Available online at www.sciencedirect.com
SCIENCE DIRECT
Physica A 353 (2005) 555–575
PHYSICA A
www.elsevier.com/locate/physa

Persuasion dynamics

Gérard Weisbuch[a,1], Guillaume Deffuant[b], Frédéric Amblard[a,b,*]

[a] *Laboratoire de Physique Statistique,[2] de l'Ecole Normale Supérieure, 24 rue Lhomond, F-75231 Paris Cedex 5, France*
[b] *Laboratoire d'Ingénierie pour les Systèmes Complexes (LISC), Cemagref - Grpt de Clermont-Ferrand, 24 Av. des Landais - BP50085, F-63172 Aubière Cedex, France*

Received 8 October 2004; received in revised form 12 January 2005
Available online 8 March 2005

Abstract

We here discuss a model of continuous opinion dynamics in which agents adjust continuous opinions as a result of random binary encounters whenever their difference in opinion is below a given threshold. We concentrate on the version of the model in the presence of few extremists which might drive the dynamics to generalized extremism. A network version of the dynamics is presented here, and its results are compared to those previously obtained for the full-mixing case. The same dynamical regimes are observed, but in rather different parameter regions. We here show that the combination of meso-scale features resulting from the first interaction steps determines the asymptotic state of the dynamics.

PACS: 87.23.Ge; 89.75.Hc

Keywords: Social dynamics; Opinion formation; Social networks

*Corresponding author. Tel.: +33 1 44 32 34 75; fax: +33 1 44 32 34 33.
E-mail address: weisbuch@lps.ens.fr (G. Weisbuch).
[1] Presently at UFR Informatique, Université de Sciences Sociales de Toulouse (UT1), 21 allées de Brienne, 31042 Toulouse, France.
[2] Laboratoire associé au CNRS (URA 1306), à l'ENS et aux Universités Paris 6 et Paris 7.

doi:10.1016/j.physa.2005.01.054

1. Introduction

The present paper is a follow up in a series of publications on continuous opinion dynamics and ‘‘extremism’’ [1].

Social psychology is often concerned with the outcome of collective decision processes in connection with individual cognitive processes and the actual dynamics of opinion exchanges in meetings. The issue of whether extremist or moderate opinions are adopted in committees is thoroughly discussed in Moscovici and Doise [2].

A connection with statistical physics and the Ising model was established, for instance, by Galam [3] who collaborated with Moscovici. They considered binary opinions as in the vast majority of the literature on binary social choice [4].

The approach here is different and rests on the fact that certain choices imply continuous opinions; typical examples are the evaluation of economic profits among different possible choices [5], or how to share profits after a collective enterprise (hunt, agriculture etc.) [6]. In the early literature on committees, opinions were simply supposed to influence each other in proportion to their difference. The described dynamics was equivalent to heat diffusion, and resulted in uniformization around some average opinion.

The notion of an interaction threshold, based on experimental social psychology, was proposed by Chattoe and Gilbert [7], and introduced in models by Deffuant et al. [8] and Hegselmann and Krause [9]. Two individuals with different opinions only influence each other when their difference in opinion is lower than a threshold. The outcome of the dynamics can then be clustering rather than uniformity. (A series of models of cultural diffusion first introduced by Axelrod and followers [10] belong to the same class: cultures are represented by vectors of integers which are brought closer by interactions under certain conditions of similarity; these authors studied how these conditions influence the outcome of the dynamics, uniformity versus diversity. Integer variables facilitate analytical approaches [10,11] via master equations.)

Fascinating results were obtained in the ‘‘extremism’’ model of Deffuant et al. [1]: when interaction thresholds are unevenly distributed, and in particular when agents with extreme opinions are supposed to have a very low threshold for interaction, extremism can prevail, even when the initially extremist agents are in very small proportion.[3] The so-called ‘‘extremist model’’ can be applied to political extremism, and a lot of the heat of the discussion generated by these models relates to our everyday concerns about extremism. But we can think of many other situations where some ‘‘inflexible’’ agents are more sure about their own opinion than others. Inflexibility can arise for instance

- because of knowledge; some agents might know the answer while others only have opinions; think of scientific knowledge and the diffusion of new theories;
- some agents might have vested interests different from others.

[3] Hegselmann and Krause [9] also implicitely propose a different ‘‘road’’ to extremism.

Although the model has some potential for many other applications, we will here use the original vocabulary of extremism.

Several subsequent papers [12] checked the genericity of these results for different interaction topologies (well-mixed systems versus social networks represented by many variants between lattices and random nets) and for different variants of the distribution of interaction intensities (see further Eq. (3)). Clustering and the possibility of extremist attractors were shown to be generic, but the phase diagrams between different dynamical regimes can be rather intricate with co-existence regions depending upon parameter values.

The purpose of the present paper is to increase our understanding of these phase diagrams obtained through systematic scanning of the parameter space, by using here simpler conditions for simulation and direct monitoring of the dynamics of single samples. (Unfortunately, we are still not very advanced in formal derivations.) The subsequent section describes the models, the simulation techniques and the monitoring of the results. We first deal with models where only one "extremist" is present. Full mixing and lattice topologies are studied. More intricate situations with many extremists can be understood from the one-extremist case. Preliminary results obtained for scale-free network topologies are then presented.

2. Models and simulations

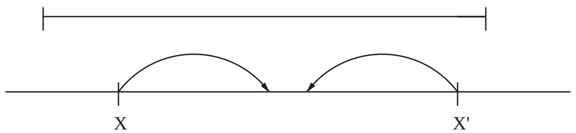
2.1. The basic model

The most basic model, later called bounded confidence model, was introduced by Deffuant et al. [8]. It supposes an initial distribution of agents with scalar opinion x.

At each time step, any two randomly chosen agents meet: they readjust their opinion when their difference in opinion is smaller in magnitude than a threshold d. Suppose the two agents have opinions x and x'. *Iff* $|x - x'| < d$ opinions are adjusted according to

$$x = x + \mu \cdot (x' - x)\,, \tag{1}$$

$$x' = x' + \mu \cdot (x - x')\,, \tag{2}$$



where μ is the convergence parameter whose values may range from 0 to 0.5. The rationale for the threshold condition is that agents only interact when their opinions are already close enough; otherwise they do not even bother to discuss.

The basic model uses

- a threshold d constant in time and across the whole population,
- a complete mixing hypothesis, and
- a random serial iteration mode.

The threshold can be interpreted as "openness", tolerance or as some uncertainty in opinion.

The choice of the random serial iteration mode models opinion diffusion in a large population which agents encounters each other in small groups such as pairs. In contrast, Hegselmann and Krause [9] chose parallel iteration since their approach is derived from earlier literature modeling formal meetings.

