Opinion dynamics of learning agents: does seeking consensus lead to disagreement?

J. Stat. Mech. (2009) P03015

(http://iopscience.iop.org/1742-5468/2009/03/P03015)

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

Opinion dynamics of learning agents: does seeking consensus lead to disagreement?

Renato Vicente[1], **André C R Martins**[1] and **Nestor Caticha**[2]

[1] GRIFE-EACH, Universidade de São Paulo, Campus Leste, 03828-080, São Paulo—SP, Brazil
[2] Departamento de Física Geral, Instituto de Física, Universidade de São Paulo, Caixa Postal 66318, 05315-970, São Paulo—SP, Brazil
E-mail: rvicente@usp.br, acrmartins@usp.br and nestor@if.usp.br

Received 13 November 2008
Accepted 12 February 2009
Published 9 March 2009

Online at stacks.iop.org/JSTAT/2009/P03015
doi:10.1088/1742-5468/2009/03/P03015

Abstract. We study opinion dynamics in a population of interacting adaptive agents voting on a set of issues represented by vectors. We consider agents who can classify issues into one of two categories and can arrive at their opinions using an adaptive algorithm. Adaptation comes from learning and the information for the learning process comes from interacting with other neighboring agents and trying to change the internal state in order to concur with their opinions. The change in the internal state is driven by the information contained in the issue and in the opinion of the other agent. We present results in a simple yet rich context where each agent uses a Boolean perceptron to state their opinion. If the update occurs with information asynchronously exchanged among pairs of agents, then the typical case, if the number of issues is kept small, is the evolution into a society torn by the emergence of factions with extreme opposite beliefs. This occurs even when seeking consensus with agents with opposite opinions. If the number of issues is large, the dynamics becomes trapped, the society does not evolve into factions and a distribution of moderate opinions is observed. The synchronous case is technically simpler and is studied by formulating the problem in terms of differential equations that describe the evolution of order parameters that measure the consensus between pairs of agents. We show that for a large number of issues and unidirectional information flow, global consensus is a fixed point; however, the approach to this consensus is glassy for large societies.

Keywords: interacting agent models

ArXiv ePrint: 0811.2099v3

1742-5468/09/P03015+14$30.00

Contents

1. Introduction

The ability of sustaining disagreement is a vital feature of a deliberative democracy [1]. Experiments show that humans seek conformity to the opinions of those that they interact with (see [2] and references therein) and observational studies of elections show that disagreement is often sustained within close communication networks of individuals [3]. Despite strong microconformity effects, democratic societies tend to sustain long term opinion diversity, generating a microconformity versus disagreement puzzle.

The recent literature on opinion dynamics has mainly focused on mechanisms for attaining consensus. Several attempts to model whether consensus will emerge in a model society have been proposed [4, 5]. Discrete models, like the Voter [6, 7] or Sznajd's [8] models always lead to consensus, at least in the long run and, in that sense, are not good choices for describing disagreement. Social Impact models [9, 10] allow for the survival of diverse opinions, given a proper choice of parameters. While still describing opinions about only one issue, in the Continuous Opinions and Discrete Actions model (CODA) [11], consensus is only obtained at the local level, with the survival of divergent extreme opinions in the society as a whole. This is achieved by distinguishing between the agent internal probabilistic continuous opinion variable, and what other agents observe, a discrete action. The internal probability is updated by Bayesian rules and results from a history of previous interactions between agents.

One important way in which the real world is different from these models is that real people debate several issues simultaneously. In the Axelrod Culture model (ACM) [12], a set of issues (culture) is represented by a vector of opinions. Agents influence their neighbors on a lattice if there is agreement on a minimum number of issues. The ACM has two absorbing states, one corresponding to complete consensus and other to global disagreement, and a non-equilibrium phase transition is observed, having the degree of initial heterogeneity as a control parameter [13]. A modification of the ACM proposed in [1] is capable of sustaining a certain level of local disagreement by introducing an auto-regressive influence, with agents keeping track of each interaction and only being influenced if there is sufficient evidence supporting the opinion being communicated.

Here, we propose to study patterns of disagreement in a population of agents that adapt their internal representations to opinions on a set of issues observed in a social neighborhood. Each agent has its own adaptive decision mechanism. In particular we consider each agent endowed with a neural network and a learning algorithm, which is used to decide on public issues. The theoretical scenario that we have put forward is not limited by the nature of the information processing units under consideration, nor by the underlying geometry of the graph of social interactions. We could well study the interaction of probabilistic units such as hidden Markov models or of deterministic neural networks. With the aim of advancing beyond general propositions and making sharp statements, it is natural to limit its scope by considering specific architectures. Therefore, in this paper we limit ourselves to modeling each agent as a simple Boolean perceptron [14] (an alternative involving Hopfield networks has been proposed in [15]), a sufficiently simple machine which affords, on one hand, the obtaining of both analytical and numerical results. On the other hand it is complex to the point of rich and interesting behavior.

Another important conceptual change with respect to other models is that issues are represented by vectors. This allows us to encompass complex issues that cannot be pinpointed by a single number but maybe can by a set of numbers, each quantifying a single feature. An issue, in this paper, is a point in a N-dimensional space. Actually we find it more difficult to defend the modeling of an issue by a single number than by a set. Consider real issues, e.g. loop gravity or strings, government control over the stock market, pregnancy termination rights, etc. Can they be represented by a single number each? More likely different aspects, such as political, cultural, religious, historical, ethical, economic, etc should be taken into account while modeling any of them. We however do not claim that we can model such complex issues with vectors, but rather make the claim that they certainly cannot be modeled by a single number. A vector is the next natural step. We should also notice that since agents debate more than an issue at a time, our model can be regarded as one of cultural dynamics, akin to the ACM. Our model is, however, distinct from the ACM in at least one very important way, namely, revealed opinions on diverse issues are the result of the application of the agent's hidden classification rule that is modified upon social interactions, changing or reinforcing these opinions.

In general terms we find that consensus is rarely attained even if agreement is strongly favored in each interaction; that disagreement is sustained locally in a probabilistic sense and when agents vote on several issues simultaneously. In the less interesting case of synchronous dynamics we find that disagreement is the norm, with consensus only emerging in some particular regimes.

This paper is organized as follows. Section 2 provides a general presentation of the model that we propose. In section 3 we analyze three simple scenarios on a social graph defined by a one-dimensional lattice with periodic boundary conditions (ring). We present our conclusions in section 4.

