Heterogeneous beliefs, segregation, and extremism in the making of public opinions

Serge Galam*
Centre de Recherche en Épistémologie Appliquée (CREA), École Polytechnique et CNRS (UMR 7656) 1 rue Descartes, 75005 Paris, France

(Received 18 September 2004; revised manuscript received 23 November 2004; published 18 April 2005)

The connection between contradictory public opinions, heterogeneous beliefs, and the emergence of majority- or minority-induced extremism is studied, extending our former two-state dynamic opinion model. Agents are attached to a social-cultural class. At each step they are distributed randomly in different groups within their respective classes to evolve locally by majority rule. In case of a tie the group adopts one or another opinion with respective probabilities k and $(1-k)$. The value of k accounts for the average of individual biases driven by the existence of heterogeneous beliefs within the corresponding class. It may vary from class to class. The process leads to extremism with a full polarization of each class along one opinion. For homogeneous classes the extremism can be along the initial minority making it minority induced. In contrast, heterogeneous classes exhibit more balanced dynamics, which results in a majority-induced extremism. Segregation among subclasses may produce a coexistence of opinions at the class level, thus averting global extremism. Insight into the existence of contradictory public opinions in similar social-cultural neighborhoods is given.

DOI: 10.1103/PhysRevE.71.046123 PACS number(s): 02.50.Ey, 05.40.−a, 89.65.−s, 89.75.−k

I. INTRODUCTION

In recent years the study of opinion dynamics has become a main stream of research in physics [1–18]. Initiated long ago [19–21] this subject is part of sociophysics [22].

Outside physics, research has concentrated on analyzing the complicated psychosociological mechanisms involved in the process of opinion forming, in particular focusing on those by which a huge majority of people gives up to an initial minority view [23,24]. The main ingredient is, for instance in the case of a reform proposal, that the prospect of losing definite advantages is much more energizing than the corresponding gains, which by nature are hypothetical.

This feature of an initial minority winning through a democratic process, voting or opinion forming, was obtained by introducing the possibility of local ties, which in turn produce a local bias either in favor of some status quo or in agreement with some cultural bias [8,12,20].

In the present work we generalize the concept of local bias driven by a tie in competing opinions. We introduce social-cultural classes which are each characterized by some common bias k that results from the class average of all heterogeneous individual beliefs. At a tie in a group, all members adopt one or other opinion with respective probabilities k and $(1-k)$. The value of k is constant within each class and may vary from class to class with $0 \leqslant k \leqslant 1$.

In the case of groups of size 4, denoting by O an opponent to the issue at stake and by S a supporter, our update rules are written

$$\text{(a) } SSSS, OSSS \rightarrow SSSS, \quad OOOO, OOOS \rightarrow OOOO,$$

$$\text{(b) } OOSS \rightarrow \begin{cases} OOOO & \text{with probability } k, \\ SSSS & \text{with probability } (1-k), \end{cases}$$

where all permutations are allowed. Rules (a) correspond to majority rule while rules (b) account for the local bias.

In our earlier works we considered homogeneous extreme cases. For high risk aversion sharing $(k=0)$ a reform proposal was found to need initial support of more than 90% of the population to survive a public debate [12]. On the contrary, a shared high prejudice against someethnic or religious group $(k=1)$ can make an initial false rumor shared by only a few percent of the population spread over the whole population [8]. A varying bias was considered in [18] and in an application to cancerous tumor growth in [25]. The existence of threshold dynamics in social phenomena was advocated long ago in social qualitative studies [26,27].

The rest of the paper is organized as follows. The model is defined in the next section. The associated dynamics is outlined. A separator determines the flow direction of the collective opinion forming. Its variation as a function of the class common bias k is calculated. In Sec. III it is shown how very small fluctuations in the initial conditions may lead to opposite extremism with contradictory public opinions in very similar areas. Segregation and mixing effects are studied in Sec. IV. Under some conditions, they are found to drive either majority-induced extremism or coexistence of opinions, thus avoiding global extremism. The last section contains some discussion.