Computer simulations show that the distribution of opinions evolves at large times towards clusters of homogeneous opinions (both iteration modes yield similar clusters under the used conditions).

For large threshold values ($d > 0.3$) only one single cluster is observed at the average initial opinion (consensus). For lower threshold values, several large clusters are observed. Consensus is then NOT achieved when thresholds are low enough. The number of clusters varies as the integer part of $1/2d$ [8], to be further referred to as the $1/2d$ rule.

Some recent literature by Stauffer and collaborators [18] consider any group of opinions as clusters, however small (even of size one). Counting all these groups yield higher figures scaling with N the number of agents. We here only monitor large clusters whose size is a finite fraction of the number N of agents. We do not care about the existence of isolated "outliers" (because of the randomness of the iteration process some agents are selected at later times and remain as "outliers" outside the main clusters). The "generic" results we here refer to, such as the $1/2d$ rule, apply to the large clusters.

Rewriting the opinion updating equation as

$$x = x + \mu \cdot f(x' - x) \cdot (x' - x)\,, \tag{3}$$

the bounded confidence model supposes a square amplitude of the interaction function $f(x' - x)$ when $|x - x'| < d$. Smoother shapes (such as trapezoidal [1] or bell-shaped [13]) were also proposed for $f(x' - x)$. The simulations show that the main dynamical features are conserved with these smoother interaction functions.

2.2. Extremism

The model for extremism introduced by Deffuant et al. [1] is based on two more assumptions.

- A few extremists with extreme opinions at the end of the opinion spectrum and with very low threshold for interaction are introduced.

- Whenever the threshold allows interaction, both opinions and threshold are readjusted according to similar expressions.

Iff $|x - x'| < d$,

$$x = x + \mu \cdot (x' - x)\,, \tag{4}$$

$$d = d + \mu \cdot (d' - d)\,. \tag{5}$$

A symmetrical condition and equations apply to the other agent of the pair with opinion x' and tolerance d' but when thresholds are different the influence can be asymmetric: the more "tolerant" agent (with larger d) can be influenced by the less tolerant (with smaller d) while the less tolerant agent is not. This "effective" asymmetry is responsible for the outcome of "extremist" attractors.

2.3. Simulation methods and displays

Computer simulations are run according to standard conditions:

- Initial conditions: uniform distribution of opinions among $[0, 1]$ among N agents with initial threshold dl. Among these a few agents are extremists, with opinions at the extreme of the opinion spectrum and with initial threshold $de \ll dl$.
- At each time step, one randomly selected pair is chosen and agents are updated according to Eqs. (4)–(5) whenever the condition on threshold is fulfilled.
- Simulations are run until an approximate state of equilibrium is reached. Histograms of agents opinion and tolerance are monitored during the simulation and we here considered that equilibrium is reached when opinion and tolerance histograms of 101 bins are stable.

The main parameters are the number and initial tolerance of extremists, and the initial tolerance dl of the other agents. Variants include different interaction networks, and different interaction functions $f(x' - x)$.

We usually first check opinion and tolerance dynamics by time plots of single simulations [1]. These time plots are clouds of points representing at each time step the opinions and tolerance of those agents chosen for eventual updating versus time along the x-axis.

The time plots display different dynamical regimes according to the eventual predominance of the extremists: sometime they remain isolated and most agents cluster as if there were no extremist (e.g., as represented in Fig. 1, left frame); otherwise extremism prevails and most agents cluster in the neighbourhood of one (e.g., as represented in Fig. 1, right frame) or both extreme.

Still, in both cases, a phase of convergence towards average opinion of most initially centrist agents is observed (for roughly 10 updating per agents). The initial convergence towards the centre is due to the much larger number of centrists as compared to extremists. After this preliminary phase, the centre clustered agents can either slowly evolve towards extremism if they still feel their influence, in other

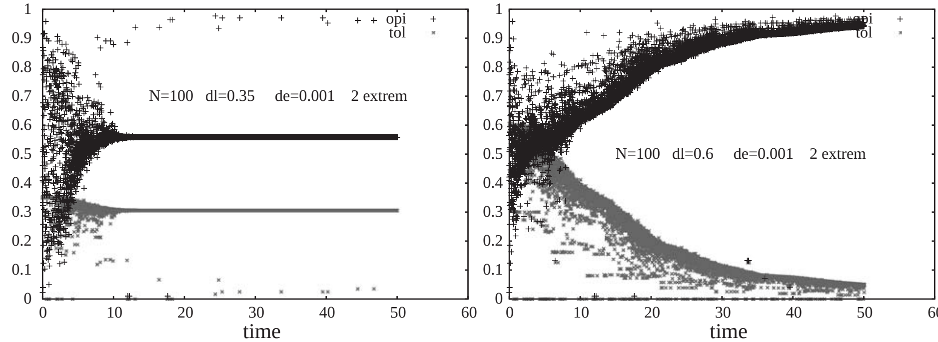


Fig. 1. Time plots of opinion (black '+') and tolerance (grey '×') dynamics exhibiting a "centrism" attractor on the left frame and a single extremist cluster on the right frame. Time is given in the average number of updating per agent. The number of agents is $N = 100$, extremists' tolerance is $de = 0.001$ and two opposite extremists are initially present. Any pair of agents can a priori interact. The centrism attractor was obtained when the initial centrist tolerance was $dl = 0.35$ and the extremist attractor when $dl = 0.6$.

words, when their tolerance is larger than their distance to extremists ($dl > 0.5$); otherwise (when $dl < 0.5$) they are not anymore under extremists influence and remain at the centre. Due to the random character of the initial opinion distribution and pair sampling, the condition $dl = 0.5$ is not a sharp boundary, but, rather indicates a dynamical crossover.

Convergence characteristic time then differs: convergence is fast for the centrism attractor and slow for the extremism attractor. The ratio in convergence time is approximately the ration in the initial fraction of centrists and extremists.

More generally, which attractor is reached depends mainly upon the parameters of the simulation (number and initial tolerance of extremists, and the initial tolerance dl of the other agents). A sketchy conclusion is that some kind of extremism prevails for larger values of the tolerance of initially non-extremist agents when $dl > 0.5$, and centrism when $dl < 0.5$. In other words, the outcome of the dynamics is largely determined by the tolerance of the non-extremists agents. But systematic studies show *co-existence* parameter regions where several attractors can be reached depending upon the specific initial distribution of opinions and upon the specific choice of updated pairs.

