# Anisotropic opinion dynamics

Juan Neirotti

*Department of Mathematics, Aston University,*

*The Aston Triangle, B4 7ET, Birmingham, UK*

## Abstract

We consider the process of opinion formation in a society of interacting agents, where there is a set $B$ of socially accepted rules. In this scenario, we observed that agents, represented by simple feed-forward, adaptive neural networks, may have a conservative attitude (mostly in agreement with $B$) or liberal attitude (mostly in agreement with neighboring agents) depending on how much their opinions are influenced by their peers. The topology of the network representing the interaction of the society's members is determined by a graph, where the agents' properties are defined over the vertexes and the inter-agent interactions are defined over the bonds. The adaptability of the agents allows us to model the formation of opinions as an on-line learning process, where agents learn continuously as new information becomes available to the whole society (on-line learning). Through the application of statistical mechanics techniques we deduced a set of differential equations describing the dynamics of the system. We observed that by slowly varying the average peer influence in such a way that the agents attitude changes from conservative to liberal and back, the average social opinion develops a hysteresis cycle. Such hysteretic behavior disappears when the variance of the social influence distribution is large enough. In all the cases studied, the change from conservative to liberal behavior is characterized by the emergence of conservative clusters, i.e. a closed knitted set of society members that follow a leader who agrees with the social *status quo* when the rule $B$ is challenged.

## I. INTRODUCTION

In this article we present a model for the process of opinion formation in rational individuals, interacting with a social circle of peers in a society with a pre-existent set of rules that determine what is socially acceptable. Our objective is to better understand observed phenomena such as the emergence and stability of opinion consensus [1–3], the social segregation through opinion polarization [4–6], hysteretic social behavior, opinion clustering and other phenomena related to properties of bi-stability of the opinion formation process [7–9], amongst others.

We propose a model with the following characteristics:

1. The model considers a mechanism for *rational* agents to assimilate information and update their opinions. The agents are explicitly adaptive.

2. The model considers the existence of a set of rules $B$ that determine what is socially acceptable. We assume the existence of a functioning society prior the beginning of the opinion formation process.

3. The model considers the interaction of the agents with their neighbors [10, 11], with a strength proportional to the credibility of the neighbors, their number and their proximity to the agent.

If by rational we mean Bayesian, two non-interacting agents should, if the information about a social issue is complete, reach the same opinion, disregarding their initial priors ([1], Theorem 1). Therefore, a model of the opinion formation process of rational agents in a functioning society must consider inter-agent interactions as means to support the emergence of different social positions.

The meta-agent $B$ determines what is socially acceptable by assigning one out of two possible values (good or bad) to social issues risen in the society. Modeling opinions with binary variables is consistent with the observation that most people opt for one out of two opposite positions while answering questions with high social impact ([7] and references therein). In mathematical terms, $B$ is a classifier that assigns labels $\sigma_B \in \{\pm 1\}$, to social instances codified in binary messages $\mathbf{S}$ of length $N$, i.e. $\mathbf{S} \in \{\pm 1\}^N$ [5]. In order to keep a balance between sophistication and analytical tractability, we choose to represent the meta-agent $B$ as a perceptron [12] with an internal representation $\mathbf{B} \in \mathbb{R}^N$, that classifies $N$-tuples $\mathbf{S}$ through the rule $\sigma_B(\mathbf{S}) = \mathrm{sgn}(\mathbf{B} \cdot \mathbf{S})$, where $\mathbf{B} \cdot \mathbf{S} \equiv \sum_{i=1}^{N} B_i S_i$ is the usual inner product in $\mathbb{R}^N$, $\mathrm{sgn}(x) = 1$ if $x > 0$, -1 if $x < 0$ and 0 if $x = 0$. By choosing a perceptron as a representative of $B$, we are singling out a particular direction in space, $\mathbf{B}$, which makes the process of opinion formation *anisotropic*.

By consistency we choose to model the agents $\{a\}$ by perceptrons with internal representations $\left\{\mathbf{J}_a \in \mathbb{R}^N\right\}$. The set of social rules $B$, which are assumed to be the product of local history, are considered to be constant, thus $\mathbf{B}$ does not change over time. $\mathbf{J}_a$, the agent's internal representation, is a plastic quantity that changes to assimilate new available information.

The use of perceptrons to model social agents has been done within the framework of moral foundation theory [13, 14]. Our approach differs from these in the fact that we consider the existence of a functioning society represented by $B$. By considering a supervised, on-line learning scenario [12] to adapt the agent-perceptrons to the information received from the society $B$ and from the other agents [15], we can mimic the opinion formation process in an environment where individual opinions are formed under local (through the neighborhood) and global (through the society) influences.

In order to construct the equations that rule the dynamics of the process and to make sense of the information provided by a large number of interacting agents, we use statistical mechanics techniques, which have been successfully applied in socially inspired problems for the last three decades ([16] and references therein).

## II. THE LEARNING ALGORITHM

By the very nature of society, social agents are influenced by others in possibly many different ways. In our model, we consider two levels of social pressure exerted on the individual agents: from the society as a whole, and from the agent's local neighborhood [10, 17, 18].

The organization of agents and neighborhoods can be modeled through 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 $a$. Clearly, there is no need to assume reciprocity ($g_{a,b} \neq g_{b,a}$) and $g_{a,b} \geq 0$ where $g_{a,b} = 0$ implies that agent $b$ is not in the neighborhood of $a$, $\mathbb{N}_a = \{c | g_{a,c} > 0\}$.

Given that the social agents are represented by perceptrons, we propose a supervised, on-line learning scenario, where agents learn continuously as new information becomes available [12, 19], to mimic the opinion formation process. The information $a$ receives is taken from the set $\mathbb{S} \equiv \{(\sigma_{B,n}, \sigma_{\mathbb{N}_a,n}, \mathbf{S}_n), n = 1, \ldots, T\}$, where the issue $\mathbf{S}_n$ is presented at time $n$ and then discarded (learning has to proceed on-line if there is only one training example available at any particular time), $\sigma_{B,n}$ is the opinion of $B$ on $\mathbf{S}_n$ (used as a reference to guide the learning process in the particular direction given by the *supervisor*'s vector $\mathbf{B}$) and $\sigma_{\mathbb{N}_a} \equiv \{\sigma_c | c \in \mathbb{N}_a\}$ is the set of opinions of the neighbors of $a$ on $\mathbf{S}$. The update equation for the internal representation of $a$ is:

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

where $\sigma_B \mathbf{S}/\sqrt{N}$ is the (unit length) Hebb vector, 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 is $T$. Based on social corroboration experiments [4, 20, 21] 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 [22] and it has been only considered for technical purposes and:

$$\Psi_a \equiv 1 - \Theta(-\sigma_B \sigma_a) \sum_{c \in \mathbb{N}_a} \frac{g_{a,c}}{f} \Theta(\sigma_a \sigma_c), \tag{2}$$

where $\Theta(x) = 1$ if $x > 0$ and 0 otherwise is the Heaviside step function. If the agent $a$ is not connected ($\mathbb{N}_a = \emptyset$), it only learns from the social rule $B$, thus $\mathbf{J}_a \to \mathbf{B}$. If $\mathbb{N}_a \neq \emptyset$, then the learning amplitude (2) is $\psi_a \propto f$ if $\sigma_a = \sigma_B$ or $\psi_a \propto f - g_{a,c_1} - \ldots - g_{a,c_m}$, if $a$ disagrees with $B$ and agrees with some of its neighbors $c_i \in \{c \in \mathbb{N}_a | \sigma_a = \sigma_c\}$. Observe that if $a$ disagrees with $B$ the added effect of $a$'s agreeing neighbors could make $\mathbf{J}_a$ grow in opposite direction to $\mathbf{B}$.