J. Stat. Mech. (2009) P03015

2. Opinion dynamics in a network of perceptrons

The model consists of K interacting agents. The interaction is defined by the fact that every agent can adapt its opinion formation unit by learning from the opinion of other agents. We now describe how an agent reaches its opinion on a given issue using its particular neural network.

Perceptrons are usually studied within scenarios that include learning a rule from a given data set (*supervised learning*) or extracting features from a given set of inputs (*unsupervised learning*) [14]. A mutual learning scenario, where perceptrons try to learn from each other, was first studied in [17]. We here extend this work to the context of opinion dynamics. Our main idea is that conformity related information processing may be modeled by adaptive agents trying to predict their social neighbor opinion σ on a (multidimensional) public issue $\boldsymbol{x}$ drawn from a set $\mathcal{X}$ of P issues. A model of this sort simultaneously allows for a number of features that are absent or only partially present in models currently described in the literature; that is to say, it allows for a dichotomy between internal representations (or beliefs) and revealed opinions, for the introduction of memory effects, for the definition of agreement in a probabilistic sense and for the attribution of intensities to beliefs.

The simple Boolean perceptron [14] is defined by a binary function $\sigma[\boldsymbol{x};\mathbb{J}] = \mathrm{sgn}[\mathbb{J}\cdot\boldsymbol{x}]$ that classifies input vectors $\boldsymbol{x} \in \mathbb{R}^{\mathbb{N}}$ by dividing the space into two subspaces, i.e. *for* or *against*, labeled $\{\pm 1\}$, with the dividing hyperplane specified by its normal vector $\mathbb{J} \in \mathbb{R}^{\mathbb{N}}$ (*internal representation*) as depicted in figure 1, to the left. Within the context of opinion dynamics, every agent is represented by a vector $\mathbb{J}$ that is not accessible to the other agents. Given a direction in the space of issues, the sign of the overlap between this and the internal representation determines whether the agent favors or opposes that issue and the size of this overlap determines how strongly the agent believes in their judgment.

A simple perceptron is capable of learning an unknown rule by processing a training data set $\mathfrak{D} = \{(\boldsymbol{x}^{\mu}, S_{\mu})\}_{\mu=1}^{P}$, with $S_{\mu} \in \{\pm 1\}$, containing P examples $(\boldsymbol{x}^{\mu}, S_{\mu})$ of classification by the unknown rule (*training pairs*). The learning process, however, only converges to the correct classification function if the unknown rule is linearly separable. The training pairs can be thought of as being generated by another neural network (*teacher*) or by natural phenomena being observed. The learning dynamics consists in adjusting the internal representation vector $\mathbb{J}$ by minimizing an error potential $V[\mathbb{J};\mathcal{D}] = \sum_{\mu} V_{\mu}[\mathbb{J};(\boldsymbol{x}^{\mu}, S_{\mu})]$, that compares the classification rule implemented by $\mathbb{J}$ with the data set $\mathfrak{D}$. This error potential *defines* a learning algorithm and must be chosen as robustness and optimality criteria are considered. To give an example, for a Boolean perceptron this potential can be simply defined as an error counting function $V_{\mu}[\mathbb{J};(\boldsymbol{x}^{\mu}, S_{\mu})] = \Theta[-S_{\mu}(\mathbb{J}\cdot\boldsymbol{x}^{\mu})]$, where $\Theta[x] = 1$ if $x \geq 0$, and 0 otherwise (the Heaviside step function). In the batch mode the perceptron extracts the rule from the entire data set at once; in the on-line mode the perceptron is presented to one data point at a time. The on-line dynamics that we study belongs to the scalar modulated Hebbian algorithms, given by [16]

$$\mathbb{J}\,(n+1) = \mathbb{J}\,(n) - \frac{1}{N}\nabla_{J} V_{\mu}. \tag{1}$$

By defining $\nabla_{J} V_{\mu} = -\mathcal{W}_{\mu}\boldsymbol{x}^{\mu} S_{\mu}$, *a modulation* function $\mathcal{W}_{\mu}$, that regulates the intensity of each modification to the internal representation vector, and a *Hebbian term* $\boldsymbol{x}^{\mu} S_{\mu}$, that

J. Stat. Mech. (2009) P03015

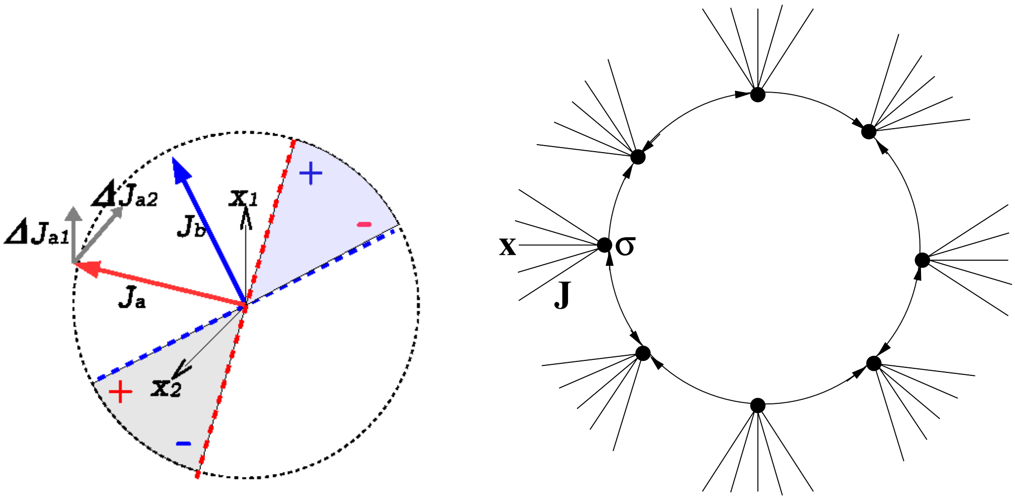

Figure 1. *Left: perceptron hyperplane and learning.* A perceptron classifies vectors ($\boldsymbol{x}_1$ and $\boldsymbol{x}_2$) in space by defining hyperplanes (thick dashed lines) normal to the internal representation vectors. In the figure two perceptrons ($\boldsymbol{J}_a$ and $\boldsymbol{J}_b$) are shown; vectors with positive (negative) overlaps are classified as $+1$ (-1). In the shaded area, perceptron classifications disagree. Vectors $\Delta\boldsymbol{J}_{a1}$ and $\Delta\boldsymbol{J}_{a2}$ represent modifications of vector $\boldsymbol{J}_a$ upon the observation that agent b agrees on classifying $\boldsymbol{x}_1$ as $+1$ but disagrees on the classification of $\boldsymbol{x}_2$. Modifying vectors are, respectively, parallel and anti-parallel to the vectors being classified with lengths proportional to the learning rate η times the modulation function $\mathcal{W}_\delta$. *Right: perceptrons on a ring.* The scenario that we study places agents on a digraph of social influences, such that a is influenced by b if b is a parent of a. In the figure each agent is represented by a perceptron (black circle) that acquires information on a public issue $\boldsymbol{x}$, classifies it following an internal representation vector $\boldsymbol{J}$ and reveals an opinion σ. Influence can be unidirectional or bidirectional, synchronous or asynchronous.

provides the direction for this vector corrections, are introduced [14]. The scale $1/N$ is chosen such that sensible macroscopic behavior is produced in the thermodynamic limit when $\boldsymbol{x}^\mu \cdot \boldsymbol{x}^\mu = \mathcal{O}(N)$.