Deffuant et al. [1] papers are filled with two dimensional regime diagrams coded according to a variant of the Derrida–Flyvbjerg parameter [14] defined as

$$Y = \sum_i \left(\frac{n_i}{N}\right)^2 \tag{6}$$

This sum of the square of the fraction of number n_i of agents in each cluster i roughly represents the inverse of a weighted number of clusters. Particular choices of monitored Derrida–Flyvbjerg parameters allow to separate dynamical regimes of attraction towards the centre, from clustering and attraction towards one or both extremes; co-existence regions that display several regimes according to initial conditions are characterized by large Y variance over initial conditions.

But these diagrams although comprehensive in terms of parameter ranges and averaging over many initial conditions are difficult to interprete and we here use a more direct approach. Rather than averaging some index over many simulations, we directly check asymptotic opinion and tolerance histograms of individual simulations. We then only vary one parameter along the x-axis, most often the initial large tolerance dl. The y-axis code the histogram of attractor clusters by vertical bars. The magnitude of the bar represents how many agents are in the asymptotic cluster(s). Clusters made of one agent are discarded to make diagrams more readable. The position of the bar represents either the opinion or the tolerance of agents in that cluster. Each bar only gives the result of one simulation.

Co-existence regions appear as dl intervals on which large fluctuations are observed in the cluster positions. Probabilities of either regime are evaluated from their frequency of observation on any interval. In contrast, pure regimes yield regular variations of cluster positions.

Further understanding of the dynamics is brought by on-line observation of the opinion and tolerance patterns, and by reporting time plots and patterns of opinions and tolerance at specific simulation times.

3. Single extremist regimes

3.1. Single extremist with full mixing topology

To easily gain more insight, let us start from a rather extreme case: one single extremist agent chosen with initial opinion 0.99 and 0.001 tolerance. The topology is full mixing. Large simulation times are used (10 000 iterations per agent) to ensure convergence under every simulation condition (Fig. 2).

The regime diagram (Fig. 2) clearly shows that the centrist agents are all attracted by the extremist when their initial tolerance dl is above 0.5: they gather in a cluster of opinion 0.99 and tolerance 0.001. The interpretation is straightforward: for this low value of extremist tolerance, interaction between the extremist and centrists are asymmetric (as we checked during the simulation). The extremist acts as a fixed source of extremism, formally equivalent to a heat source at constant temperature (Eqs. (1) and (2) can be thought of as a randomly discretized version of a Euler relaxation algorithm solving a diffusion equation [15]. Here, opinion is the equivalent of temperature).

Below $dl = 0.5$ the influence of the extremist decreases and agents cluster near the centre opinion keeping roughly their initial tolerance. Asymptotic extremism is sometimes observed around $dl = 0.4$ for instance: this happens when several initial pairing of tolerant agents with extremists give them a role of ‘‘active intermediates’’ in promoting extremism (as we observed by checking the intermediate stages during the simulation).

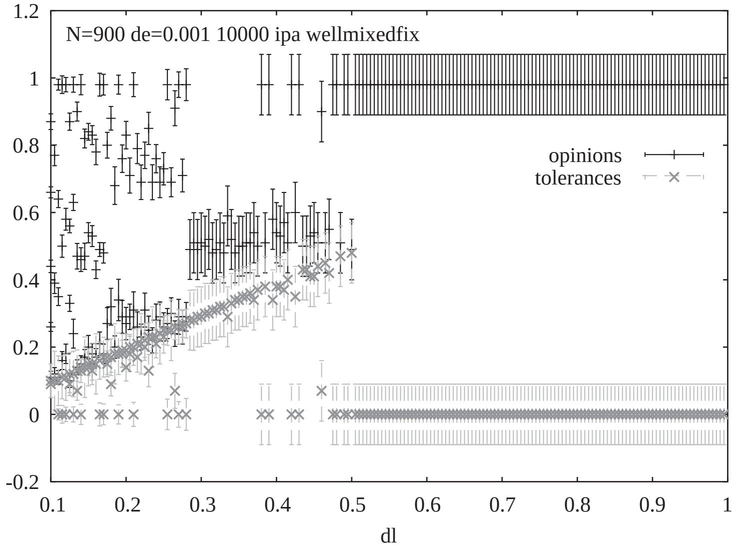


Fig. 2. Histograms of asymptotic clusters (one single extremist, full mixing topology). The y-axis code the histogram of attractor clusters by vertical bars which magnitude represents how many agents are in the asymptotic cluster. The position of the bar represents either the opinion (black '+') or the tolerance (grey '×', dashed line) of the agents in that cluster. The horizontal axis gives the initial tolerance parameter of the "centrist" agents. One single extremist present, $N = 900$, $de = 0.001$, average number of iterations per agent 10 000, any pair of agents can a priori interact.

When $dl < 0.27$, the diagram shows the same increase in cluster number that can be observed in the absence of extremist (the "$1/2d$ rule"), except for a partial extremism clustering below $dl = 0.27$ which is easily understood.

The region $0.37 < dl < 0.52$ is thus a co-existence region where both asymptotic regimes, centrism or extremism can be observed, depending upon initial conditions and pair sampling.

3.2. Single extremist with square lattice topology

In many cases, we expect interactions to occur across some social network. Such would be the case for political discussions, especially in the absence of an open discussion forum. Many model topologies of social networks have recently been proposed. Here, we report simulation results for square lattices, when interactions are only possible among nearest neighbours (each node can only interact with his four neighbours). The boundaries of the lattice are connected to each other; the diagrams represent, in fact, the unfolding of a torus (see Fig. 3).

Although the regularity of connections on a square lattice make it a poor candidate to model a social network, the existence of short interaction loops is shared with many empirical social nets. But again, the purpose of this paper is to

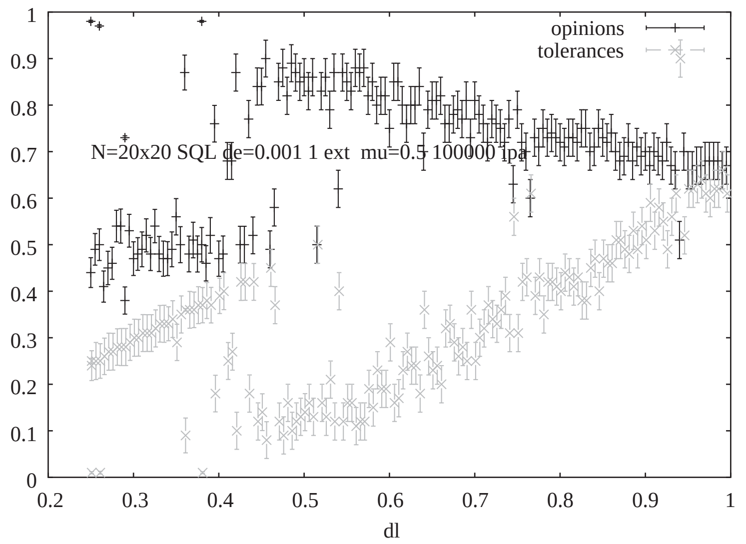


Fig. 3. Histograms of asymptotic clusters for interactions across a square lattice topology. Same coding as the previous figure. One single extremist present, $N = 400$, $de = 0.001$, average number of iterations per agent 100 000.

increase our understanding of the dynamics of more complicated cases and the possibility to observe patterns determined our choice of lattice topology. The relevance of the results to other topologies will be further discussed.