### A. The learning equations

When the number of examples in the training set $\mathbb{S}$ is large ($T \to \infty$) the state of the society can be assessed by measuring the overlaps between synaptic vectors:

$$R_a \equiv \frac{\mathbf{J}_a \cdot \mathbf{B}}{|\mathbf{J}_a||\mathbf{B}|} = \cos(\theta_a), \tag{3}$$

where $\theta_a$ is the angle between $\mathbf{B}$ and $\mathbf{J}_a$. $R_a$ represents the level of agreement of agent $a$ with the society $B$. Similarly, the level of agreement between two agents can be represented by:

$$W_{a,b} \equiv \frac{\mathbf{J}_a \cdot \mathbf{J}_b}{|\mathbf{J}_a||\mathbf{J}_b|} = \cos(\gamma_{a,b}) \tag{4}$$

where $\gamma_{a,b}$ is the angle between $\mathbf{J}_a$ and $\mathbf{J}_b$. By defining the vector $\mathbf{J}_{c,\perp} = \mathbf{J}_c - |\mathbf{B}|^{-2}(\mathbf{B} \cdot \mathbf{J}_c)\mathbf{B}$ which is the component of $\mathbf{J}_c$ in hyper-plane perpendicular to $\mathbf{B}$, we can define the overlaps:

$$Y_{a,b} \equiv \frac{\mathbf{J}_{a,\perp} \cdot \mathbf{J}_{b,\perp}}{|\mathbf{J}_{a,\perp}||\mathbf{J}_{b,\perp}|} = \cos(\varphi_{a,b}), \tag{5}$$

where $\varphi_{a,b}$ is the angle between $\mathbf{J}_{a,\perp}$ and $\mathbf{J}_{b,\perp}$. The relationship between (3), (4) and (5) is:

$$W_{a,b} = R_a R_b + Y_{a,b}\sqrt{(1 - R_a^2)(1 - R_b^2)}. \tag{6}$$

$Y_{a,b}$ represents the level of agreement between agents $a$ and $b$ on issues $\mathbf{S}_0$ that are *socially neutral,* i.e. $\mathbf{B} \cdot \mathbf{S}_0 = 0$. Therefore, the state of the society can be described by the sets of overlaps $\{R_a\}$, defined on the vertexes and $\{Y_{a,b}\}$, defined on the bonds of the graph $\boldsymbol{G}$, to describe the state of the society.

The components of the issue vectors $\mathbf{S}$ are random variables with $\mathcal{P}(S_j = 1) = \frac{1}{2}$. To average out this disorder we define the variables:

$$\phi_a \equiv \sigma_B \frac{\mathbf{J}_a \cdot \mathbf{S}}{|\mathbf{J}_a|} \tag{7}$$

$$\beta \equiv \sigma_B \frac{\mathbf{B} \cdot \mathbf{S}}{|\mathbf{B}|} \tag{8}$$

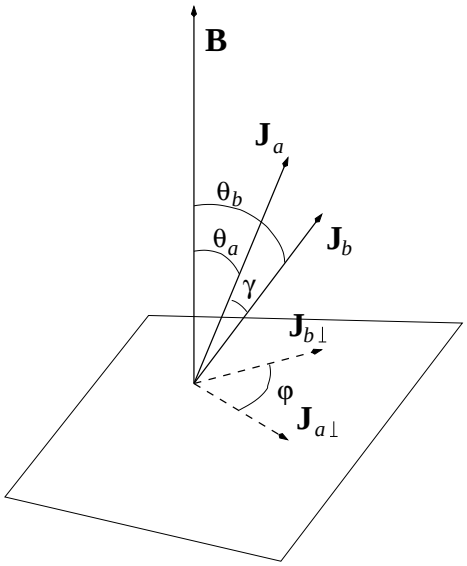


Figure 1: Pictorial representation of the synaptic vectors $\mathbf{B}$, $\mathbf{J}_a$ and $\mathbf{J}_b$ and the angles between them.

which are the agent's and society's post-synaptic fields respectively.

The data accessible to the agent $a$ is $(\sigma_B, \phi_a, \phi_{\mathbb{N}_a}, \mathbf{S})$ where $\phi_{\mathbb{N}_a} \equiv \{\phi_c | c \in \mathbb{N}_a\}$. With this information we have to construct the equations for the opinion formation process local to agent $a$. By defining the norm $Q_a \equiv \mathbf{J}_a \cdot \mathbf{J}_a / N$ and from equation (1) we have that, disregarding terms of $\mathcal{O}(f^2 N^{-1})$:

$$\frac{1}{\sqrt{Q_{a,n+1}}} \simeq \frac{1}{\sqrt{Q_{a,n}}} \left(1 - \Psi_{a,n} \phi_{a,n} \frac{f}{N}\right). \tag{9}$$

Using the definition of $Q_{a,n}$ and the expansion (9) we have that:

$$R_{a,n+1} \simeq R_{a,n} + \Psi_{a,n} \left(\beta_n - \phi_{a,n} R_{a,n}\right) \frac{f}{N}. \tag{10}$$

The length of the training set is $T = \alpha_{\max} N$, which implies that $\alpha_{\max} = T/N$. For a given number $1 \leq n < N$ of examples presented to the perceptrons there is an $0 < \alpha < \alpha_{\max}$ 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 $\Delta t \equiv f \Delta\alpha = f/N$ we have that the equation for the evolution of the overlap $R_a$ is:

$$\frac{\Delta R_a}{\Delta t} \simeq \Psi_a \left(\beta - \phi_a R_a\right) + \mathcal{O}(f). \tag{11}$$

From (5), (3), (4), the definitions of $Q_{a,n}$ and $\Delta t$ we have that:

$$\begin{aligned} \frac{\Delta Y_{a,b}}{\Delta t} = & \frac{\Psi_a}{\sqrt{1 - R_a^2}} \left[\frac{\phi_b - R_b \beta}{\sqrt{1 - R_b^2}} - Y_{a,b} \frac{\phi_a - R_a \beta}{\sqrt{1 - R_a^2}}\right] + \\ & + \mathrm{IT}_{b,a} + \mathcal{O}(f), \end{aligned} \tag{12}$$

where $\mathrm{IT}_{b,a}$ represents a term, identical to the first one in (12) with the indexes $a$ and $b$ interchanged.

In the limit of $N \to \infty$ the overlaps $\{R_a\}$ and $\{Y_{a,b}\}$ are self-averaging [23], thus $\lim_{N\to\infty} \langle R_a \rangle = R_a$ and $\lim_{N\to\infty} \langle Y_{a,b} \rangle = Y_{a,b}$. The particular form the averages of the equations (11) and (12) have in the large system ($N \to \infty$) optimal algorithm ($f \to 0$) limit depend on the distribution $\mathcal{P}(\beta, \{\phi_a\})$, used to compute the expectation values represented by $\langle \cdot \rangle$ . In the following we present different settings where the distribution $\mathcal{P}(\beta, \{\phi_a\})$ acquires different forms.

### 1. *Dimer*

From the developments of the appendix A we have that in the case of a society formed by only two agents the equation for the vertex and bond overlaps are:

$$\begin{aligned} \dot{R}_a &= (1 - R_a^2)\frac{2 - \eta_{a,b}}{2} + \\ &\quad + \frac{\eta_{a,b}}{2}\left[(1 - R_a^2)\frac{\varphi_{a,b}}{\pi} + \rho_{a,b}\,(R_b - W_{a,b}R_a)\right] \end{aligned} \tag{13}$$

$$\dot{Y}_{a,b} = (1 - Y_{a,b}^2)\left[\sqrt{\frac{1 - R_b^2}{1 - R_a^2}}\eta_{a,b}\rho_{a,b} + \mathrm{IT}_{b,a}\right] \tag{14}$$

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

$$\rho_{a,b} \equiv \frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{R_a - W_{a,b}R_b}{\sqrt{(1 - R_a^2)(1 - R_b^2)(1 - Y_{a,b}^2)}}\right). \tag{15}$$

The RHS of equation (14) is non-negative, thus the stable solution for this set of equations is $Y_{a,b} = 1$, which implies total agreement between $a$ and $b$ on socially neutral issues. As a consequence, equation (15) becomes $\lim_{Y_{a,b}\to 1} \rho_{a,b} = \Theta(R_b - R_a)$. Thus

$$\begin{aligned} \dot{R}_a &= \frac{2 - \eta_{a,b}}{2}(1 - R_a^2) + \\ &\quad + \frac{\eta_{a,b}}{2}\Theta(R_b - R_a)(R_b - W_{a,b}R_a). \end{aligned} \tag{16}$$

The equivalent equation for $R_b$ can be obtained by interchanging subindexes $a$ and $b$.