Despite the simplicity, perceptron learning scenarios present a vast richness of different behaviors, which is derived in part from the interesting properties of the modulation function $\mathcal{W}$. For the supervised learning scenario, the modulation function has been studied exhaustively. One can determine using a variational method an optimal modulation function $\mathcal{W}^*$, in the sense that leads to the maximal decrease of generalization error [18] per example. A more fundamental approach in [19] permitted identifying $\mathcal{W}^*$ as a maximum entropy approximation to on-line Bayesian learning. Evolutionary programing studies [20, 21] showed that, under evolutionary dynamics, learning algorithms evolved to essentially the optimal algorithms. During evolution a sequence of dynamical phase transitions signaled the *discovery* of structures that lead to efficient learning. Interestingly the time order of the appearance of these structures was stable and robust, occurring in the same order whenever evolution was successful in obtaining efficient learning algorithms. This corroborated analytical work concerning temporal ordering in [22].

J. Stat. Mech. (2009) P03015

Within the class of learning algorithms described by equation (1), the static learning algorithms, which do not incorporate any annealing schedule, can be described as learning by the incorporation of two basic ingredients. These are correlation and error correction. The pure Hebbian algorithm ($\mathcal{W} = 1$) is the classic algorithm where learning is driven by incorporating correlations between input and output. The perceptron algorithm ($\mathcal{W} = 1$ if the output is in error and 0 if not) is the prime example of learning by correction of errors. There is a vast literature contrasting learning by error correction to learning by incorporating correlations in psychology [23]. In this paper we consider learning algorithms which are mixtures of these two strategies, in the context that we now describe.

We here suppose social interactions to be stable and represented by a directed graph $\mathcal{G}$ with adaptive agents, represented by perceptrons, at vertices $a \in V(\mathcal{G})$ (see figure 1, to the right). As agent a interacts, it is influenced by information that flows from parent b to child a of the arc $(b, a) \in E(\mathcal{G})$ (set of arcs of $\mathcal{G}$). This influence takes place as agent a learns on line the opinion of b, $\sigma_b(n)$, on the public issue at step n, $\boldsymbol{x}^{\mu(n)}$:

$$\mathbb{J}_a(n+1) = \mathbb{J}_a(n) + \frac{1}{N}\eta\mathcal{W}_\delta\left[h_a(n)\sigma_b(n)\right]\boldsymbol{x}^{\mu(n)}\sigma_b(n), \tag{2}$$

where η is the learning rate and $h_a(n) = \mathbb{J}_a(n) \cdot \boldsymbol{x}^\mu$ is a field that represents the degree of belief on the current opinion $\sigma_a(n) = \mathrm{sgn}[h_a(n)]$ in a manner akin to the CODA model [11]. Notice that in this scenario a training pair $(\boldsymbol{x}^{\mu(n)}, \sigma_b(n))$ is composed of the public issue being discussed and the opinion of agent b on this issue. Training pairs are provided at each social interaction, with $\sigma_b(n)$ playing the role of the classification label S_μ defined in the general supervised learning scenario described above. We consider that at each time step n one issue $\mu(n)$ is chosen for debate from the quenched set of P issues $\mathcal{X} = \{\boldsymbol{x}^\mu\}_{\mu=1}^{P}$. Information on revealed opinions can then flow through the arcs of the graph either synchronously, with all agents accessing neighbor opinions simultaneously, or asynchronously.

We assume that agents modulate learning according to their agreement $h_a(n)\sigma_b(n) > 0$ (or disagreement) before internal representations are modified. The modulation function is thus a mixture of correlation seeking and error correction defined as

$$\mathcal{W}_\delta[h_a\sigma_b] = 1 - (1-\delta)\Theta\left[h_a\sigma_b\right], \tag{3}$$

where δ represents how an agent a weights agreement with its neighbor b in relation to disagreement and $\Theta[x]$ is the Heaviside step function.

In section 3 we study three simple scenarios on a social structure defined by a one-dimensional lattice with periodic boundary conditions (a *ring*), to say: (1) a single public issue with asynchronous bidirectional information flow; (2) P simultaneous public issues with asynchronous and bidirectional information flow; (3) $P \to \infty$ simultaneous issues with synchronous and unidirectional information flow.

3. Patterns of disagreement on a ring

3.1. Single issue with asynchronous dynamics

We start by analyzing the case in which K agents influence each other asynchronously on a single issue represented by a random vector $\boldsymbol{x}$ with components drawn i.i.d. from a standard normal distribution such that $\boldsymbol{x} \cdot \boldsymbol{x} = \mathcal{O}(N)$. At each time step one arc