All simulations display the same initial behaviour: the first steps of the dynamics ($\simeq 10$ iterations per agent) result in averaging centrist opinions. Extremists are situated in isolated extremist islands surrounded by a nearly homogeneous sea of centrism. The ultimate fate of these islands depend on their ability to convince their neighbours. We might a priori expect that their influence might be weak for $dl < 0.5$ since their difference in opinion with their neighbours is above the tolerance threshold. On the contrary, when $dl > 0.5$, they are able to influence them. But local fluctuations in the initial opinions of the neighbours might blur this obvious frontier and give rise to some co-existence regions.

Let us first discuss the lattice dynamics when $dl > 0.5$.

Fig. 4 displays opinion and tolerance patterns after 27 average updates per site, i.e., after the initial local averaging transient. $d_l = 0.55$. The initial extremist, at the centre of the lattice, is surrounded by a spot of "new converts"; outside the spot, most sites have reached the centrist cluster configuration unstable at $d_l = 0.55$ with respect to attraction by extremist; in between these two regions, a narrow diffusion front propagates outwards. In the beginning, the interactions are asymmetric across the front as long as the opinion difference across the front is larger than the tolerance of the "new extremists" inside. The extremist spot widens until the time when the front becomes less sharp, the difference in opinions across the front decreases and

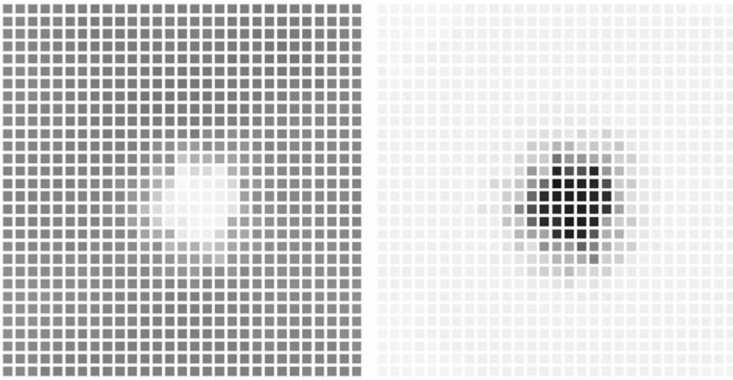


Fig. 4. Propagation of extremism from a single initial extremist at the centre of the lattice. Left frame: Opinions after 27 iterations per site. Colour scale: black 0, grey 0.5, white 1. Right frame: Tolerance relative to d_l after 27 iterations per site. Initial centrist's tolerance $d_l = 0.55$, initial extremist tolerance 0.001.

the tolerance of the "new extremists" inside increases. When they cross, interactions become symmetrical and the centrists themselves start convincing extremists. Eventually, the asymptotic homogeneous cluster position reflects this balance of influence between the two regions at the onset of symmetric interactions.

Direct observation by simulations show that one can approximate the opinions and tolerances inside the extremist spot by the initial extremist opinion and tolerance x_e and d_e, and outside the spot by the average centrists opinion and tolerance 0.5 and d_l. The balance equations (equivalent to physics' calorimetric equations) relate asymptotic opinion x_∞ and tolerance d_∞ to the size s of the extremist spot at the onset of symmetric interactions. They are written as

$$s \times x_e + (S - s) \times 0.5 = S \times x_\infty \,, \tag{7}$$

$$s \times d_e + (S - s) \times d_l = S \times d_\infty \,, \tag{8}$$

where S is the lattice area. Since s only depends upon *dl*, the balance equations explain the variations of x_∞ and d_∞ with S and *dl*:

- Larger *dl* values favour an earlier emergence of symmetric interactions, and a smaller extremist spot at the onset of symmetric interaction. As a consequence, the position of the asymptotic cluster get closer to centrism when *dl* increases as seen in Fig. 3.
- The shift towards centre when $dl > 0.5$ is density dependant: for a smaller lattice, $N = 100$ instead of 400, the deviation of extremism towards the centre is much weaker for the same initial tolerances (e.g. the opinion cluster is at 0.9 rather than 0.7 at $dl = 0.99$).

On the other hand, these equations allow to deduce s from the measurement of x_∞ and d_∞, a method which we will further use in Section 5 to study the dynamics in the presence of several extremists.

Fig. 3 for a square lattice is the equivalent of Fig. 2. It represents the histograms of asymptotic opinions and tolerance and also displays a co-existence region when $0.37 < dl < 0.5$ with the occurrence of either type of attractor. Starting from dl just above 0.5, the extremism regime seems to reappear, but for larger value of dl, the asymptotic extremist cluster continuously moves towards centrist opinions, but with a lower tolerance than dl.

In conclusion, the main difference in dynamics between well-mixed systems and the square lattice structures then occurs in the $dl > 0.5$ region. The relative "return towards centrism" experienced by the transient extremists in the neighbourhood of the initial extremist is the most important network effect that we observed. In contrast, in the case of full mixing, all agents are attracted by the extremist when it acts as a source; the absence of screening shells in the vicinity of the extremist results in full convergence towards extreme opinion and tolerance.

4. Simulation with several extremists

4.1. Several extremists and full mixing

Deffuant et al. [1] report the existence of several dynamical regimes for the full mixing case in the presence of extremists of both kind:

- When $dl < 0.5$, extremists are not important and clustering follows the standard $1/2d$ rule.
- When $dl > 0.5$ they determine the dynamics:
  - at high extremist initial density, clusters of extremists appear at both end of the spectrum; we speak of double extreme attractors;
  - at low extremist density, instabilities of the centrist attractor arise, and the system evolves either in a single asymmetric extreme attractor at one end of the spectrum, or can reach an attractor with centrist opinion but extremist low tolerance.

All time plots and regime diagrams are given in Ref. [1].

4.2. Several extremists on a square lattice topology

On-line observation of simulations show that the outcome of the dynamics depends upon which meso-scale features are able to develop at some intermediate stage of the dynamics, after the initial averaging of centrist opinions far from the extremists (50 < number of iterations per agent < 500) depending on μ and d_l.