If agent $a$ has a conservative attitude $2 > \eta_{a,b}$ (the adjective *conservative* refers to the attitude of an agent $c$ for which the socially accepted opinion $\sigma_B$ weighs more than the opinion $\sigma_d$ of its neighbor, i.e. $\eta_{c,d} < 2$), the RHS of (16) is always positive and $R_a \to 1$ asymptotically. Suppose that $R_b < R_a$. By defining:

$$K_b \equiv \frac{2 - \eta_{b,a}\,[1 - \Theta(R_a - R_b)R_a]}{\eta_{b,a}\Theta(R_a - R_b)\sqrt{1 - R_a^2}}, \tag{17}$$

the stable solution of (16) at the current time is:

$$R_{b,0} = \frac{K_b}{\sqrt{1 + K_b^2}}. \tag{18}$$

If $R_a = 1 - \varepsilon_a$, with $0 < \varepsilon_a \ll 1$, the root $R_{b,0} \approx 1 - \frac{1}{4}\eta_{b,a}^2\varepsilon_a$. Therefore, one agent with a conservative attitude is sufficient to build up consensus with the society.

Suppose both agents have a liberal attitude, $\eta_{a,b}, \eta_{b,a} > 2$ and, without loss of generality, suppose that $R_b < R_a$. As far as this inequality holds, $R_a \to -1$ and $R_b \to R_{b,0}$ given by (18). If the relationship between agents change, i.e. $R_a < R_b$ at a posterior time, $R_b \to -1$ and $R_a \to R_{a,0}$ equivalent to the given in (18). In both cases the agents ratchet their overlaps down, until their respective values of $K$, when well defined, become negative. If the largest overlap is close to -1, the smallest is even closer. Clearly the only stable solution is $R_a, R_b \to -1$ asymptotically and the system gets polarized with respect to $B$.

### B. Beyond the dimer

To go beyond the dimer we will consider societies with $M^2$ members, organized on a graph $\boldsymbol{G}$ without loops. This allows us to factorize the global joint probabilities into factors involving pairs of linked post-synaptic field variables, which is consistent with a description where agents have access to local information only.

To calculate the terms that appear in the equation for the overlaps $Y_{a,b}$ we observe that the locally estimated field $\tilde{\beta}_a$ can be decomposed into $\tilde{\beta}_a = \tilde{\beta}_{a,b} + \Delta\beta_{a,b}$ where $\tilde{\beta}_{a,b}$ is the estimation of $B$'s post-synaptic field inferred using only $\phi_a$ and $\phi_b$ and $\Delta\beta_{a,b}$ is the correction due to the other neighborhood variables $\{\phi_c | c \in \mathbb{N}_a, c \neq b\}$ (see the complete derivation in Appendix B). By disregarding the average interaction of the learning algorithm and these corrections, i.e. $\langle \Psi_a \Delta\beta_{a,b} \rangle \sim 0$, the equation for the overlap $Y_{a,b}$ becomes identical to (14). By making $Y_{a,b} = 1$ [24], the opinion formation process becomes controlled by the expression (see appendix B):

$$\begin{aligned} \dot{R}_a \;\leq\; & (1 - R_a^2) \left( 1 - \sum_{c \in \mathbb{N}_a} \frac{\eta_{a,c}}{2} \right) + \\ & + \sum_{c \in \mathbb{N}_a} \frac{\eta_{a,c}}{2} \Theta(R_c - R_a)(R_c - R_a W_{a,c}) \end{aligned}$$

which is very similar to (16). As in the dimer case, the equation that determines the opinion formation process consists of two terms, the first is an effective learning from $B$, with an effective attitude given by $\sum_{c \in \mathbb{N}_a} \eta_{a,c}$. The second term is a perturbation formed by a linear superposition of terms depending on the differences $\theta_a - \theta_c$.

## III. THE DISORDER INTRODUCED THROUGH THE SOCIAL STRENGTHS

Let us assume that our model is reduced to the set of equations:

$$\begin{aligned} \dot{R}_a \;=\; & (1 - R_a^2) \left( 1 - \sum_{c \in \mathbb{N}_a} \frac{\eta_{a,c}}{2} \right) + \\ & + \sum_{c \in \mathbb{N}_a} \frac{\eta_{a,c}}{2} \Theta(R_c - R_a)(R_c - R_a W_{a,c}). \end{aligned} \tag{19}$$

This case is different from the dimer, given that there is more than one neighbor that contributes to the effective attitude of the agent.

To better understand the role of the disorder introduced by the social strengths we consider the case where the $\eta_{a,c}$ are drawn from a Gaussian distribution centered at $\eta$ and with variance $\Delta^2$, i.e. $\eta_{a,c} \sim \mathcal{N}(\eta_{a,c}|\eta, \Delta^2)$. $\eta$, 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 $\Delta$ controls the level of variation, or disorder, in the set of social strengths.

Suppose the system is defined on a loop-free graph and where each vertex has precisely $n$ neighbors (Cayley tree). Let us also assume that the system has been set with an average social strength such that $2/(n-1) > \eta > 2/n$ and small disorder $\Delta \ll \eta$. It may occur with probability:

$$\mathbb{P}_0 = \mathcal{H}\left(\frac{n\eta - 2}{\sqrt{n}\Delta}\right) \ll \frac{1}{2},$$

where $\mathcal{H}(x) \equiv \int_x^\infty \mathrm{d}y\, \mathrm{e}^{-y^2/2}/\sqrt{2\pi}$ is the Gardner error function, that a vertex $o$ has a conservative attitude $2 > \sum_{c\in\mathbb{N}_o} \eta_{o,c}$, in which case, for times not smaller than a sufficiently large $t_0$, $R_o > R_b$ for all $b \in \mathbb{N}_o$. Clearly $R_o \to 1$ and each neighbor $b$ of $o$ will have a well defined $K_b$:

$$K_b \equiv \frac{2 - \sum_{c\in\mathbb{N}_b} \eta_{b,c}\left[1 - \Theta(R_c - R_b)R_c\right]}{\sum_{c\in\mathbb{N}_b} \eta_{b,c}\Theta(R_c - R_b)\sqrt{1 - R_c^2}} \tag{20}$$

and a stable solution $R_{b,0} = K_b/\sqrt{K_b^2 + 1}$. Observe that the probability:

$$\mathbb{P}(K_b > 0) \approx 1 - \mathcal{H}\left(\frac{2 - (n-1)\eta}{\sqrt{n-1}\Delta}\right), \tag{21}$$

is close to 1 and $R_b \to R_{b,0} \to 1$. Assuming that the neighbors of conservative agents have an effective conservative attitude with a probability given by (21), the probability of a conservative cluster with a center surrounded by $\ell$ concentric shells of neighbors is:

$$\mathbb{P}_\ell \approx \mathbb{P}_0\left[1 - n^\ell \mathcal{H}\left(\frac{2 - (n-1)\eta}{\sqrt{n-1}\Delta}\right)\right].$$

The largest cluster size $L$ is defined by $\mathbb{P}_L = 0$, that, for $\Delta \ll \eta$ scales like:

$$L \approx \frac{\Delta_0^2}{\Delta^2} + \mathcal{O}(\log \Delta).$$

Observe that both, the constant $\Delta_0$ and the exponent of $\Delta$, are determined by the topology of the graph (Cayley tree in this case). The constant $\Delta_0$ is the maximum value of the disorder for which the system supports conservative clusters with a minimum size of $L = 1$. The largest number of agents a conservative cluster may have is $M^2$, the whole population. If the cluster covers the whole system, $L \sim \mathcal{O}(M)$ and the maximum value of the disorder for a system with a maximum possible number of clusters equal to one is $\Delta_m = \Delta_0/M^{1/2}$.