II. SETTING THE PROBLEM

We consider one social-cultural class with a social bias k and N members facing some issue like a reform proposal, a behavior change (stopping smoking), a foreign policy decision, or the belief in some rumor. We discriminate between two levels in the process of formation of the global opinion.

*Email address: galam@shs.polytechnique.fr

First there is an external level, which accounts for the net result from the global information available to everyone, the various private information, and the influence of mass media. The second level is internal and concerns the dynamics driven by people freely discussing the issue among themselves. Both levels are interpenetrated but here we decoupled them to study specifically the latter, focusing on the laws governing the internal dynamics.

Accordingly, at a time t prior to the public debate the issue at stake is given support by $N_S(t)$ individuals (denoted S) and opposition from $N_O(t)$ agents (denoted O). Each person is supposed to have an opinion with $N_S(t)+N_O(t)=N$. Associated individual probabilities to be in favor of or against the proposal at time t are

$$p_{S,O}(t) \equiv \frac{N_{S,O}(t)}{N}, \tag{1}$$

with

$$p_S(t) + p_O(t) = 1. \tag{2}$$

From this initial configuration, people start discussing the project. However, they do not hold ongoing discussions all the time. We assume people are diffusing randomly without interaction. But at a series of discrete times $\{t, t+1, \ldots, t+n\}$, local interactions are activated randomly within small groups of four agents following the above rules (a) and (b). The probability to find one supporter S after n successive updates becomes

$$p_S(t+n) = p_S(t+n-1)^4 + 4p_S(t+n-1)^3\{1-p_S(t+n-1)\} + 6(1-k)p_S(t+n-1)^2\{1-p_S(t+n-1)^2\}, \tag{3}$$

where $p_S(t+n-1)$ is the proportion of supporters at a distance of $(n-1)$ updates from the initial time t. The last term includes the tie case contribution $(2S-2O)$ weighted with the probability k.

Equation (4) exhibits two attractors at $p_{S,0}=0$ and $p_{S,1}=1$ and an unstable fixed point $p_{c,4}$, the separator, at the nonsymmetric value,

$$p_{c,4} = \frac{(6k-5)+\sqrt{13-36k+36k^2}}{6(2k-1)}, \tag{4}$$

which equals $\frac{1}{2}$ at $k=\frac{1}{2}$. Figure 1 shows the variation of $p_{c,4}$ as a function of k. Depending on the initial values of opinions, the effect of the value of k on the direction of the final polarization of public opinion can be seen explicitly. The separator is located between $p_{c,4}\approx 0.23$ for $k=0$ and $p_{c,4}\approx 0.77$ at $k=1$. Dealing with probabilities, the results are independent of N. For very small systems, fluctuations are expected according to $\pm 1/\sqrt{N}$.

Inverting Eq. (4) allows us to define a critical value k_c in the common bias as a function of the initial support p_S for a given issue like a rumor with

$$k_c = \frac{-1+5p_S-3p_S^2}{6p_S(1-p_S)}, \tag{5}$$

where k_c determines two phases for any initial support p_S. For all classes that have a common belief $k<k_c$, a rumor