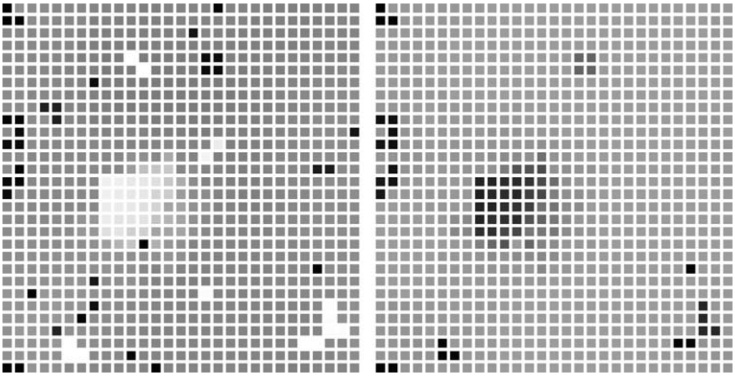


Fig. 5. Upper frames: opinions (left panel) and tolerances (right panel) after 300 iterations per site. Same colour scale as Fig. 4. Initial centrist's tolerance $d_l = 0.38$.

The patterns of Fig. 5 exemplify the different possible influence of initial extremists, according to the first events occurring in their immediate neighbourhood. One can distinguish two different ''geographic'' configurations:

- Mesas: A few extremist islands, e.g. in the upper parts of the patterns, with opinion and tolerance close to the initial extremist values survive, but their influence on their neighbourhood is zero: the difference in opinion at the edge between those agents on the mesa and all their neighbours is larger than $dl = 0.38$. Obviously small values of $dl < 0.5$ favour mesas, which disappear whenever $dl > 0.5$.
- Hills: The success of the extremists in the lower-left corner is due to local fluctuations in opinions: there exists in their neighbourhood some centrists which distance in opinion is less than tolerance. Since the dynamics results in decreasing local opinion gradients, the diffusion process once started carries on across the lattice.

Whether extremist spots are able to develop into hills is a probabilistic process when $dl < 0.5$. But for any value of dl, whether hills will be able to fully develop and the outcome of the dynamics also depend upon how far on the lattice are the initial extremists.

We are now in a better position to interprete Fig. 6, which displays the standard asymptotic histograms as in Figs. 2 and 4.

5. The global picture and the ''hopeful monster hypothesis''

The former mesa/hill analysis leads to some predictions about which global attractor is reached when several extremists of either side are randomly scattered. A basic hypothesis is that when the density of extremists is low, the initial growth of ''extremist spots'' are independent events which occurrence only depends upon a

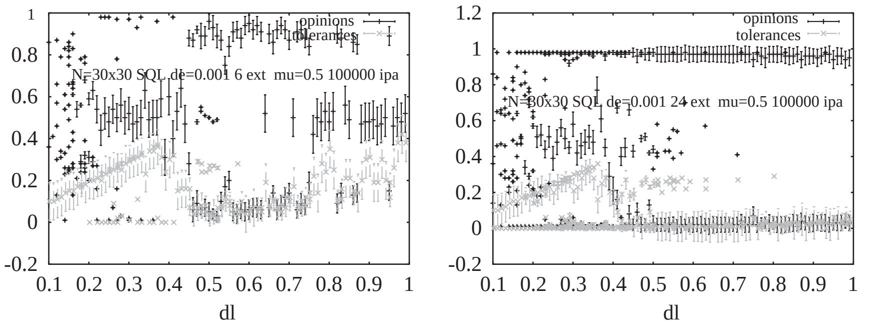


Fig. 6. Clustering as a function of centrists' threshold *dl*. Vertical bars represent clusters: their centre gives cluster's average opinion and their length give the number of agents belonging to the cluster. The left plot corresponds to a small number of initial extremists, 6, while the right plot corresponds to a larger number of initial extremist, 24.

restricted neighbourhood of each initial extremist. We call these "extremist spots" "hopeful monsters" since they are susceptible to grow and invade the lattice.

5.1. Large initial tolerance

Above $d_l = 0.5$, extremists are always able to influence some centrists in their immediate neighbourhood. Spots of extremism start to develop; but their ultimate fate depends on the density of extremists.

- When the spots are far apart, they have enough space to reach the hill status and opinion exchanges with their centrist neighbours become symmetrical. The centrists region acts as "heat/opinion" conductors which "transfer" heat between extremist spots. The final cluster is characterized by opinions averaging around 0.5 (for initial equal numbers of extremists of either kind) and relatively low tolerance with respect to *dl* (Fig. 5, left, at larger values of *dl*).
- For larger extremist densities (Fig. 5, right), the extremist spots collide before reaching a hill profile. The asymptotic opinion pattern displayed in Fig. 7 is composed of extremist mesas of both kinds. Bi-extremism is thus observed with a few isolated spots of centrists. Asymptotic tolerances are low.

The crossover line in the (centrist tolerance)/(extremist density) plan can be guessed by checking the extremist density such that the mesas at their larger extension, i.e., before their evolution into hills, cover all the lattice. The mesas' larger extension s is computed from the balance equation, using the determination of asymptotic opinions x_∞ and tolerance d_∞ from Fig. 3. These two independent determinations of s coincide. For instance, the crossover between bi-extremism and centrism attractor is observed with 6 initial extremists (Fig. 6 on the left) at $d_l =$ 0.75; the balance equations predicts a mesa size of 200 agents and a crossover of 4.5 extremists for a lattice size of 900 (the size used for Fig. 6). We consider this

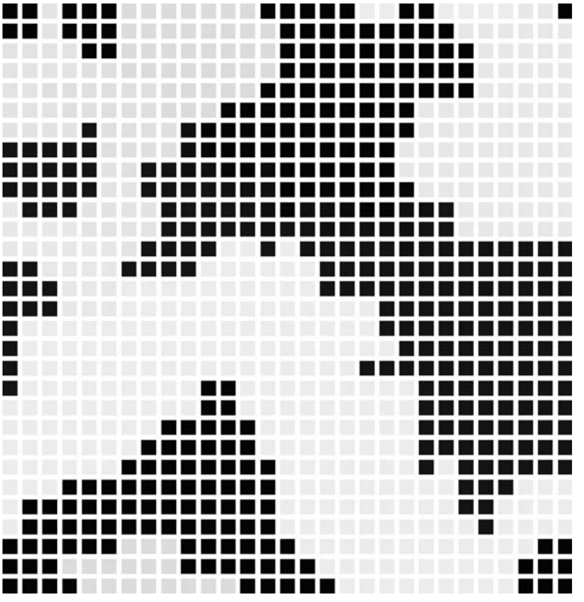


Fig. 7. A bi-extremism opinion pattern observed after 200 iterations per agent for initial tolerances of $dl = 0.6$ and $de = 0.001$; 24 extremists on each side 0 or 1 were initially present. Same colour scale as Fig. 4.

prediction (4.5 instead of 6) as satisfactory in view of the strong simplifications that we made.