## IV. EXPERIMENTS IN THE SQUARE LATTICE

In order to gauge the system's behavior we integrate the equations (19) up to an arbitrary long time $t_{\max}$, for a given set $\{\eta_{a,c}\}$ of social strengths on a society with $M^2$ members and compute the following observable:

$$\mu[\{\eta_{a,c}\}] \equiv \frac{1}{M^2}\sum_a R_a(t_{\max}). \tag{22}$$

$R_a(t_{\max})$ represents the level of agreement of agent $a$ with $B$ at the end of the learning process of length $t_{\max}$, therefore $\mu$, the magnetization, measures the social agreement with $B$ on the current social issue. Assuming that $\eta_{a,b} \sim \mathcal{N}(\eta_{a,b}|\eta, \Delta^2)$ and a graph with co-ordination number $n$, if $n\eta < 2-3n\Delta$ almost all agents have a conservative attitude $\sum_{c\in\mathbb{N}_a}\eta_{a,c} < 2$ that results on a consensus with $B$ at the end of the learning process $R_a(t_{\max}) \approx 1$. If we increment $\eta$ in $\delta\eta \ll \Delta$ and integrate the system (19) again up to $t_{\max}$, considering the initial condition $\{R_a(0, \eta+\delta\eta) = R_a(t_{\max}, \eta)\}$, we may observe a behavioral change in those agents which have suffered a large enough increment in their attitude, becoming liberal $\sum_{c\in\mathbb{N}_a}\eta_{a,c} > 2$, resulting on a reduction of the overlap $R_a(t_{\max}) < 1$. Such changes may occur, according to the predictions of Section III, after conservative clusters appear, keeping the value of the magnetization close to one even for values of $n\eta > 2$. Eventually, for values of $\eta$ sufficiently large, the system's magnetization saturates to a polarized state $\mu \approx -1$. The path back from polarized to consensual behavior can be constructed by starting with a large enough average strength $n\eta > 2 + 3n\Delta$, a polarized initial condition $\{R_a(0, \eta) = -1\}$ and by reducing the average strength $\eta$ in steps $\delta\eta \ll \Delta$. Although there is only one stable solution to the system of equations (19), the time it takes for the conservative clusters to emerge from a polarized initial condition is longer than the one it takes the system to reach the stable solution from the consensual condition and, effectively, the path from a polarized to a consensual behavior may appear identical to the consensual-to-polarized path, but shifted to the left.

To explore this phenomenon we perform a number of numerical integrations of the system (19), defined over a $M^2 = 100 \times 100$ square lattice with periodic boundary conditions. An agent is placed at each vertex of the lattice and its neighborhood is form by its first nearest neighbors ($n = 4$ for all $a$). The system so constructed forms an array $\mathbf{R} \in [-1, 1]^{M\times M}$, with entries $(\mathbf{R})_{j,k} = R_{a(j,k)}$, where $a = j + (M-1)k$. Although a square lattice is not a loop free graph, considering first neighbors only satisfies the condition $\mathbb{N}_a \cap \mathbb{N}_b = \emptyset$ for any pair of agents $a$ and $b$.

We integrate numerically the system (19) up to $t_{\max} = 100$ time units using a second order Runge-Kutta method, with consensual initial conditions $\{R_a(0) = 1\}$, for initial values of the average strength $\eta = \frac{1}{2} - 3\Delta$, for various values of $\Delta$. We expect this initial run to saturate at consensus $\mu \approx 1$. We slowly incremented $\eta$ by an amount $\delta\eta \sim \Delta/10$ always taking as initial condition the final configuration of the previous run $\{R_a(0, \eta+\delta\eta) = R_a(t_{\max}, \eta)\}$, until reaching values of $\eta$ such that

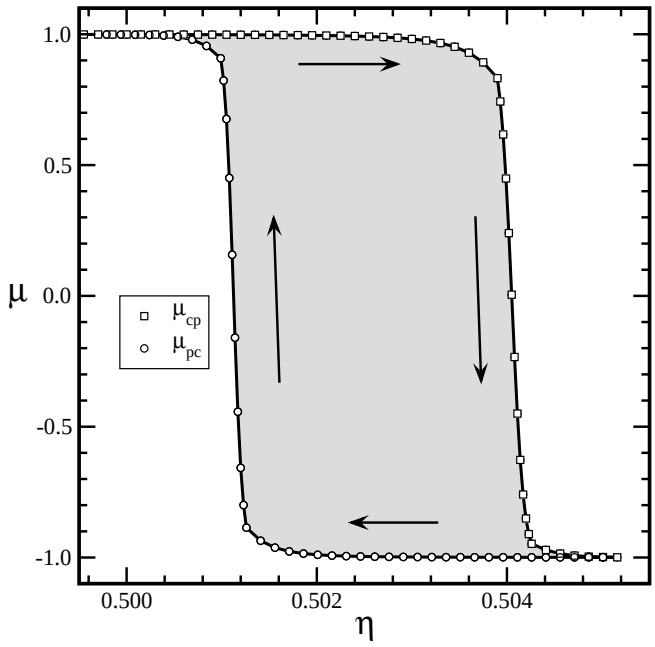


Figure 2: Hysteresis cycle for a system with disorder parameter $\Delta = 1.5 \times 10^{-3}$. The square symbols represent the $\mu_{\rm cp}$ curve and the circles represent the $\mu_{\rm pc}$ curve.

the system's final magnetization reaches polarized saturation $\mu \approx -1$. The magnetization curve so obtained represents the path from a consensual to a polarized position with respect to $B$, $\mu_{\rm cp}(\eta)$. We take this final (polarized) configuration as the initial configuration for the polarized to consensual path, that can be constructed by reducing adiabatically $\eta$ in amounts of $\delta\eta$. The magnetization curve so obtained is labeled $\mu_{\rm pc}(\eta)$.

We follow the evolution of the system by plotting $\mu$ as a function of the social field $\eta$ for several values of the disorder parameter $10^{-4} \leq \Delta \leq 10^{-1}$. For all the values of $\Delta$ analyzed we found that the magnetization $\mu$ develops hysteresis cycles with negative (clockwise) areas, $\mu_{\rm cp}(\eta) > \mu_{\rm pc}(\eta)$. We define two critical values of $\eta$ through the equations $\mu_{\rm cp}(\eta_{\rm cp}) = 0$ and $\mu_{\rm pc}(\eta_{\rm pc}) = 0$. In figure 2 we present a hysteresis cycle for a system with a disorder $\Delta = 1.5 \times 10^{-3}$.

In figure 3 we present a plot of the right shift in the curve $\mu_{\rm cp}$, $|\eta_{\rm cp} - 2/n|$ and the left shift in the curve $\mu_{\rm pc}$, $|\eta_{\rm pc} - 2/n|$ as a function of the disorder $\Delta$. We observe that the right shift in $\mu_{\rm cp}$ obeys a scaling low proportional to $\Delta^{0.8}$, whilst the left shift in $\mu_{\rm pc}$ asymptotically approaches $\mu_{\rm cp}$ for values of the disorder $1.8 \times 10^{-3} \leq \Delta$.

