1 + \lambda)$. If the effective strengths H_{o_L} of agents located at a distance L of the nucleating center are sufficiently large, it may occur that $R_{o_L}^\star < 0$ and a border of polarizing agents emerge around a cluster of consensual agents. Such clusters have been observed emerging in the numerical experiments presented in Ref. [1]. In the following section we will explore the distribution of cluster sizes as a function of the strengths $\{\eta_{a,b}\}$ in the square lattice.

II. ESTIMATED OBSERVABLES AND EXPERIMENTS IN THE SQUARE LATTICE

We consider the case of the agents sitting in the vertices of a square lattice and where the $\eta_{a,c}$'s are drawn from a Gaussian distribution centered at η and with variance Δ^2. η, the average social strength, represents the average influence neighbors have on each other, and in an indirect form it also represents a level of discontent with B. The parameter Δ controls the level of variation, or disorder, in the set of social strengths. In the following we present a theoretical analysis based on the distribution of social strengths, disregarding the initial conditions of the variables $\{R_a(0)\}$.

On a square lattice each agent is linked to precisely four neighbors. The expression,

$$\mathbb{P}_0 := \int_{-\infty}^{2(1+\lambda)} \frac{dx}{\sqrt{2\pi(2\Delta)^2}} \exp\left\{-\frac{1}{2}\left(\frac{x - 4\eta}{2\Delta}\right)^2\right\}$$
$$= \mathcal{H}\left(\frac{4\eta - 2(1+\lambda)}{2\Delta}\right), \quad (50)$$

where $\mathcal{H}(x) \equiv \int_x^\infty dy\, \mathcal{N}(y)$ is the Gardner error function represents the probability that a vertex o has a conservative attitude $2(1 + \lambda) > \sum_{c \in \mathbb{N}_o} \eta_{o,c}$ in which case for times not smaller than a sufficiently large t_0, $R_o > R_b$ for all $b \in \mathbb{N}_o$. At a link away from o there are the neighbors $\mathbb{N}_o = \{o_1, o_2, o_3, o_4\}$. The probability that these sites have a conservative attitude is in similar manner and given that the center o

is consensual,

$$\mathbb{P}_{1,0} := \mathcal{H}\left(\frac{3\eta - 2(1+\lambda)}{\sqrt{3}\Delta}\right). \quad (51)$$

The vertices of the square that complete the first layer around o have only two neighbors each that may not be consensual thus,

$$\mathbb{P}_{1,1} := \mathcal{H}\left(\frac{2\eta - 2(1+\lambda)}{\sqrt{2}\Delta}\right), \quad (52)$$

thus the probability of having a consensual cluster with only one layer of agents around the nucleating center is

$$\mathbb{P}_1 = \mathbb{P}_0 \mathbb{P}_{1,0}^4 \mathbb{P}_{1,1}^4. \quad (53)$$

By repeating this process L times, the probability of having a consensual cluster with L layers is

$$\mathbb{P}_L(\eta, \Delta) = \mathbb{P}_0(\eta, \Delta)[\mathbb{P}_{1,0}(\eta, \Delta)]^{4L}[\mathbb{P}_{1,1}(\eta, \Delta)]^{4L^2}. \quad (54)$$

In order to measure the distribution of clusters in a square lattice of side $\sqrt{M}$, we propose the following quantities:

$$\langle L \rangle \equiv \frac{\sum_{j=1}^{\sqrt{M}/2} j \mathbb{P}_j(\eta, \Delta)}{\sum_{j=1}^{\sqrt{M}/2} \mathbb{P}_j(\eta, \Delta)}, \quad (55)$$

$$\sigma_L \equiv \sqrt{\langle L^2 \rangle - \langle L \rangle^2}. \quad (56)$$

$\langle L \rangle$ is the estimated size of the cluster for the given values of the social strength and the disorder parameter, and σ_L its standard deviation.