5.2. Low initial tolerance

Between $d_l = 0.5$ and $d_l = 0.25$, extremists are not always able to influence centrists in their immediate neighbourhood. Chances of conversion to extremism depend upon the existence of neighbours close enough in opinion (i.e., with difference in opinion smaller than centrist tolerance). Since an initial extremist is only one among four neighbours of its centrist neighbours, the random sampling of pairs might result in initial interactions of these neighbours with centrists (with a 0.75 probability), thus making it harder for the extremist and its neighbour to later interact.

- For a small number of extremists in a large lattice, there is a tolerance region such that the probability of having only one diffusing extremist ‘‘hill’’ is large enough to observe single extremism convergence. This is also true if there are several ‘‘hills’’ on the same extremist side, either close to 0 or close to one.
- At large extremist density, one obtains several hills and bi-extremism is by far the most frequent attractor.

Chances of extremists to influence their neighbours anyway decrease with d_l.

Below $d_l = 0.25$, the lattice is highly divided between many clusters some of which are extremists.

If we call P_0 the probability of occurrence of a hill around one extremist, we can obtain the probabilities of observing any of the three attractors by simple combinatorics. In the presence of $2n_e$ initial extremists (n_e extremists close to 1, n_e close to 0) on a large lattice (to ensure independence), these probabilities are given by

$$P_c = Q_0^{2n_e} , \tag{9}$$

$$P_{bie} = (1 - Q_0^{n_e})^2 , \tag{10}$$

$$P_{moe} = 2(1 - Q_0^{n_e})Q_0^{n_e} , \tag{11}$$

- where $Q_0 = 1 - P_0$ is the probability of any initial extremist to give a "sterile" mesa;
- P_c is the probability of getting a centrist cluster (due to the absence of any extremist hill);
- P_{bie} is the probability of getting clusters of extremists of both kind (two kinds of hills present);
- P_{moe} is the probability of getting a single extremist cluster (only hills of the same extreme grow).

These expressions are immediately generalized to the asymmetric case when the initial numbers of extremists close to 0 and to 1 are different. They imply that the initial number of extremists is important, not their density, at least in the limit of low densities (densities such that intermediate meso-scale features are generated independently of each other).

The exact calculation of P_0 as a function of the threshold d_l involves a rather intricate combinatorics on the possible initial configurations of the extremist's neighbourhood and on the initial sequence of iterations. But P_0 is easily evaluated by simulations. We did it on a 32×32 square lattice with a single extremist in the centre.

Knowing P_0 allows to check the "independent hopeful monster hypothesis" which predicts the occurrence of attractors with probabilities given by Eqs. (9)–(11). Let 's take the case of $n_e = 3$. Eq. (11) predicts a maximum P_{moe} probability of occurrence of a single extremist attractor of 0.5 at $Q_0^3 = 0.5$, which corresponds to $Q_0 = 0.79$ and approximately to $d_l = 0.35$ according to Fig. 8. The statistics plotted in Fig. 9 roughly confirm this prediction: the maximum of P_{moe} is around 0.5 and occurs around $d_l = 0.38$.

Fig. 8 and expressions (9)–(11) then give a clear prediction of the succession of the most frequent attractors when centrist initial tolerance is decreased from 0.5 to 0.25.

- Bi-extremism is predominant until $(1 - P_0)^{n_e}$ reaches 0.5.

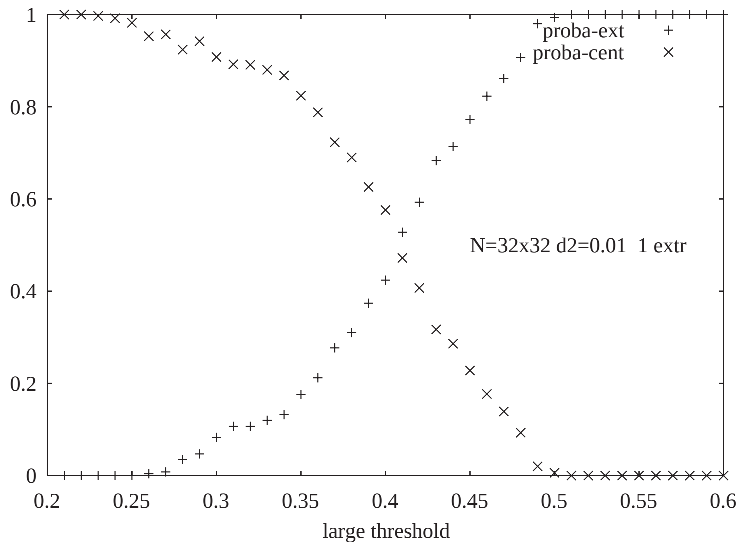


Fig. 8. Statistics of extremist attractors ('+') and centrists attractor ('×') as a function of centrists' threshold for one initial extremist on a square lattice. Each point corresponds to 1000 samples on a 32×32 lattice.

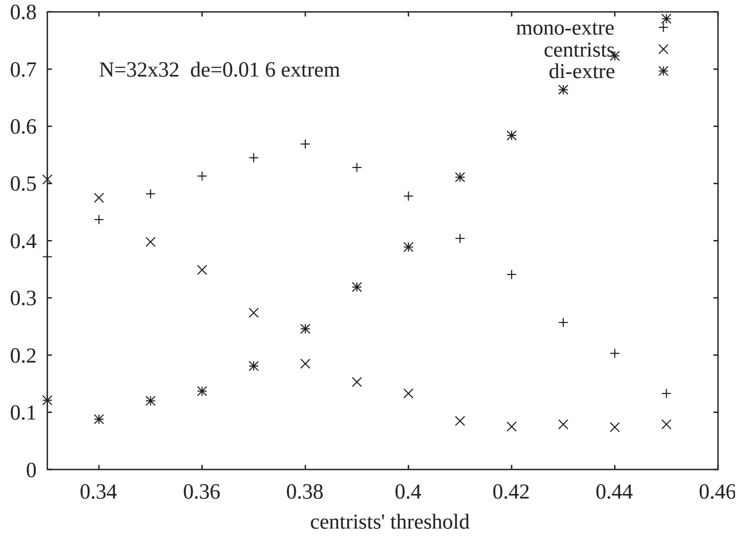


Fig. 9. Statistics of attractors: '+' are single extreme, '*' are double extreme and '×' are centrist attractors; as a function of centrists' threshold. Each point corresponds to 1000 samples on a 32×32 lattice with 6 extremists.