We expect to have only one macroscopic conservative cluster for systems with sufficiently small disorder $\Delta$. We found that for systems with disorder $\Delta \lesssim \Delta_0 = 1.8 \times 10^{-3}$ there is mostly only one conservative cluster and the curves $\mu_{\rm cp}(\eta)$ as a function of $x \equiv (\eta - \eta_{\rm cp}(\Delta))/\Delta^{\lambda}$ and $\mu_{\rm pc}(\eta)$ as a function of $x \equiv (\eta - \eta_{\rm pc}(\Delta))/\Delta^{\lambda}$ collapse into a single function when $\lambda \sim 0.3$. This result for systems with $\Delta < \Delta_0$ is presented in figure 4. In the insets we present snapshots of the systems at $x = 0$. For larger values of $\Delta > \Delta_0$ we have systems that can support more than one conservative cluster, with cluster sizes that get smaller, the larger the value of the disorder $\Delta$ (figure 5). The exponent

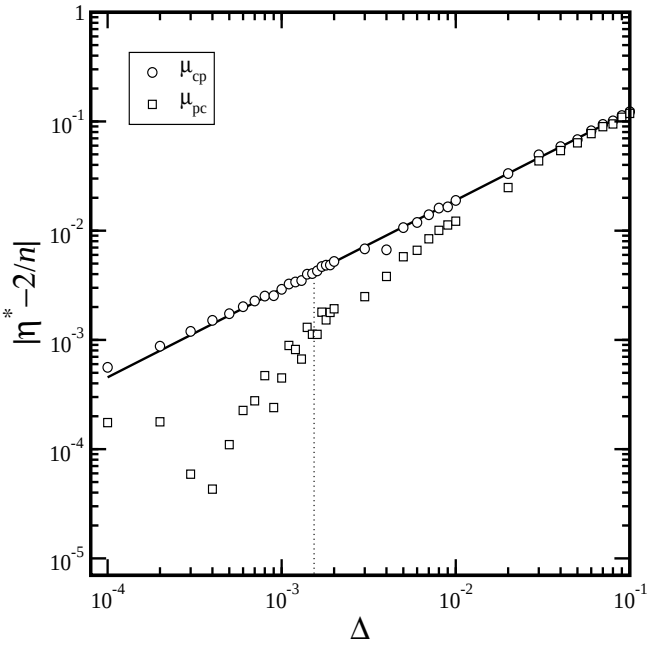


Figure 3: Right shift in the curve $\mu_{\rm cp}$, $|\eta_{\rm cp} - 2/n|$ and the left shift in the curve $\mu_{\rm pc}$, $|\eta_{\rm pc} - 2/n|$ as a function of the disorder $\Delta$. The label in the vertical axis is $|\eta^\star - 2/n|$ where $\eta^\star$ stands for either $\eta_{\rm cp}$ or $\eta_{\rm pc}$. The straight line is the best fit for the right shift of $\mu_{\rm cp}$, which is a scaling law proportional to $\Delta^{0.8}$. The left shift in $\mu_{\rm pc}$ asymptotically approaches this behavior for values of $\Delta$ larger than $1.8 \times 10^{-3}$ (indicated by the dotted line).

for the latter case is $\lambda \sim 1$.

The area of the hysteresis cycle as a function of the disorder parameter $\Delta$ is:

$$\begin{aligned} \mathrm{A}(\Delta) &\equiv \oint \mathrm{d}\eta\, \mu(\eta, \Delta) \\ &= \int_{1/2-3\Delta}^{1/2+3\Delta} \mathrm{d}\eta\, \left[\mu_{\rm cp}(\eta, \Delta) - \mu_{\rm pc}(\eta, \Delta)\right]. \end{aligned}$$

The hysteresis cycle, if approximated by a rectangle, should cover an area proportional to the disorder parameter $\Delta$. Therefore the quantity defined as $a(\Delta) \equiv \mathrm{A}(\Delta)/\Delta$ should be a constant if the cycles are stable for large values of $\Delta$. We expect to see a decrement in the area of the cycle associated with a decrement in the size of the conservative clusters in the $\mu_{\rm cp}$ path, as $\Delta$ gets larger. The measured areas obey the following power law $a(\Delta) = A_0 \Delta^{-\lambda}$ with $\lambda = 0.50 \pm 0.02$ for values of the disorder in the range $1 \times 10^{-4} \leq \Delta \leq 1 \times 10^{-1}$. the curves $\mu_{\rm cp}$ and $\mu_{\rm pc}$ collapse into each other and the hysteresis disappears.

## V. 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. We observed that by means of the self-averaging property of the relevant parameters $\{R_a\}$ and $\{Y_{a,b}\}$, the description of the system is given

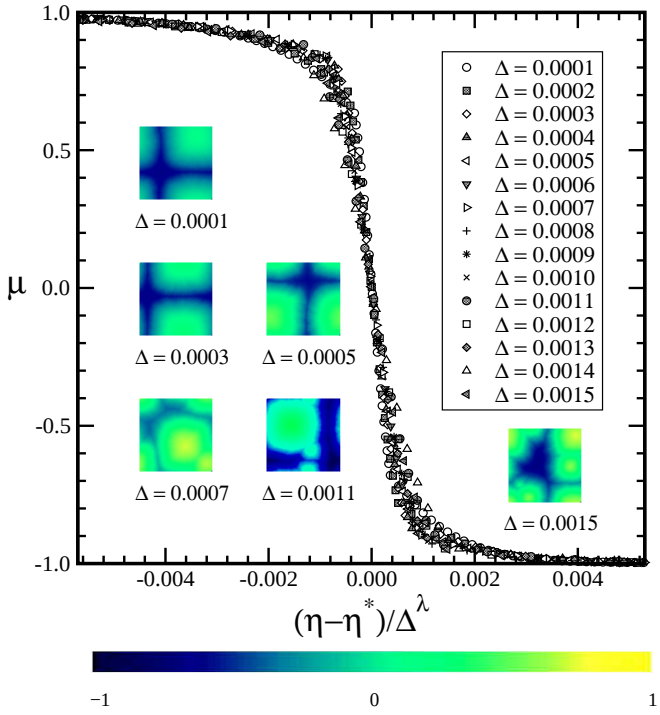


Figure 4: Data collapse of $\mu_{\rm cp}$ as a function of $x \equiv (\eta - \eta_{\rm cp}(\Delta))/\Delta^{0.3}$ and $\mu_{\rm pc}$ as a function of $x \equiv (\eta - \eta_{\rm pc}(\Delta))/\Delta^{0.3}$ for several values of $\Delta < 1.8 \times 10^{-3}$. In the inset we have the correspondent snapshots of the arrays $\mathbf{R}$ at $x = 0$. In all the cases the number of conservative clusters supported by the system is mostly 1 (color on-line).

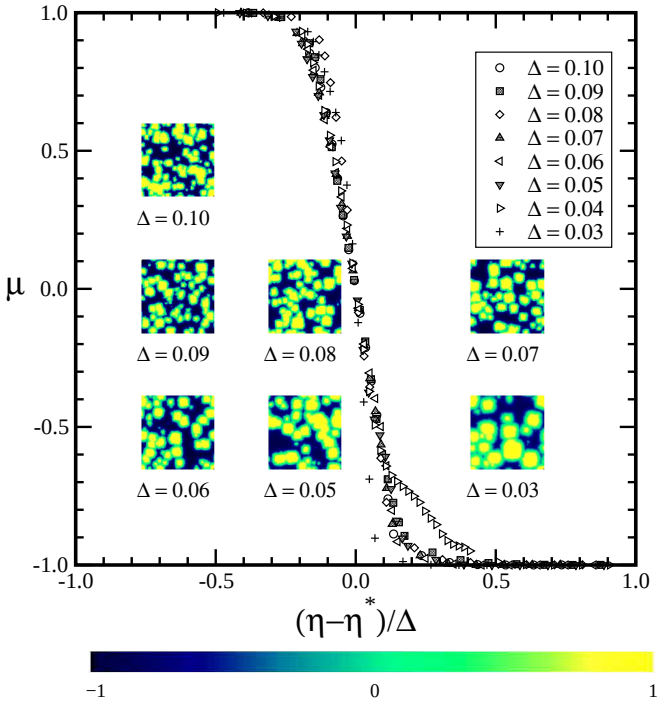


Figure 5: Data collapse of $\mu_{\rm cp}$ as a function of $x \equiv (\eta - \eta_{\rm cp}(\Delta))/\Delta$ and $\mu_{\rm pc}$ as a function of $x \equiv (\eta - \eta_{\rm pc}(\Delta))/\Delta$ for several values of $\Delta > 1.8 \times 10^{-3}$. In the inset we have the correspondent snapshots of the arrays $\mathbf{R}$ at $x = 0$. In all the cases the number of conservative clusters supported by the system is larger than one and the larger the disorder parameter $\Delta$ the smaller the size of the clusters observed (color on-line).

by equation (19). In this approximation the social agents agree in all issues that are socially neutral ($Y_{a,b} = 1$). One can suppose that this inter-agent agreement in the hyper-plane perpendicular to $\mathbf{B}$ would add up to a social consensus different from the imposed by $B$. To understand this point, we have to realize that the agreement conveyed by $\{Y_{a,b} = 1\}$ is particular to the present opinion formation process, and that the social consensus $B$ is the integrated result of many similar processes previously undergone by the society. The rigidity of $B$ is the rigidity observed in the written law which, despite the opinion of the individuals on the conjunctural social issues, must be obeyed and can only be modified after building up a new, comprehensive consensus.