J. Stat. Mech. (2009) P03015

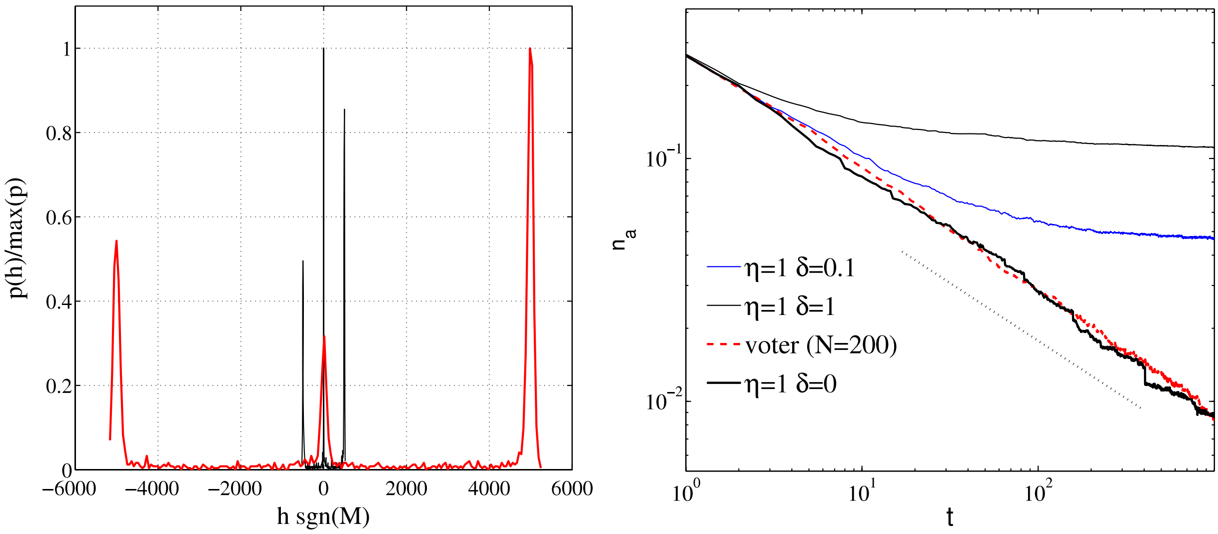

Figure 2. Figures represent five simulation runs of an asynchronous dynamics for $N = 200$ and $K = 100$ and random initial internal representation vectors with $\mathbb{J}\cdot\mathbb{J} = 1$. *Left: distribution of fields.* Fields are multiplied by the collective opinion and normalized by the maximum values. The emergence of extremist factions separated by interfaces, where fields fluctuate around $h = 0$, is revealed by a three-peaked distribution. *Right: dynamics of the density of active interfaces.* For $\delta = 0$ the system relaxes to consensus by coarsening, like for the 1D Voter model, that is also shown. The dotted line indicates an slope $t^{-1/2}$ expected in a coarsening process with non-conserved order parameter. For $\delta > 0$ the dynamics slows down as fields are mutually reinforced by the neighbors inside a faction.

(b, a) is randomly chosen and agent a is influenced by agent b. As the issue vector is quenched we can use equation (3) to rewrite equation (2) as the following (deterministic) one-dimensional map:

$$h_a(n+1) = h_a(n) + \tfrac{1}{2}\eta\,\{\delta\,[\sigma_b(n) + \sigma_a(n)] + [\sigma_b(n) - \sigma_a(n)]\}\,. \qquad (4)$$

This map is straightforward to simulate in any social graph $\mathcal{G}$; however, in this paper we will always assume that the social graph is a ring. For $\delta = 1$, agents are in a purely correlation seeking mode and we recover the CODA model [11]. For $\delta > 0$ the equilibrium state is composed by domains (or extremist factions) where fields are reinforced by agreement and interfaces that undergo a random walk resulting (for the one-dimensional ring) in a three-peaked distribution $p(h)$ with a peak in $h = 0$ produced by the interfaces and two peaks drifting away symmetrically as $h(n) \sim n\eta\delta$. In figure 2, to the left, we show an average over five field distributions obtained from the simulation of $N = 200$, $K = 100$; system for $\delta = 0.1$ and 1, with $\eta = 1$ in both cases. Initial internal representation vectors are chosen to be random with $\mathbb{J}\cdot\mathbb{J} = 1$. In order to keep results comparable among systems of different size K, t is measured in MC steps defined as $t = n/K$. Distributions are normalized by maximum values and multiplied by the collective opinion given by the sign of $M = \sum_{a=1}^{K}\sigma_a$, to avoid artificially symmetrizing the result upon averaging.

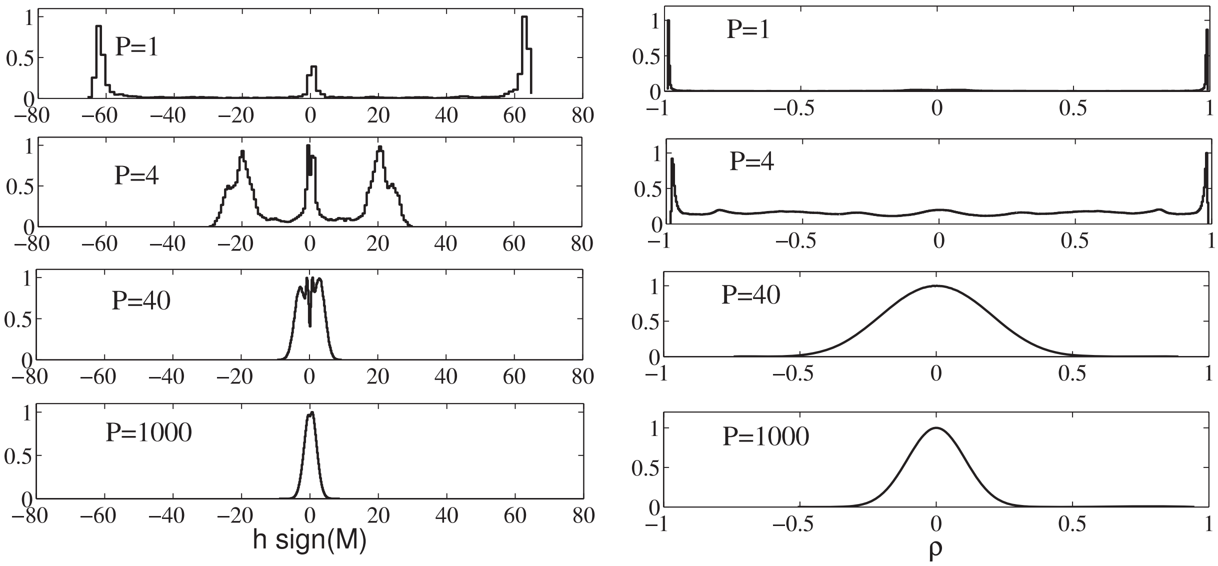

Figure 3. Figures show the results of a single run of simulations of asynchronous dynamics with $P = 1, 4, 40$ and 1000, $N = 100$, $K = 10^4$, $\eta = 1$ and $\delta = 0.01$. Measurements are taken for $t = 10^6$ MC steps measured as $t = n/K$. *Left: distribution of fields with number of issues.* The histograms show the distribution of belief fields (h_a). Distributions are normalized by maximum values and multiplied by the collective opinion, given by the sign of $M = \sum_{a=1}^{K} \sigma_a$, to avoid artificially symmetrizing the result upon averaging. As P increases, the dynamics slows down and opinions become less extreme inside factions. *Right: distribution of overlaps with number of issues.* Opinions become more diverse as P increases. For $P = 1000$ the distribution corresponds to what is expected to be seen in the case of K random vectors.