- The probability of observing the single extremist attractor then increases to a maximum of 0.5 according to Eq. (11) which also predicts the width of the single extremist region, W_{moex} (the region where the probability of the single extremist attractor is above one half of the maximum $P_{moe} > 0.25$). According to Eq. (11)

$$W_{moex} \propto (n_e)^{-1} . \tag{12}$$

- Below the single extremist region, centrist attractors are predominant.

So according to the above analysis, single extreme attractors should be observable even in the presence of many initial extremists, but the width of their region of existence decreases in $1/n_e$.

Large fluctuations of the statistics of "hills" and "mesa" are observed in Figs. 8 and 9 due to the vicinity of regime transitions (these two figures represent averages over 1000 samples, and fluctuations are still noticeable). These fluctuations reduce the occurrence of single extreme attractors at larger n_e values. Furthermore increasing n_e decreases the distance between sources of intolerance that cannot be considered as independent anymore: the probability of a "centrist" to be early influenced by an extremist is increased by having more than one extremist neighbour.

A rapid survey of the d_l region most favourable to single extreme attractor, $0.34 < d_l < 0.40$, when the number of extremists is increased from 0.4 to 2 perc. show that the probability of observing single extreme attractors decreases from 50 to 4 perc. This is consistent with Amblard and Deffuant [12], who report the absence of any single extreme attractor for extremist densities higher than 2.5 perc.

5.3. Late appearance of extremism

Up to now, we have supposed that extremists are always present and in particular before the general convergence process. We might also wonder what happens when the extremists appear later, after the centrists agents have clustered. Such a situation could be encountered in a consensual society after some event (war, economic crisis) triggers resentment and extremism. We then ran simulations from initial conditions such that all centrists'opinions are initially clustered at $x = 0.5$. Fig. 10 results,

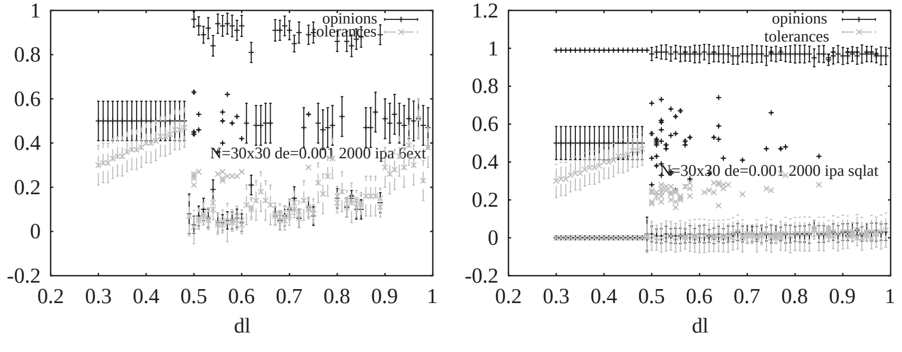


Fig. 10. Clustering as a function of centrists' threshold *dl* starting from clustered centrists' opinions at $x = 0.5$. All other conditions similar to those of Fig. 6.

obtained from initial consensus of centrists, are to be compared with those of Fig. 6, obtained from randomly scattered initial opinions.

- At smaller values of $dl < 0.5$, centrists are outside the influence zone of extremists, no hill develops and centrism always maintains itself.
- At larger values of $dl > 0.5$, the general picture observed for random initial distribution is maintained with some increased probability of observing centrism at low extremist densities.

5.4. Scale-free networks

What about more realistic topologies? Since the successive neighbourhood structure (i.e. there exist a succession of layers of neighbours at distance 1, 2, 3 etc. of each node) is preserved in all networks topologies, except fully connected networks, we expect that the same intermediate scale features which drive the dynamics, such as mesa or hills in opinions or boundaries across domains are present for different topologies. Can we expect for more random networks equivalent phase diagrams, with possibly more irregularities such as outliers and co-existence phases?

We then run the extremist dynamics on scale-free networks [17] to test the above prediction (equivalent phase diagram). (After small worlds networks were introduced by Watts and Strogatz [16], scale-free networks became recently the strongest contenders as models of social networks.) Scale-free networks differ from lattices by the inhomogeneity of connectivity and by their smaller diameter.

We used a standard construction method to generate scale-free networks, see e.g., Ref. [18].

Starting from a fully connected network of 3 nodes, we add iteratively nodes (in general up to 900 nodes) and connect them to previously created nodes in proportion to their degree. We have chosen (among many possibilities) to draw two symmetrical connections per new added node in order to achieve the same average connection degree (4) as in the 30×30 square lattice taken as reference. But obviously the obtained networks are scale-free as shown by Barabasi and Albert [17].

In fact, scale-free networks [17] display a lot of heterogeneity in nodes connectivity. In the context of opinion dynamics, well connected nodes might be supposed more influential, but not necessarily more easily influenced. At least this is the hypothesis that we choose here. We have then assumed asymmetric updating: a random node is first chosen, and then one of its neighbours. But only the first node in the pair might update his position according to Eq. (1), not both. As a result, well connected nodes are influenced as often as others, but they influence others in proportion to their connectivity. This particular choice of updating is intermediate between what Stauffer and Meyer-Ortmanns [18] call directed and undirected versions.

The cluster diagram obtained with 24 initial extremists out of 900 agents (with the same parameters as for Fig. 6, right frame) is represented in Fig. 11.

- Similarity with lattice dynamics. Below $dl = 0.45$ this cluster diagram closely resembles those we obtained for square lattice, with predominance of centrism.

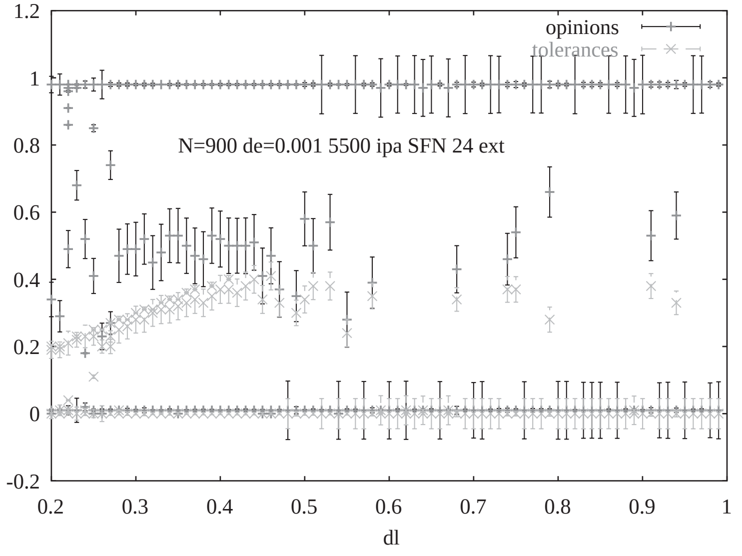


Fig. 11. Clustering as a function of centrists' threshold *dl* for a scale-free networks with the same parameters as for the diagram displayed for the square lattice in Fig. 5. The size of the bars representing tolerance clusters is reduced by a factor two for clarity reasons.