Consider now the following observables:

$$\mu(\eta, \Delta) \equiv \frac{1}{M} \sum_a R_a(t_{\max}), \quad (57)$$

the magnetization, which measures the average social agreement with B, and the correlations defined as

$$C_\ell(\eta, \Delta) = \frac{1}{M} \frac{1}{8\ell} \sum_a \sum_{b \in \mathbb{N}_a(\ell)} R_a(t_{\max}) R_b(t_{\max}) - \mu^2(\eta, \Delta),$$
$$(58)$$

where $\mathbb{N}_a(\ell)$ is the set of agents located in the ℓth shell of agent a. The behavior of the correlation can be modeled by the expression,

$$C_\ell(\eta, \Delta) = C_0(\eta, \Delta) \exp\left(-\frac{\ell}{\xi(\eta, \Delta)}\right), \quad (59)$$

where $\xi(\eta, \Delta)$ is the correlation length and $C_0(\eta, \Delta)$ is the susceptibility.

To better understand the behavior of a system described by Eq. (41), we perform a number of numerical experiments on a square grid where we placed the member of a society of size $M = 10^4$ with periodic boundary conditions. The agents are connected to their first neighbors only with social strengths $\{\eta_{ab}\}$ extracted from a Gaussian distribution with η mean and variance Δ^2. Observe that, although the $\{\eta_{a,b}\}$'s so produced are allowed to have negative values, the values of η and Δ used are such that the chances of a negative social strength occurring are negligible. The social update constant was kept constant at $\lambda = 0.1$. The evolution of the agents'
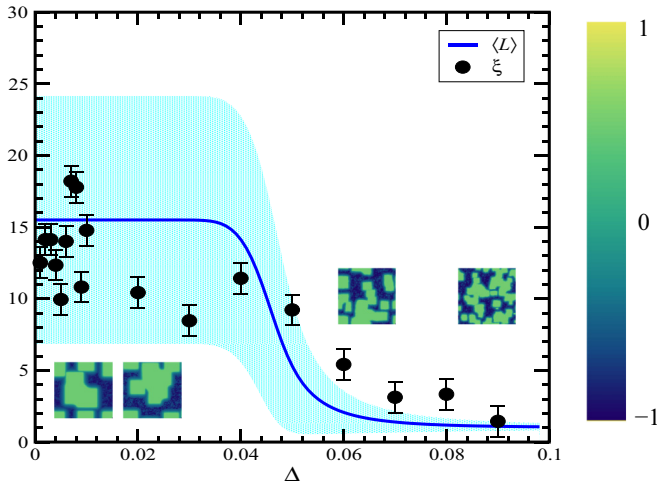
FIG. 2. Expected number of layers in the typical cluster $\langle L \rangle \pm \sigma_L$ (full line) within one standard deviation (light shadow) and correlation lengths ξ [full circles, averaged over ten independent numerical integrations of (41)] as functions of the disorder Δ for values of the social strength $\eta^\star(\Delta)$, $\lambda = 0.1$ for a society of $M = 10\,000$ agents arranged on a square lattice with periodic boundary conditions. In the insets we have snapshots of the system at, from left to right, $\Delta = 0.001$, 0.005, 0.06, and 0.09, respectively. The pixel at position (i, j), $1 \leqslant i, \ j \leqslant 100$ in the insets represents the value of $R_{i,j}$ with values between -1 and 1 according to the provided scale.

agreement with B is obtained from the system (41), which was numerically integrated using a second order Runge-Kutta method for a maximum time $t_{\max} = 100$ time units with initial conditions uniformly distributed in $[-1, 1]$ and for a variety of values for the parameters η and Δ. For each $\Delta = 0.001$, 0.01 in increments of 0.001 and from 0.01 to 0.09 in increments of 0.01 we found the value of $\eta^\star(\Delta)$ that satisfies the equation $\mu(\eta^\star, \Delta) = 0$. For those values of $[\eta^\star(\Delta), \Delta]$ we computed $\langle L \rangle$ (55) and σ_L (56) and estimated $\xi[\eta^\star(\Delta), \Delta]$ averaged over ten independent runs of the Runge-Kutta integration process. The results of these experiences are presented in Fig. 2.