In the case $\delta = 0$ agents are purely error correcting and the system is driven towards consensus by coarsening (for a social structure in one dimension) with domains growing with $\sqrt{t}$, like in Voter dynamics [24]. In figure 2, to the right, we show the simulated evolution (average over five runs) over time of the density of active interfaces defined as $n_a = (1 - \langle \sigma_a \sigma_b \rangle)/2$, with $(b, a) \in E(\mathcal{G})$, for $\delta = 0$, 0.1 and 1 (CODA model). For $\delta > 0$ the system is still driven to order by coarsening; however, the dynamics becomes glassy due to mutual reinforcing inside factions, leading to $n_a(t \to \infty) > 0$ and asymptotically frozen domains.

3.2. Set of P issues with asynchronous dynamics

In the second scenario that we analyze, a (quenched) set $\mathcal{X} = \{\boldsymbol{x}^\mu\}_{\mu=1}^{P}$ of vectors, with components sampled i.i.d. from a standard normal distribution, represents public issues to be debated. At each time step n one issue $\boldsymbol{x}^{\mu(n)} \in \mathcal{X}$ and one arc (b, a) are randomly chosen. Agent a, then, tries to learn on-line agent b's internal representation $\mathbb{J}_b$, employing the learning dynamics described by equation (2).

The restricted quenched set of issues leads to a dynamics that is not trivially amenable to analytical treatment. In order to get some insight we have, therefore, relied on a number of simulations on a 1D ring. The general behavior that emerges can be seen on figure 3

J. Stat. Mech. (2009) P03015

(left) that shows, for $P = 1, 4, 40$ and 1000, the distribution of fields multiplied by the collective opinion for a single run with $N = 100$, $K = 10^4$, $\delta = 0.01$ and $\eta = 1$. The averages are taken at time $t = 10^6$ MC steps measured as $t = n/K$. If agents are correlation seekers in some measure ($\delta > 0$), identifiable extremist factions emerge for a number P of simultaneous issues that is not too large. However, as P increases, extremism via field reinforcement inside a faction weakens. Given that the set of issues is produced randomly and that $P \ll N$, the probability that two agents will disagree on a given issue is well defined by $\mathbb{P}\{a \text{ disagrees with } b\} = (1/\pi) \arccos[\rho_{ab}]$, where $\rho_{ab} = (\mathbb{J}_a \cdot \mathbb{J}_b)/(J_a J_b)$ is the overlap between a and b; J_a and J_b being vector norms. In figure 3, to the right, we show histograms for overlaps under the same conditions as are described in the left panel of the same figure. For P not too large, agents either agree or disagree with certainty, which produces strong mutual reinforcement effects within a faction. As P increases, internal representations diversify and the probability of disagreement increases, reducing extremist trends. For P very large, no convergence of opinions is observed and the distribution of overlaps is the same expected for random vectors with N dimensions (i.e. a dispersion that is $\mathcal{O}(1/\sqrt{N})$ is observed).

No phase transition with P as control parameter takes place though. Instead, the dynamics slows down continuously with $1/\sqrt{P}$. A simple argument shows that this is so. Inside a faction, two agents agree on most of the issues, which implies a positive overlap with neighbors, as depicted in figure 4, to the left. As agent a is influenced by agent b, it corrects its internal representation following the learning dynamics of equation (2), which implies that $\mathbb{J}_a$ is corrected by $(1/N)\eta\sigma_b \boldsymbol{x}^{\mu(n)}$, to say, by a vector that is parallel or anti-parallel to the issue depending on the opinion of agent b being $\sigma_b = 1$ or $\sigma_b = -1$, respectively. The overlap changes by the average displacement due to the components of $(1/N)\eta\sigma_b \boldsymbol{x}^{\mu(n)}$ orthogonal to the direction defined by the internal representation $\mathbb{J}_a$. Considering that the dimensionality of the issue space N is very large ($N \to \infty$ in the thermodynamic limit), the P random issues can be considered to be orthogonal to order $1/\sqrt{N}$. The average displacement due to a sequence of random issues, therefore, scales with $1/\sqrt{P}$.

The collective result of the mutual learning of opinions is the alignment inside a faction of internal representations to a direction defined by the average issue $\bar{\boldsymbol{x}} = (1/P)\sum_{\mu=1}^{P} \boldsymbol{x}^\mu$. The typical alignment of agent a's internal representation and the average issue is $\rho_{ax} = (\mathbb{J}_a \cdot \bar{\boldsymbol{x}})/(J_a \bar{x})$, with the mutual learning dynamics being symmetrical with respect to the hyperplane defined by the average issue. This alignment as a function of P can be estimated as follows. The norm of the internal representation grows as $J_a \sim \sqrt{n/N}$ and the norm of the average issue scales as $\bar{\boldsymbol{x}} \sim \sqrt{N/P}$. On the other hand, the scaling of the overlap $\mathbb{J}_a \cdot \bar{\boldsymbol{x}}$ is estimated as

$$\mathbb{J}_a \cdot \left(\frac{1}{P} \sum_{\mu} \boldsymbol{x}^\mu \right) = \frac{1}{N} \sum_{j=1}^{n} \eta \sigma_b(j) \boldsymbol{x}^{\mu(j)} \cdot \left(\frac{1}{P} \sum_{\mu} \boldsymbol{x}^\mu \right)$$
$$= \frac{1}{NP} \sum_{j=1}^{n} \eta \sigma_b(j) \left(\boldsymbol{x}^{\mu(j)} \cdot \sum_{\mu} \boldsymbol{x}^\mu \right) \sim \frac{1}{P} \sum_{j=1}^{n} \sigma_b(j) \sim \frac{\sqrt{n}}{P}. \tag{5}$$

Putting everything together we can finally see that $\rho_{ax} \sim (1/\sqrt{P})$. In the right panel of figure 4, we plot the result of five simulations for $|\rho_{ax}|$ averaged over the vertices,