By considering social strengths drawn from a Gaussian distribution centered at $\eta$ and with standard deviation $\Delta$, the curves we obtained for the overall opinion of the system $\mu$ developed hysteresis cycles, as functions of $\eta$. These cycles can be associated to the delay in the emergence of conservative clusters in the path from polarized to consensual behavior. The delay (left shift) in the emergence of the conservative clusters in the polarized-to-conservative path is due to the relatively large distance of the polarized initial condition to the consensual stable solution of the system (19) and it is more pronounced for systems that develop only one macroscopic conservative cluster (see figure 3). Conservative clusters may have a size equal to the size of the system for small enough disorder and they decrease in size as the value of the disorder $\Delta$ increases. By studying the areas of the hysteresis cycles, we observed that, for large enough $\Delta$, the curves $\mu_{\rm cp}(\eta)$ and $\mu_{\rm pc}(\eta)$ collapsed on each other and the cycles disappeared. The decay of the relative area of the hysteresis cycle was empirically observed to obey the power law $a \propto \Delta^{-\lambda}$ with $\lambda = 0.50 \pm 0.02$.

These disorder-driven, zero temperature changes, are very similar to the ones occurring in the random field Ising model, as reported in [26]. Moreover, $\mu_{\rm cp}(\eta)$ curves for several values of $\Delta$ have shown evidence of crackling noise [27], as it is observed in the tails of $\mu \sim -1$ presented in figure 5.

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 $\eta$). 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* [7] and conservative clusters emerge and remain, even if $\eta$ is increased by the emergence of more evidence in favor of the challenging policy.

### Appendix A: Dimer

By supposing that the entries of $\mathbf{S}$ are iid variables with $\mathcal{P}(S_i = 1) = \frac{1}{2}$ we can compute the joint probability $\mathcal{P}(\beta, \{\phi_a\}) = \int \mathrm{d}\mathbf{S}\mathcal{P}(\beta, \{\phi_a\}|\mathbf{S})\mathcal{P}(\mathbf{S})$. In particular, for a dimer, i.e. a society composed by only two agents, and for a large enough $N$ the joint probability $\mathcal{P}(\beta, \phi_a, \phi_b)$ can be computed exactly:

$$\begin{aligned}\mathcal{P}(\beta, \phi_a, \phi_b) &= \mathcal{N}\left(\beta \left| \Sigma_{a,b}\Lambda(\phi_a, \phi_b), \Sigma^2_{a,b}\right.\right) \\ &\quad \mathcal{N}\left(\phi_b|W_{a,b}\phi_a, 1 - W^2_{a,b}\right)\mathcal{N}(\phi_a) \qquad \text{(A1)} \\ \mathcal{P}(\phi_a, \phi_b) &= 2\mathcal{H}\left(-\Lambda(\phi_a, \phi_b)\right) \\ &\quad \mathcal{N}\left(\phi_b|W_{a,b}\phi_a, 1 - W^2_{a,b}\right)\mathcal{N}(\phi_a), \qquad \text{(A2)}\end{aligned}$$

where

$$\begin{aligned}\Lambda(\phi_a, \phi_b) &\equiv \frac{(R_a - W_{a,b}R_b)\phi_a + (R_b - W_{a,b}R_a)\phi_b}{\Omega_{a,b}\sqrt{1 - W^2_{a,b}}} \\ \Sigma^2_{a,b} &\equiv \frac{\Omega^2_{a,b}}{1 - W^2_{a,b}}\end{aligned}$$

and where $\mathcal{N}(x|\mu, \sigma^2)$ is a Gaussian distribution for $x$ centered at $\mu$ with variance $\sigma^2$. We also use the following notation $\mathcal{N}(x) \equiv \mathcal{N}(x|0, 1)$, $\mathcal{H}(x) \equiv \int_x^\infty \mathrm{d}y\mathcal{N}(y)$ is the Gardner error function and $\Omega^2_{a,b} = (1 - R^2_a)(1 - R^2_b)(1 - Y^2_{a,b})$. By using (A1) and (A2) we can average out the dependency on the training set (through the variables $\phi_a$, $\phi_b$ and $\beta$) in the equations (11) and (12). By using that in the large system size limit ($N \to \infty$) the overlaps $\{R_a\}$ and $\{Y_{a,b}\}$ are self-averaging quantities [23], and by optimizing the learning algorithm (1) by taking the limit of $f \to 0$, with $\eta_{a,b} \equiv \lim_{f\to 0} g_{a,b}/f$, we have that $\lim_{N\to\infty}\lim_{f\to 0} = \lim_{\Delta t\to 0}$ and the evolution of the overlaps is ruled by:

$$\begin{aligned}\frac{\mathrm{d}R_a}{\mathrm{d}t} &\equiv \lim_{\Delta t\to 0}\int \mathrm{d}\phi_a\mathrm{d}\phi_b\mathrm{d}\beta\,\mathcal{P}(\beta, \phi_a, \phi_b)\frac{\Delta R_a}{\Delta t} \\ &= \left\langle \Psi_a\left(\tilde{\beta}_{a,b} - \phi_a R_a\right)\right\rangle,\end{aligned}$$

where the external angular brackets represent expectation over the distribution $\mathcal{P}(\phi_a, \phi_b)$ and the estimated social post-synaptic field, represented by the conditional average:

$$\begin{aligned}\tilde{\beta}_{a,b} &\equiv \int \mathrm{d}\beta\,\beta\,\mathcal{P}(\beta|\phi_a, \phi_b) \\ &= \Sigma_{a,b}\left[\mathcal{F}\left(\Lambda(\phi_a, \phi_b)\right) + \Lambda(\phi_a, \phi_b)\right], \qquad \text{(A3)}\end{aligned}$$

where $\mathcal{F}(x) \equiv \mathcal{N}(x)/\mathcal{H}(-x)$. By using the results (B6) and (B7) and by re-scaling the time by $t \to \sqrt{\frac{2}{\pi}}t$, we have that the equation for the vertex variables is:

$$\begin{aligned}\dot{R}_a &= (1 - R^2_a)\frac{2 - \eta_{a,b}}{2} + \\ &\quad + \frac{\eta_{a,b}}{2}\left[(1 - R^2_a)\frac{\varphi_{a,b}}{\pi} + \rho_{a,b}\left(R_b - W_{a,b}R_a\right)\right], \qquad \text{(A4)}\end{aligned}$$

where

$$\rho_{a,b} \equiv \frac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{R_a - W_{a,b}R_b}{\Omega_{a,b}}\right). \tag{A5}$$

In similar manner:

$$\dot{Y}_{a,b} \;=\; (1 - Y_{a,b}^2)\left[\sqrt{\frac{1-R_b^2}{1-R_a^2}}\eta_{a,b}\rho_{a,b} + \mathrm{IT}_{b,a}\right]. \tag{A6}$$