III. CONCLUSIONS

We proposed a model of opinion formation in societies of adaptive agents where there is a set of rules B that determined what is socially acceptable. In the present paper we allow B to adjust according to the average position of the population with a constant of proportionality λ. By the application of statistical mechanics techniques we constructed a description of the system's behavior based on a set of differential equations ruling the evolution of the parameters $\{R_a\}$, that represent the agreement of the agents $\{a\}$ with B, and $\{Y_{ab}\}$ that represent the agreement between two connected agents a and b on neutral issues (i.e., issues for which B has no opinion). For the case of a society with only two individuals, the system can be described by Eqs. (16)–(18). For this system there are only two stable fixed points, dubbed the *conservative point* $C^\star$ with coordinates $R_a^\star = R_b^\star = -Y_{a,b}^\star = 1$ and the *liberal point* $L^\star$ with coordinates $R_{\max}^\star = -R_r$, $R_{\min}^\star = -R_v$, and $Y_{a,b}^\star = 1$, where $R_{\min(\max)} \equiv \min(\max)\{R_a, R_b\}$ and R_r and R_v are given by (20) and (21), respectively. We observe that in

the case of $\lambda > 0$ and opposite to the case reported in Ref. [1] where $\lambda = 0$, consensual agents converge to the position of B, represented by $C^\star$ with opposite positions on neutral issues. This effect is due to the action of B to adapt to the average position of a and b. The vector obtained by averaging the internal representations of a, and b is a coplanar vector bisecting the angle subtended between $\mathbf{J}_a$ and $\mathbf{J}_b$. If the internal representation of B is modified, it eventually positions itself in between $\mathbf{J}_a$ and $\mathbf{J}_b$ and, given that the plane perpendicular to the internal representation $\mathbf{B}$ has also changed, $\mathbf{J}_{a,\perp}$ and $\mathbf{J}_{b,\perp}$ become collinear and opposite. Thus, the effective action of B adapting to the average position of the voters favors a consensus constructed over the polarization of the agents' positions on the plane of neutral issues. Only when the social strengths are sufficiently high, both agents become polarized with respect to B and with equal positions on socially neutral issues. In this case, the adaptation of B is only sufficient to reduce the final position of a and b to $-R_r > -R_v > -1$ (remember that in the case $\lambda = 0$, $R_a = R_b = -1$).

For a larger population $M > 2$, the equations obtained by averaging the divided difference equations (11) and (12) using the probability distribution from the independent bond approximation are Eqs. (34) and (35). We observed that, in the homogeneous approximation (37), the average $\overline{Y}$ is constructed mostly with terms that correspond to passive bonds. Such a quantity is ruled by a differential equation with two fixed points $\overline{Y}^\star = 0, 1$. The fixed point $\overline{Y}^\star = 0$ appears for sufficiently low values of the social strengths. In this phase, it is also observed that the connected bond parameters converge to $Y_{a,b}^\star = 1$. Thus, in this conservative phase, the effect of the adaptation of B to the average position of the population produces an overall disagreement on socially neutral issues between neighborhoods that are mutually disconnected, producing $\overline{Y} = 0$. For large enough values of the social strengths, we observe an overall agreement among the agents that adds up to $\overline{Y} = 1$. The model as it stands from (34) and (35) becomes difficult to analyze for intermediate values of the social strengths $\nu\eta \sim 2(1 + \lambda)$. This problem arises from the construction of the joint probability $\mathcal{P}(\beta, \{\phi_{a,}\})$, which is an approximation based on independent bonds. In this approximation, the effect of neighbors is overestimated, producing differential equations with no stable steady points close to the boundary between the liberal and the conservative phases. To overcome this problem we considered the effective model presented in Eq. (41) where the values of $\{Y_{a,b}\}$ for active bonds are set to 1. Such an equation is not obtained by averaging (11) with a suitable joint probability but by generalizing the effective system of equations for the dimer [(32) and (33)].

The effective model presented in Eq. (41) also admits one stable point (conservative) for sufficiently low values of the social strengths and two stable points (one conservative and one liberal) for sufficiently large values of them. One of the main characteristics of this model is probably that the length of the learning process (i.e., the time required to reach a fix point in the phase space) is finite if the initial conditions are in the conservative basin of attraction [Eqs. (45)–(48)].

By considering the social strengths to be drawn from a Gaussian distribution with mean η and variance Δ^2 for a population of interactive agents placed on a square lattice,

we estimated the expected size of the conservative clusters $\langle L \rangle$ and its standard deviation σ_L, Eqs. (55) and (56), to be formed around nucleating centers as defined by (43). We contrasted this quantity with the correlation length (59) obtained from numerical integrations of (41) at the point of transition between the conservative (low average social strength) and liberal (high average social strength) phases, defined by $\mu(\eta, \Delta) = 0$. The results, presented in Fig. 2, show that these quantities are strongly correlated.