J. Stat. Mech. (2009) P03015

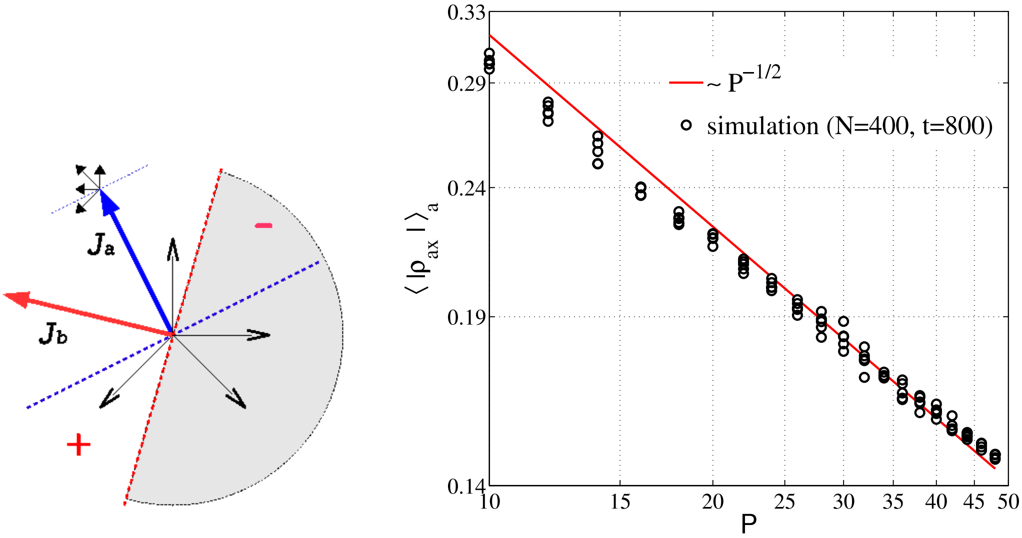

Figure 4. *Left: mutual learning representation.* Inside a faction, overlaps are positive and corrections to the internal representation $\mathbf{J}_a$ due to learning are parallel or anti-parallel to issues depending on the opinion of agent b (represented by $\mathbf{J}_b$). In the figure the shaded area indicates the half-space where $\sigma_b = -1$. Overlaps between agents only change during learning due to issue components that are orthogonal to the internal representation being influenced. The hyperplane orthogonal to $\mathbf{J}_a$ is represented by a thin dashed line in the figure. In the thermodynamic limit $N \to \infty$ the average displacement due to a sequence of P random issues scales with $1/\sqrt{P}$. *Asynchronous dynamics. Right: average overlap to the mean issue as a function of the number of issues.* Five simulations (circles) with $N = 400$, $K = 100$, $\eta = 1$ and $\delta = 1$ are compared to the theoretical prediction for the scaling with the number of issues P. Opinions become more diverse as the number of simultaneous issues being debated increases.

where we have dropped the sign to deal with the learning dynamics symmetry that allows for two factions with $\pm\rho_{ax}$. In the same figure we also plot a curve $\sim 1/\sqrt{P}$ for comparison. Though not precise, the general agreement is reasonable considering that we have ignored non-trivial correlations and finite size effects in the argument above.

Following along the same lines as above it is also interesting to notice that the dispersion of fields $h_a(n) = \mathbb{J}_a \cdot \boldsymbol{x}^{\mu(n)}$ depicted in figure 3, to the left, scales as $\sigma_h \sim \sqrt{n/P}$. It is, therefore, possible to rescale the dispersion of fields by redefining time as $n_P = n/P$. Both the mean field and the mean overlap, however, still decay with $1/\sqrt{P}$.

Considering that time is not rescaled with the number of issues, namely, that the time allowed for discussion is *not* a function of the number of issues being discussed, we expect that, inside a faction, agent beliefs become less extreme. The general message is also independent of the exact decay of $|\rho_{ax}|$. As the number of issues being simultaneously debated increases, we expect that inside a faction agent beliefs become not only less extreme but also more diverse.

3.3. Synchronous dynamics with random issues

In the third scenario we propose that the social graph is again a one-dimensional lattice with arcs of the form $E(\mathcal{G}) = \{(a, a+1(\mathrm{mod}\ K)), a = 1, \ldots, K\}$ and that information flows unidirectionally between nodes with internal representations being updated at once in the whole graph as public random issues are presented. In this case, at least in principle, semi-analytical solutions are available for P finite along the lines of [25]. For $P \to \infty$ the scenario is amenable to semi-analytical treatment by a direct adaptation of [14, 16, 17] that we summarize in this section.

The technique relies upon the assumption that the macroscopic properties of the system, defined by the overlaps $R_{ab} = \mathbb{J}_a \cdot \mathbb{J}_b$, are self-averaging in the thermodynamic limit $(N \to \infty)$ [26] and an integration over the distribution of $P \to \infty$ random issues yields a system of deterministic differential equations. By supposing that at each time step n an issue $\boldsymbol{x}^n$ is produced with components sampled i.i.d. from a standard normal distribution we can write from equation (2) difference equations for the overlaps. By scaling time as $\alpha = n/N$ and averaging over the issues we can write a system of $\frac{1}{2}(K^2 + K)$ differential equations whose integrals should describe the macroscopic evolution of the system, given that the self-averaging property holds. Clearly, the same procedure can be implemented to describe synchronous dynamics on arbitrary graphs $\mathcal{G}$. On a ring with unidirectional information flow, the first $\frac{1}{2}K(K-1)$ equations describing internal representation overlaps are given by

$$\begin{aligned}\frac{\mathrm{d}R_{ab}}{\mathrm{d}\alpha} = {} & \frac{\eta}{2}(\delta+1)\left[\langle h_a\sigma_{b+1}\rangle + \langle h_b\sigma_{a+1}\rangle\right] + \frac{\eta}{2}(\delta-1)\left[\langle h_b\sigma_a\rangle + \langle h_a\sigma_b\rangle\right] \\ & + \frac{\eta^2}{4}(\delta-1)^2\langle\sigma_a\sigma_b\rangle + \frac{\eta^2}{4}(\delta+1)^2\langle\sigma_{a+1}\sigma_{b+1}\rangle \\ & + \frac{\eta^2}{4}(\delta+1)(\delta-1)\left[\langle\sigma_a\sigma_{b+1}\rangle + \langle\sigma_b\sigma_{a+1}\rangle\right],\end{aligned} \tag{6}$$

where $\langle\cdots\rangle$ denotes averages over the distribution of issues $\boldsymbol{x}$ that, for our choice, can be calculated analytically to give

$$\begin{aligned}\langle\sigma_a\sigma_b\rangle &= 1 - \frac{2}{\pi}\arccos\left(\frac{R_{ab}}{J_a\,J_b}\right), \\ \langle h_a\sigma_b\rangle &= \sqrt{\frac{2}{\pi}}\frac{R_{ab}}{J_b},\end{aligned} \tag{7}$$

with $J_a = \sqrt{\mathbb{J}_a \cdot \mathbb{J}_a}$. The remaining K equations describe the evolution of the norms of internal representations, yielding

$$\begin{aligned}\frac{\mathrm{d}J_a}{\mathrm{d}\alpha} = {} & \frac{\eta}{\sqrt{2\pi}}(\delta-1) + \frac{\eta^2}{2J_a}\delta^2 + \frac{\eta}{\sqrt{2\pi}}\frac{(\delta+1)}{2}\left[\frac{R_{(a-1)a}}{J_{a-1}\,J_a} + \frac{R_{a(a+1)}}{J_a\,J_{a+1}}\right] \\ & - \frac{\eta^2}{4\pi J_a}\left(\delta^2-1\right)\left[\arccos\left(\frac{R_{(a-1)a}}{J_{a-1}\,J_a}\right) + \arccos\left(\frac{R_{a(a+1)}}{J_a\,J_{a+1}}\right)\right],\end{aligned} \tag{8}$$

when the averages in equation (7) are appropriately replaced.