**Appendix B: Beyond the dimer**

To go beyond the dimer we will consider societies with $M^2$ members. This allows us to factorize the global joint probabilities into factors involving pairs of linked surprise variables and local social post-synaptic fields only. This is consistent with a description where the agents have access to local information only. From agent $a$'s perspective, the information available is represented by the surprise variables $\{\phi_a, \phi_{\mathbb{N}_a}\}$ and the information to be inferred is represented by the local field $\beta_a$. The probabilities are modeled by $\mathcal{P}(\beta_a, \phi_a, \phi_{\mathbb{N}_a}) = \int \prod_{c\in\mathbb{N}_a} \mathrm{d}\beta_{a,c}\mathcal{P}(\beta_a|\beta_{\mathbb{N}_a})\mathcal{P}(\beta_{a,c}, \phi_a, \phi_c)$, where $\beta_{\mathbb{N}_a} = \{\beta_{a,c}|c \in \mathbb{N}_a\}$, $\beta_{a,c}$ is the field inferred from the knowledge of $\phi_a$ and $\phi_c$ alone and $\mathcal{P}(\beta_{a,b}, \phi_a, \phi_b)$ is given by (A1). By supposing equal probability *a priori:* $\mathcal{P}(\beta_a|\beta_{a,\mathbb{N}_a}) = \frac{1}{|\mathbb{N}_a|}\sum_{c\in\mathbb{N}_a}\delta(\beta_a - \beta_{a,c})$, thus $\mathcal{P}(\beta_a, \phi_a, \phi_{\mathbb{N}_a}) = \frac{1}{|\mathbb{N}_a|}\sum_{c\in\mathbb{N}_a}\mathcal{P}(\beta_a, \phi_a, \phi_c)$ and

$$\mathcal{P}(\beta_a|\phi_a, \phi_{\mathbb{N}_a}) = \frac{\sum_{c\in\mathbb{N}_a}\mathcal{P}(\beta_a, \phi_a, \phi_c)}{\sum_{c\in\mathbb{N}_a}\mathcal{P}(\phi_a, \phi_c)}$$

where $\mathcal{P}(\beta_a, \phi_a, \phi_c)$ is given by (A1) and $\mathcal{P}(\phi_a, \phi_c)$ by (A2). Therefore the inferred local field is:

$$\begin{aligned}\tilde{\beta}_a \;&\equiv\; \int \mathrm{d}\beta_a \frac{\sum_{c\in\mathbb{N}_a}\mathcal{P}(\beta_a, \phi_a, \phi_c)\beta_a}{\sum_{c\in\mathbb{N}_a}\mathcal{P}(\phi_a, \phi_c)} \\ &=\; \sum_{d\in\mathbb{N}_a} A_{a,d}\tilde{\beta}_{a,d},\end{aligned} \tag{B1}$$

where

$$A_{a,d} \equiv \frac{\mathcal{H}\left(-\Lambda(\phi_a, \phi_d)\right)\mathcal{N}\left(\phi_d|W_{a,d}\phi_a, 1-W_{a,d}^2\right)}{\sum_{c\in\mathbb{N}_a}\mathcal{H}\left(-\Lambda(\phi_a, \phi_c)\right)\mathcal{N}\left(\phi_c|W_{a,c}\phi_a, 1-W_{a,c}^2\right)}.$$

The distribution $\mathcal{N}\left(\phi_c|W_{a,c}\phi_a, 1-W_{a,c}^2\right)$ is sharply picked at $\phi_c = W_{a,c}\phi_a$, thus we can estimate the factor in the sum at the RHS of (B1) by

$$\begin{aligned}A_{a,d} \;\approx\; &\frac{\mathcal{H}(-R_a\phi_a\sqrt{1-W_{a,d}^2}/\Omega_{a,d})}{\sum_{c\in\mathbb{N}_a}\mathcal{H}(-R_a\phi_a\sqrt{1-W_{a,c}^2}/\Omega_{a,c})} + \\ &+\sum_{c\in\mathbb{N}_a}\mathcal{O}(W_{a,c}^2 - W_{a,d}^2) + \sum_{c\in\mathbb{N}_a}\mathcal{O}(\phi_c - W_{a,c}\phi_a).\end{aligned}$$

By assuming that $Y_{a,c} \approx 1$ for all $c \in \mathbb{N}_a$ (see bellow) we have that $1 \ll \sqrt{1-W_{a,c}^2}/\Omega_{a,c}$, thus $A_{a,d} \approx 1/|\mathbb{N}_a|$. Finally we have that: $\tilde{\beta}_a \approx \frac{1}{|\mathbb{N}_a|}\sum_{d\in\mathbb{N}_a}\tilde{\beta}_{a,d}$ and:

$$\dot{R}_a \;=\; \left\langle \Psi_a\left(\tilde{\beta}_a - \phi_a R_a\right)\right\rangle \tag{B2}$$

$$\dot{Y}_{a,b} = \left\langle \frac{\Psi_a}{\sqrt{1-R_a^2}} \left[ \frac{\phi_b - R_b\tilde{\beta}_b}{\sqrt{1-R_b^2}} - Y_{a,b}\frac{\phi_a - R_a\tilde{\beta}_a}{\sqrt{1-R_a^2}} \right] \right\rangle + \\ + \mathrm{IT}_{b,a}. \tag{B3}$$

By defining $\Gamma_a \equiv R_a/\sqrt{1-R_a^2}$ we can write:

$$\mathcal{P}(\phi_a) \equiv 2\mathcal{H}(-\Gamma_a\phi_a)\mathcal{N}(\phi_a)$$
$$\mathcal{P}(\phi_c|\phi_a) \equiv \frac{\mathcal{H}\left(-\Lambda(\phi_a,\phi_c)\right)}{\mathcal{H}(-\Gamma_a\phi_a)}\mathcal{N}\left(\phi_c \left| W_{a,c}\phi_a, 1-W_{a,c}^2\right.\right)$$

and we can compute the following integrals:

$$\langle \phi_c - W_{a,c}\phi_a |\, \phi_a \rangle = \frac{R_c - W_{a,c}R_a}{\sqrt{1-R_a^2}}\mathcal{F}(\Gamma_a\phi_a) \tag{B4}$$

$$\langle \mathcal{F}\left(\Lambda(\phi_a,\phi_c)\right) |\, \phi_a \rangle = \frac{\Sigma_{a,c}}{\sqrt{1-R_a^2}}\mathcal{F}(\Gamma_a\phi_a) \tag{B5}$$

and by defining $\Psi_{a,b} \equiv 1 - \eta_{a,b}\Theta(-\phi_a)\Theta(-\phi_b)$ we have that

$$I_1 \equiv \langle \Psi_{a,b}(\phi_b - W_{a,b}\phi_a)\rangle \\ = \sqrt{\frac{2}{\pi}}(R_b - W_{a,b}R_a)\left[1 - \frac{\eta_{a,b}}{2\pi}\arccos(-Y_{a,b})\right] + \\ + \sqrt{\frac{2}{\pi}}\eta_{a,b}\frac{1-W_{a,b}^2}{2}\rho_{a,b}. \tag{B6}$$

$$I_2 \equiv \langle \Psi_{a,b}\mathcal{F}\left(\Lambda(\phi_a,\phi_b)\right)\rangle \\ = \sqrt{\frac{2}{\pi}}\Sigma_{a,b}\left[1 - \frac{\eta_{a,b}}{2\pi}\arccos(-Y_{a,b})\right]. \tag{B7}$$

Therefore $\left\langle \tilde{\beta}_a - \phi_a R_a \right\rangle = \sqrt{\frac{2}{\pi}}(1-R_a^2)$ and

$$I_3 \equiv \left\langle \Theta(-\phi_a)\Theta(-\phi_c)\left(\tilde{\beta}_a - \phi_a R_a\right)\right\rangle \\ = \sqrt{\frac{2}{\pi}}\frac{1-R_a^2}{2|\mathbb{N}_a|}\left[\frac{\pi - \varphi_{a,c}}{\pi} - \frac{R_c - W_{a,c}R_a}{1-R_a^2}\rho_{a,c}\right] + \\ + \frac{|\mathbb{N}_a|-1}{|\mathbb{N}_a|})\sqrt{1-R_a^2}\left\langle \Theta(-\phi_a)\Theta(-\phi_c)\mathcal{F}(\Gamma_a\phi_a)\right\rangle. \tag{B8}$$