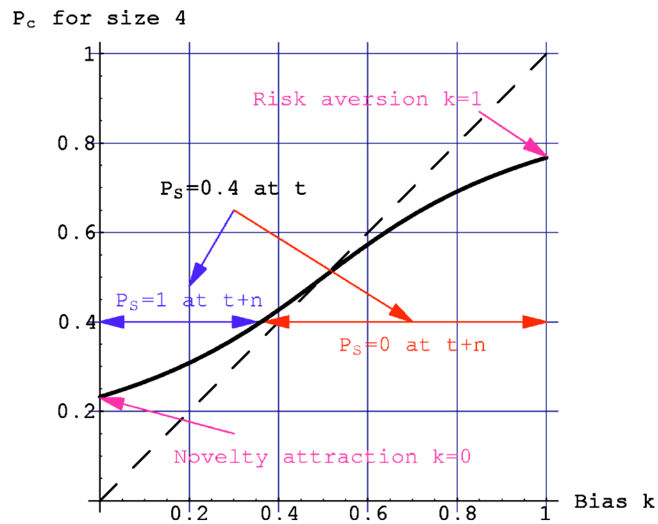

FIG. 1. Variation of $p_{c,4}$ as function of k. For collective risk aversion $(k=1)$, $p_{c,4}\approx 0.77$ while for collective novelty attraction $(k=0)$, $p_{c,4}\approx 0.23$. In the case of no collective bias $(k=1/2)$, $p_{c,4}=1/2$. An initial support of $p_s=0.40$ is shown to lead to an extremism in favor $(p_S=1)$ for the range of bias $0\leqslant k\leqslant 0.36$. In contrast the extremism is against $(p_S=0)$ for the whole range $0.36\leqslant k\leqslant 1$.

supported by a fraction p_S of the agents will just fade away by discussion among the people alone; there is no need for an external intervention. However, when $k>k_c$ the rumor will invade the whole population unless some external action is taken.

For instance, an initial support of $p_S=0.40$ leads to an extremism in favor of the issue at stake for the whole range of bias $0\leqslant k\leqslant 0.36$. In contrast when $0.36\leqslant k\leqslant 1$ the final extremism is against the issue.

III. SENSITIVITY TO INITIAL CONDITIONS AND CONTRADICTORY PUBLIC OPINIONS IN SIMILAR AREAS

We now discuss the situation of two different neighboring areas characterized by very similar common beliefs, for instance, a city area and its suburb as shown in Fig. 2. We are in the presence of two sociocultural classes; one covers the

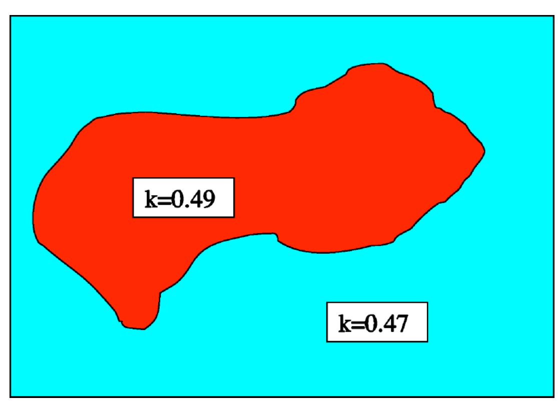

FIG. 2. A city with $k=0.49$ and its surrounding suburb with $k=0.47$.

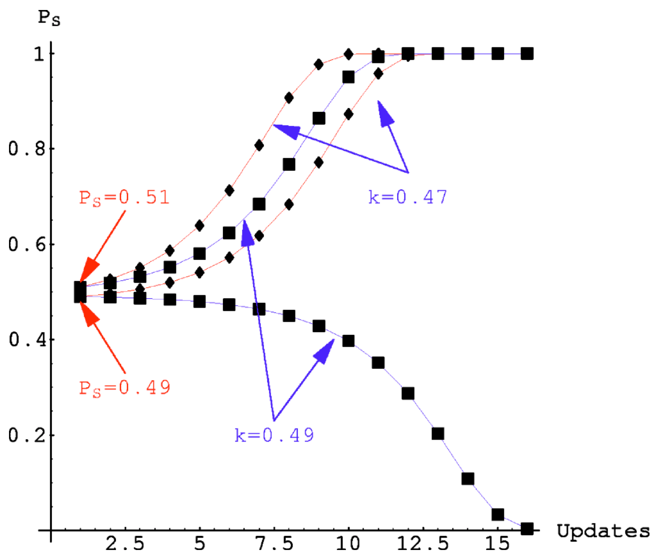

FIG. 3. Evolution of $p_S(t)$ as a function of updates for groups of size 4 with $k=0.49$ and 0.47. Two initial supports are considered. For $p_S=0.51$ both $k=0.49$ and 0.47 lead to $p_S=1$. But for $p_S=0.49$ only $k=0.47$ leads to $p_S=1$ while $k=0.49$ leads to $p_S=0$.