Along the lines of [17], the fixed point of equation (8) can be found for the case of agents that are pure error correctors ($\delta = 0$). By observing that in this case a synchronous

J. Stat. Mech. (2009) P03015

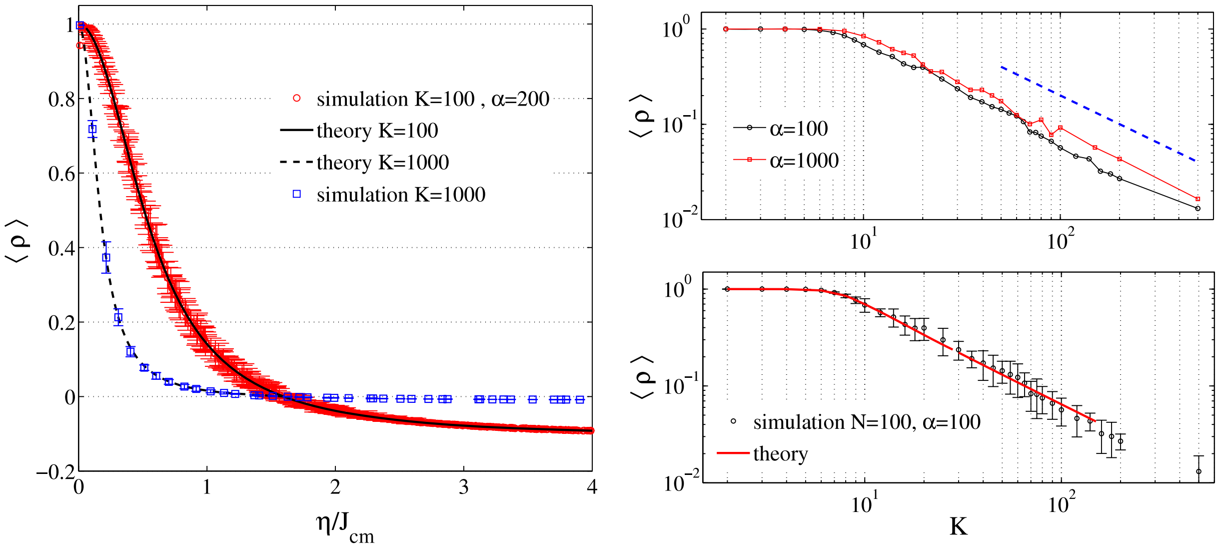

Figure 5. *Synchronous dynamics. Left: average overlap versus learning rate for error correctors* ($\delta = 0$). Symbols show overlaps averaged over arcs $(a, b) \in E(\mathcal{G})$ in ten simulations with $N = 100$, $K = 100$ (circles) and $K = 1000$ (squares) at $\alpha = 200$. Lines represent the fixed points given by equation (10). Consensus is favored by small learning rates ($\eta \to 0$). *Right: average overlap versus system size for correlation seekers* ($\delta = 1$). Top: averages are over vertices and 10 simulation runs with $N = 100$, $\eta = 1$, $\delta = 1$ at, from the bottom up, $\alpha = 100$ (black) and $\alpha = 1000$ (red). The dashed line indicates the slope of K^{-1}. Bottom: numerical integration of equations (6) and (8) compared to simulations at $\alpha = 100$. Consensus is attained even for $P \to \infty$, given that the system size is not too large. For large K the asymptotic average overlap scales as $\langle \rho \rangle \sim 1/K$.

dynamics preserves the center of mass $\mathbb{J}_{\rm cm} = (1/K) \sum_{a=1}^{K} \mathbb{J}_a$ and by supposing a symmetric evolution, with $R = R_{ab}$ and $J = J_a$ for all a and b, we can rewrite equation (8) as

$$\frac{\mathrm{d}J}{\mathrm{d}\alpha} = -\frac{\eta}{\sqrt{2\pi}} + \frac{\eta}{\sqrt{2\pi}} \frac{R}{J^2} + \frac{\eta^2}{2\pi J} \arccos\left(\frac{R}{J^2}\right), \qquad (9)$$

which has a fixed point given by

$$\frac{\eta}{J_{\rm cm}} = \frac{\sqrt{2\pi}}{\arccos(\rho)} \frac{1-\rho}{\sqrt{1+(K-1)\rho}}, \qquad (10)$$

where $\rho = R/J^2$. In the left panel of figure 5, we show the average overlap as a function of the normalized learning rate $\eta/J_{\rm cm}$. Ten simulations (symbols) with $N = 100$, $K = 100$ (circles) and $K = 1000$ (squares) at $\alpha = 200$ are shown and compared to fixed points predicted by equation (10). Interesting features of the synchronous case are that consensus of error correcting agents is fostered by slow learning ($\eta \to 0$), which is also observed in [27], and that a system of adaptive agents under this condition tend towards a random distribution of opinions ($\rho \to 0$) as the system grows larger.

For $\delta > 0$ internal representations are constantly reinforced by confirmation inside a faction and no fixed point can be found for equation (8). To simplify the analysis we

J. Stat. Mech. (2009) P03015

assume the extreme case of agents that are pure correlation seekers $\delta = 1$. Introducing $\lambda_a = \eta/J_a$ and $\rho_{ab} = R_{ab}/(J_a J_b)$ and considering the limit of long times (λ_a small) leads to

$$\frac{\mathrm{d}\rho_{ab}}{\mathrm{d}\alpha} \approx \sqrt{\frac{2}{\pi}}\lambda_a \left(\rho_{(a+1)b} - \rho_{ab}\rho_{a(a+1)}\right) + \sqrt{\frac{2}{\pi}}\lambda_b \left(\rho_{a(b+1)} - \rho_{ab}\rho_{b(b+1)}\right) \tag{11}$$

and

$$\frac{\mathrm{d}\lambda_a}{\mathrm{d}\alpha} \approx -\sqrt{\frac{2}{\pi}}\lambda_a^2 \rho_{a(a+1)}. \tag{12}$$

These equations imply that the state of consensus ($\rho_{ab} = 1$ for all a and b) is always a fixed point; however, the dynamics slows down with λ_a as the system drifts towards the fixed point defined by an arbitrary distribution of overlaps and $\lambda_a = 0$.