To compute the expectations that appear in the equation of the overlaps $Y_{a,b}$ we observe that the local learning amplitude and the local estimated field can be decomposed into $\Psi_a = \Psi_{a,b} - \Delta\Psi_{a,b}$, where $\Delta\Psi_{a,b} \equiv \sum_{c\in\mathbb{N}_a/b}\eta_{a,c}\Theta(-\phi_a)\Theta(-\phi_c)$ and $\tilde{\beta}_a = \tilde{\beta}_{a,b} + \Delta\beta_{a,b}$ where $\Delta\beta_{a,b} \equiv \frac{1}{|\mathbb{N}_a|}\sum_{c\in\mathbb{N}_a/b}\left(\tilde{\beta}_{a,c} - \tilde{\beta}_{a,b}\right)$. The equation for the overlaps $\{Y_{a,b}\}$ can be written as:

$$\dot{Y}_{a,b} = F_{a,b}^{(0)} - F_{a,b}^{(1)} + \Delta F_{a,b}$$

where $F_{a,b}^{(0)}$ is identical to the RHS of (14)

$$F_{a,b}^{(1)} \equiv \left\langle \frac{\Delta\Psi_{a,b}}{\sqrt{1-R_a^2}}\left[\frac{\phi_b - R_b\tilde{\beta}_{a,b}}{\sqrt{1-R_b^2}} - Y_{a,b}\frac{\phi_a - R_a\tilde{\beta}_{a,b}}{\sqrt{1-R_a^2}}\right]\right\rangle +$$

$$\Delta F_{a,b} \equiv \begin{array}{l} +\mathrm{IT}_{b,a} \\ \left\langle \dfrac{\Psi_a}{\sqrt{1-R_a^2}} \left[ \dfrac{R_b \Delta\beta_{b,a}}{\sqrt{1-R_b^2}} - Y_{a,b} \dfrac{R_a \Delta\beta_{a,b}}{\sqrt{1-R_a^2}} \right] \right\rangle + \\ +\mathrm{IT}_{b,a}. \end{array}$$

$F_{a,b}^{(1)} = 0$ and we consider $\Delta F_{a,b}$, the term that accounts for the interaction of the learning amplitude with the contribution of the neighborhood to the estimate of the local field, to be negligible. In such a case the stable solution to the set of equations (B3) is $\{Y_{a,b} = 1\}$ for all pair of vertexes $a$ and $b$ sharing a bond.

In such a case we have to compute

$$I(R_a, R_b) \equiv \lim_{Y \to 1} \langle \Theta(-\phi_a)\Theta(-\phi_b)\mathcal{F}(\Gamma_a \phi_a) \rangle$$

which, by using the definition of the Gardner error function and taken the limit inside the remaining integral, leads to:

$$I(R_a, R_b) = \sqrt{\frac{1-R_a^2}{2\pi}} \left\{ 1 - 2\Theta(R_b - R_a) \int_0^\infty \mathcal{D}\phi \frac{\mathcal{H}(D_{a,b}\phi)}{\mathcal{H}(R_a \phi)} \right\},$$

where

$$D_{a,c} \equiv \frac{(1-R_a^2)\sqrt{1-R_c^2} + R_a R_c \sqrt{1-R_a^2}}{R_c \sqrt{1-R_a^2} - R_a \sqrt{1-R_c^2}} > R_a.$$

By defining $\nu_a \equiv 1/|\mathbb{N}_a|$, the equation for the vertexes variables becomes:

$$\begin{aligned} \dot{R}_a \;=\; & (1-R_a^2) \left\{ \left( 1 - \sum_{c \in \mathbb{N}_a} \frac{\eta_{a,c}}{2} \right) + \right. \\ & + \sum_{c \in \mathbb{N}_a} \frac{\eta_{a,c}\Theta(R_c - R_a)}{2} \left[ \nu_a \frac{R_c - R_a W_{a,c}}{1-R_a^2} + \right. \\ & \left. \left. + (1-\nu_a)\, 2 \int_0^\infty \mathcal{D}\phi_a \frac{\mathcal{H}(D_{a,b}\phi_a)}{\mathcal{H}(R_a \phi_a)} \right] \right\}. \end{aligned} \tag{B9}$$

Observe that equation (B9) is identical to (16) for $|\mathbb{N}_a| = 1$.

It is possible to demonstrate that:

$$R_c - R_a W_{a,c} \geq (1-R_a^2) 2 \int_0^\infty \mathcal{D}\phi_a \frac{\mathcal{H}(D_{a,b}\phi_a)}{\mathcal{H}(R_a \phi_a)}$$

where the equal sign is satisfied for $R_c = R_a$ or $R_c = 1$ only. We can therefore propose an upper bound for the derivative

$$\begin{aligned} \dot{R}_a \;\leq\; & (1-R_a^2) \left\{ \left( 1 - \sum_{c \in \mathbb{N}_a} \frac{\eta_{a,c}}{2} \right) + \right. \\ & \left. + \sum_{c \in \mathbb{N}_a} \frac{\eta_{a,c}\Theta(R_c - R_a)}{2} \frac{R_c - R_a W_{a,c}}{1-R_a^2} + \right\}. \end{aligned}$$

**Acknowledgments**

The author would like to acknowledge the comments of Dr I. Yurkevich at the early stages of this manuscript. The constructive discussions with Dr. R. C Alamino, Dr M. Stich and Dr L. E. Rebollo-Neira are also warmly appreciated.

---

[1] D. Acemoglu and A. Ozdaglar, Dyn. Games. Appl. **1**, 3 (2011).

[2] M. H. DeGroot, J. Am. Stat. Assoc. **63**, 118 (1974).

[3] F. Giardini, D. Vilone and R. Conte, arXiv:1502.06430v4[physics.soc-ph] (2015).

[4] C. G. Lord, L. Ross and M. R. Lepper, J. Personality Soc. Psychology **37**, 2098 (1979).

[5] R. Vicente, A. C. R. Martins and N. Caticha, J. Stat. Mech. doi:10.1088/1742-5468/2009/03/P03015 (2009).

[6] P. Dadenkar, A. Goel and D. T. Lee, Proc. Natl. Acad. Sci. USA **110**, 5791 (2013).

[7] K. Kacperski and J. A. Holyst, J. Stat. Phys. **84**, 169 (1996).

[8] M. Scheffer, F. Westley and W. Brock W, Ecosystems **6**, 493 (2003).

[9] P. D. Finley et al. Opinion Dynamics Modeling in Tobacco Control Policy, Joint Statistical Meeting, Montréal, Québec, Canada (2013).

[10] B. Latané, American Psychologist **36**, 343 (1981).

[11] M. Lewenstein, A. Nowak and B. Latané, Phys. Rev. A **45**, 763 (1992).

[12] A. Engel and C. Van den Broeck, Statistical mechanics of learning, Cambridge: CUP (2001).

[13] N. Caticha and R. Vicente, Advs. Complex Syst. **14**, 711 (2011).

[14] R. Vicente, A. Susemihl, J. P. Jericó and N. Caticha, Physica A **400**, 124 (2014).

[15] R. Metzler, W. Kinzel and I. Kanter, Phys. Rev. E **62**, 2555 (2000).

[16] C. Castellano, S. Fortunato and V. Loretto, Rev. Mod. Phys. **81**, 591 (2009).

[17] N. Eisenberg, M. D. Lieberman and K. D. Williams, Science **302**, 290 (2003).

[18] S. E. Asch, Groups, leadership and men; research in human relations, H. Guetzkow ed. 177 (1951).

[19] D. O. Hebb, The organization of behavior, Wiley, New York (1949).

[20] R. S. Baron et al., J. of Experimental Social Psychology **32**, 537 (1996).

[21] E. Gilbert, T. Bergstrom and K. Karahalios, Proceedings of the Hawaii International Conference on System Sciences, ed. R. J. Sprague, IEEE Computer Society, Washington DC, 1 (2009).

[22] N. Caticha and O. Kinouchi, Philos. Mag. **77**, 1565 (1998).

[23] G. Reents and R. Urbanczik, Phys. Rev. Lett. **80,** 5445 (1998).

[24] A. J. Healy, N. Malhotra and C. H. Mo, Proc. Natl. Acad. Sci. USA **107**, 12804 (2010).

[25] A. Soulier and T. Halpin-Healy, Phys. Rev. Lett. **90**, 258103 (2003).

[26] K. A. Dahmen and J. P. Sethna, Phys. Rev. B **53**, 14872 (1996).

[27] J. P. Sethna, K. A. Dahmen and C. R. Myers, Nature **410**, 242 (2001).