Above $dl = 0.45$ one also observes the predominance of some kind of extremism, with a limited final tolerance.

- Differences. But the only two kinds of observed attractors are low tolerance centrism and single sided extremism. We don't observe two sided extremism attractors as with lattices. The large inhomogeneity in nodes connectivity favours well connected nodes: most often, the best connected extremist impose his view nearly everywhere. And there are still minority clusters around the other extremists. The asymptotic clusters with central opinion are probably obtained when the initial sampling of extremists does not contain highly connected nodes.

The present result is still preliminary: the distribution of the connectivity of initial extremists is only a rough predictor of the outcome of the dynamics. More complete studies, outside the scope of the present paper are still needed.

6. Conclusions

The above series of simulations give a clearer picture of the phenomena occurring in this strongly simplified model of opinion dynamics.

The most important result, already established in Ref. [1], is also true for lattice and scale-free networks topologies: the existence of extremist regimes is largely due to the large tolerance of agents which were initially centrists.

Restricting possible interactions to a network structure noticeably changes the dynamics:

- The eventual influence of extremists is screened due to spatial effects.
- Hill configurations allow the spreading of this influence across the whole network.
- Larger extremist densities result in collisions between meso-scale structures.

A general conclusion is that social networks can limit the propagation of extremism. Major differences with the well-mixed topology are:

- Single extremism is only observed when extremists are few, and when initial centrist tolerance is below 0.5; as opposed to the well mixed case where it occurs even at moderate densities above 0.5.
- Attractors with centrist opinion and lower tolerance are observed with probability close to one in a well defined low density high tolerance region, as opposed to the well mixed case where they occur with a limited probability as the negative outcome of some instability.
- Single sided extremism is favoured by scale-free networks. Even our preliminary results allow to understand why extremists (or market strategists) should first convince leader figures to establish their influence on a social network.

At this stage, we might try to generalize some of the predictions of the model. As we said in our introduction, one can model the influence of a minority of agents with strong views based either on knowledge or vested interests. We thus start from initial random distribution of agents' opinion plus a few agents concentrated in initial opinion with much lower tolerance. These low tolerance agents do not necessarily have extreme opinions. But most of our qualitative conclusions apply:

- The spread of the minority agents opinion and tolerance is favoured by the large tolerance of agents which were initially centrists. The most important predictor of the outcome of the dynamics is the comparison between the tolerance of the majority with the difference between the minority opinion and the average majority opinion. If the tolerance of the majority is larger, minority opinion can spread, otherwise it will most probably not spread.
- The eventual influence of the minority agents is screened due to spatial effects.
- Minority chances to spread depend upon the development of meso scale features such as hills.

Acknowledgements

We acknowledge partial support from the FET-IST grant of the EC IST 2001-33555 COSIN. The present work was completed at the Institute for Scientific Interchange, Torino, which we thank for its hospitality.

References

[1] G. Deffuant, F. Amblard, G. Weisbuch, T. Faure, J. Artif. Soc. Soc. Simulat. 5 (4) (2002) 1 http://jasss.soc.surrey.ac.uk/5/4/1.html.

[2] S. Moscovici, W. Doise, Dissension et consensus, PUF, Paris, 1992.

[3] S. Galam, S. Moscovici, Towards a theory of collective phenomena: consensus and attitude changes in groups, Eur. J. Soc. Psychol. 21 (1991) 49–74;
S. Galam, S. Moscovici, Compromise versus polarization in group decision making, Defense decision making, Russ. Psychol. J. 13 (1991) 93–103.

[4] H. Föllmer, J. Math. Economics 1/1 (1974) 51–62;
A. Orléan, J. Econ. Behav. Organ. 28 (1995) 257–274.

[5] G. Deffuant, Final Report of Project FAIR 3 CT 2092, 2001, pp. 56–69, http://www.lisc.clermont.-cemagref.fr/ImagesProject/freport.pdf.

[6] H.P. Young, M.A. Burke, Am. Econ. Rev. 91 (2001) 559–573.

[7] E. Chattoe, N. Gilbert, Private communications, 1998.

[8] G. Deffuant, D. Neau, F. Amblard, G. Weisbuch, Adv. Complex Syst. 3 (2000) 87;
G. Weisbuch, G. Deffuant, F. Amblard, J.-P. Nadal, Complexity 7 (3) (2002) 55
D. Neau, Révisions des croyances dans un système d'agents en interaction, Rapport d'option de l'école polytechnique, 2000, available at http://www.lps.ens.fr/~weisbuch/rapneau.ps.

[9] R. Hegselmann, U. Krause, J. Artif. Soc. Soc. Simulat. 5 (3) (2002) paper 2 (jasss.soc.surrey.ac.uk);
U. Krause, in: U. Krause, M. Stöckler (Eds.), Modellierung und Simulation von Dynamiken mit vielen interagierenden Akteuren, Bremen University, 1997, p. 37.

[10] R. Axelrod, J. Conflict Resolut. 41 (1997) 203;
C. Castellano, M. Marsili, A. Vespignani, Phys. Rev. Lett. 85 (2000) 3536–3539.

[11] E. Ben-Naim, P.L. Krapivsky, S. Redner, Physica D 183 (2003) 190 (cond-mat/0212313).

[12] F. Amblard, G. Deffuant, The role of network topology on extremism propagation with relative agreement opinion dynamics, Physica A 343 (2004) 725–738.

[13] G. Deffuant, F. Amblard, G. Weisbuch, Modelling group opinion shift to extremism: the smooth bounded confidence model, Proceedings of Vallalodid ESSA 2004 Conference, 2004, to appear.

[14] B. Derrida, H. Flyvbjerg, J. Phys. A 19 (1986) L1003.

[15] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, Numerical Recipies: The Art of Scientific Computing, Cambridge University Press, Cambridge, 1994.

[16] D.J. Watts, S.H. Strogatz, Collective dynamics of small-world networks, Nature 393 (1998) 440–442.

[17] R. Albert, A.L. Barabási, Rev. Mod. Phys. 74 (2002) 47.

[18] D. Stauffer, H. Meyer-Ortmanns, Int. J. Mod. Phys. C 15 (2) (2004);
D. Stauffer, A. Sousa, C. Schulze, Discretized opinion dynamics of the Deffuant model on scale-free networks, J. Artif. Soc. Soc. Simulat. 7 (3) (2004) http://jasss.soc.surrey.ac.uk/7/3/7.html.