In figure 5, to the top right, we show the average overlaps (over vertices and over ten runs) as a function of the system size K for simulations with $N = 100$, $\eta = 1$, $\delta = 1$ at, from the bottom up, $\alpha = 100$ and 1000. In figure 5, to the bottom right, we show a comparison between the result of simulations at $\alpha = 100$ and the numerical solution for equations (6) and (8). In contrast with the asynchronous case, if the update is synchronous, consensus can be attained even for $P \to \infty$, given that the system size K is not too large. For large K the asymptotic average overlap decays as $\langle \rho \rangle \sim 1/K$ (dashed line to the top right of figure 5).

4. Conclusions

A central puzzle in the study of information processing in social systems is the persistence of disagreement in a system composed by consensus seekers. This puzzle can be solved by considering a dichotomy between internal representations and revealed opinions, as the introduction of memory effects may allow for competition between a local trend toward uniform opinions and long range heterogeneity. On top of that, in the model for opinion dynamics with learning that we have introduced, agreement can be attained in one issue while disagreement persists in other issues.

In the model that we have proposed, agents classify any issue presented to discussion by using a single internal representation $\mathbb{J}$ which implies the strong presupposition that agents are always consistent when issuing opinions. It seems more realistic also considering cases with multiple internal representations or with noisy decision processes allowing for inconsistent judgments. Such directions, however, still have unclear consequences and might be pursued in further work.

We have shown that some simple instances of the model are amenable to analytical treatment with a number of non-trivial phenomena emerging. For a single issue with asynchronous update of information and a one-dimensional social structure, we observe behavior akin to that of the Voter model as long as agents are pure error correctors ($\delta = 0$), with full consensus being reached by a coarsening process. When agents are also correlation seekers ($\delta \neq 0$), extremist factions with opposite opinions survive due to the mutual reinforcement of confirmed beliefs. That is, extreme disagreement becomes a characteristic of the system, even in the long run. The survival of disagreement, however, comes in this case at the cost of the appearance of extremist agents.

By introducing a number of issues to be simultaneously debated, we observe that the tendency towards extremism weakens. This seems to suggest that a richer cultural setting, where the agents debate a larger number of different subjects, is less likely to develop extremists than a setting where very few issues (or highly correlated issues) are debated. As the number of issues grows larger, internal representations converge to distributions that are indistinguishable from random vectors in a space of equal dimensionality. Disagreement also survives in this case, but with the tendency to extremism tamed.

We also have investigated semi-analytically the case of synchronous information update with a very large number of simultaneous issues. A dynamics distinct from the asynchronous case is observed, with consensus being fostered either by slowing down the adaptation process, when agents are error correctors, or by allowing agents to be correlation seekers, given that the system is not too large.

Acknowledgments

This work has been funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), under grant 550981/2007-1 (RV), and by Fundação de Amparo á Pesquisa do Estado de São Paulo (FAPESP), under grant 2008/00383-9 (ACRM). We also wish to thank the anonymous referee for the thoughtful suggestions which contributed to improvement of our manuscript.

References

[1] Huckfeldt R, Johnson P E and Sprague J, 2004 *Political Disagreement: The Survival of Diverse Opinions within Communication Networks* (Cambridge: Cambridge University Press)
[2] Sunstein C R, 2005 *Why Societies Need Dissent* (Cambridge, MA: Harvard University Press)
[3] Huckfeldt R *et al*, 1995 *Am. J. Polit. Sci.* **39** 1025
[4] Castellano C, Fortunato S and Loreto V, 2007 arXiv:0710.3256 [physics.soc-ph]
[5] Galam S, Gefen Y and Shapir Y, 1982 *J. Math. Sociol.* **9** 1
[6] Clifford P and Sudbury A, 1973 *Biometrika* **60** 581
[7] Holley R and Liggett T M, 1975 *Ann. Probab.* **3** 643
[8] Sznajd-Weron K and Sznajd J, 2000 *Int. J. Mod. Phys.* C **11** 1157
[9] Latané B, 1981 *Am. Psychol.* **36** 343
[10] Lewenstein M, Nowak A and Latané B, 1992 *Phys. Rev.* A **45** 763
[11] Martins A C R, 2008 *Int. J. Mod. Phys.* C **19** 617
[12] Axelrod R, 1997 *J. Conf. Resolut.* **41** 203
[13] Castellano C, Marsili M and Vespignani A, 2000 *Phys. Rev. Lett.* **85** 3536
[14] Engel A and Van den Broeck C, 2001 *Statistical Mechanics of Learning* (Cambridge: Cambridge University Press)
[15] Stauffer D, Grabowicz P A and Holyst J A, 2008 arXiv:0712.4364v2 [physics.soc-ph]
[16] Vicente R, Kinouchi O and Caticha N, 1998 *Mach. Learn.* **32** 179
[17] Metzler R, Kinzel W and Kanter I, 2000 *Phys. Rev.* E **62** 2555
[18] Kinouchi O and Caticha N, 1992 *J. Phys. A: Math. Gen.* **25** 6243
[19] Opper M, 1996 *Phys. Rev. Lett.* **77** 4671
[20] Neirotti J P and Caticha N, 2003 *Phys. Rev.* E **67** 041912
[21] Caticha N and Neirotti J P, *The evolution of learning systems: to Bayes or not to be*, 2006 *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, 26th Int. Workshop* ed A Mohammad-Djafari
[22] Caticha N and Kinouchi O, 1998 *Phil. Mag.* B **77** 1565
[23] Houghton G (ed), 2005 *Connectionist Models in Cognitive Psychology* (Hove/New York: Psychology Press)
[24] Redner S, 2001 *A Guide to First-Passage Processes* (Cambridge: Cambridge University Press)
[25] Coolen A C C and Saad D, 2000 *Phys. Rev.* E **62** 5444
[26] Reents G and Urbanczik R, 2000 *Phys. Rev. Lett.* **80** 5445
[27] Stark H-U, Tessone C J and Schweitzer F, 2008 *Phys. Rev. Lett.* **101** 018701