The most relevant effect observed, the emergence of conservative clusters when the average importance to the peers' opinions is increased, has a clear interpretation in the context of opinion formation. Let us assume we live in a society where the *status quo B* is well established. Suppose there is evidence in support of an action against the established order, and in consequence a policy is made to challenge B. Such evidence may produce a change in attitude in the social members, trying to corroborate their opinions by contrasting them with their peers (increase in η). Such corroboration is not sought when there is general conformity with B. Members of the society that remain in agreement with B have the effect of *leaders* [10] and conservative clusters emerge and remain, even if η is increased by the emergence of more evidence in favor of the challenging policy. It has been observed that the effect of a positive λ for large enough values of the social strengths is the stabilization of the conservative fixed point $R = 1$. The analysis of a bistable system under periodic perturbations [11] will be presented in a following article.

APPENDIX: DEMONSTRATION OF THE ORDER ASSOCIATED WITH THE SOCIAL STRENGTHS

Observe that our analysis is based on the supposition that, if $\eta_{a,b} < \eta_{b,a}$, then at the stable point and in a neighborhood of it we have that $R_b < R_a$. Observe that by the definitions of A and C [Eqs. (22) and (23), respectively] we have that

$$\eta_{b,a} > \eta_{a,b} > 2(1 + \lambda),$$

$$\eta_{b,a} R_r + \eta_{b,a} - 2 > \eta_{b,a} R_r + \eta_{a,b} - 2 > 0,$$

$$\left(\frac{2\lambda}{\eta_{a,b} - 2}\right)^2 (\eta_{b,a} R_r + \eta_{b,a} - 2)^2 > \left(\frac{2\lambda}{\eta_{a,b} - 2}\right)^2 (\eta_{b,a} R_r + \eta_{a,b} - 2)^2,$$

$$\left(\frac{2\lambda}{\eta_{a,b} - 2}\right)^2 A^2 > (C R_r + 2\lambda)^2,$$

$$(1 - R_r^2) A^2 > C^2 R_r^2 + 4\lambda C R_r + 4\lambda^2,$$

$$(A^2 + C^2) A^2 > (A^2 + C^2)[(A^2 + C^2) R_r^2 + 4\lambda C R_r + 4\lambda^2],$$

$$(A^2 + C^2 - 4\lambda^2) A^2 > (A^2 + C^2)^2 R_r^2 + 4(A^2 + C^2)\lambda C R_r + 4\lambda^2 C^2,$$

$$(A^2 + C^2 - 4\lambda^2) A^2 > [(A^2 + C^2) R_r + 2\lambda C]^2,$$

$$\sqrt{A^2(A^2 + C^2 - 4\lambda^2)} > (A^2 + C^2) R_r + 2\lambda C,$$

$$\lim_{t \to \infty} R_a(t) = -R_r > \frac{2\lambda C - \sqrt{A^2(A^2 + C^2 - 4\lambda^2)}}{A^2 + C^2} = -R_v = \lim_{t \to \infty} R_b(t) \ \square.$$

Thus, for a sufficiently large t, we have that $R_a(t) > R_b(t)$ due to the continuity of these functions.

[1] J. Neirotti, Phys. Rev. E **94**, 012309 (2016).
[2] J. Neirotti, Phys. Rev. E **95**, 062305 (2017).
[3] A. Engel and C. Van den Broeck, *Statistical Mechanics of Learning* (Cambridge University Press, Cambridge, UK, 2001).
[4] D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).
[5] C. G. Lord, L. Ross, and M. R. Lepper, J. Personality Soc. Psychol. **37**, 2098 (1979).
[6] R. S. Baron *et al.*, J. Exp. Soc. Psychol. **32**, 537 (1996).
[7] E. Gilbert, T. Bergstrom, and K. Karahalios, in *Proceedings of the 42nd Hawaii International Conference on System Sciences, Big Island, HI, 2009*, edited by R. J. Sprague (IEEE Computer Society, Washington DC, 2009), p. 1.
[8] N. Caticha and O. Kinouchi, Philos. Mag. B **77**, 1565 (1998).
[9] G. Reents and R. Urbanczik, Phys. Rev. Lett. **80**, 5445 (1998).
[10] K. Kacperski and J. A. Holyst, J. Stat. Phys. **84**, 169 (1996).
[11] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, Rev. Mod. Phys. **70**, 223 (1998).