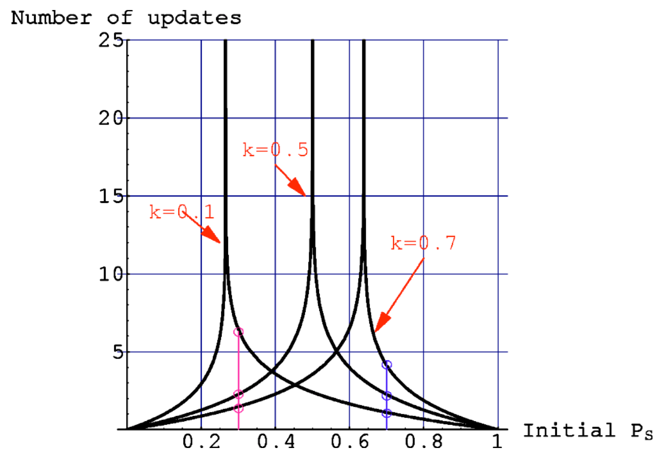

FIG. 4. Number of required updates to reach extremism for several values of the local bias with $k=0.10$, 0.50, and 0.70. Associated values are shown for an initial support of, respectively, 30% and 70%.

city and the other the suburb. We assume two similar biases with, respectively, $k=0.49$ and 0.47. Such a minor difference is not expected to be explicitly felt, in particular when crossing from one area to another. They are perceived as identical. However, a study of the dynamics of opinion starting from the same initial conditions may lead to huge differences in the two areas.

We consider the possible outcomes of two very similar initial conditions where the issue at stake would have either 51% or 49% support within both the city and suburb populations. For 51% support of the issue at stake both populations end with full support of it, making both geographical areas identical as seen in Fig. 3. The process is completed within an estimate of ten updates. However, a tiny decrease of 2% in the initial support down to 49% would split the two neighboring similar areas into opposed extremisms. The city is totally opposed to the issue while the suburb is in full support of it (see Fig. 3).

The above case may shed light on situations in which contradictory feelings or opinions are sustained in areas which are nevertheless very similar, like for instance the feeling of safety. It shows how an insignificant change in either the initial support or the bias driven by the common beliefs may yield drastic differences in the outcome of public opinion. The same initial support is shown to lead to totally different outcomes as a function of k. A value $p_S=0.30$ leads to $p_S=0$ for both $k=0.50$ and 0.70, while it yields $p_S=1$ for $k=0.10$. For $p_S=0.70$ all three cases lead to $p_S=1$.

Performing a Taylor expansion from Eq. (4), the number of updates to reach full polarization can be calculated as

$$n \simeq \frac{1}{\ln[\lambda]} \ln\left(\frac{p_c - p_S}{p_c - p_+(t)}\right), \tag{6}$$

where λ is the first derivative of $p_S(t+1)$ with respect to $p_S(t)$ taken at $p_S(t)=p_c$ and $p_S=0$ if $p_+(t)<p_c$ while $p_S=1$ when $p_+(t)>p_c$. The number of updates being an integer, its value is obtained from Eq. (6) by rounding to an integer. The number of updates diverges at p_c. The situation is symmetric with respect to $k=0$ and 1 with the divergence at, respectively, $p_c=0.23$ and 0.77. It is smaller and occurs at $p_c=0.50$ for $k=3$ as shown in Fig. 4.

IV. SEGREGATION, MAJORITY-INDUCED EXTREMISM, AND COEXISTENCE

Up to now we have considered at a tie an average local bias k, which results from a distribution of heterogeneous beliefs within a population. It means that all members of that population do mix together during the local group updates, whatever is the individual belief. At this stage it is worth noting that different situations may arise in the distribution of the individual k_i.

We discuss two cases for which either all k_i are equal, i.e., a homogeneous population, or they are all distributed among two extreme values, for instance, 0 and 1. There the existence of subclasses as a result of individual segregation may become instrumental in producing drastic changes in the final global public opinion of the corresponding class.

Consider first two different homogeneous classes A and B in two different areas with $k_i=0$ for all i in A and $k_j=1$ for all j in B. From Eq. (5) an initial support $p_S=0.25$ yields $p_S=1$ for A and $p_S=0$ for B as seen in Fig. 5. For A the extremism in support of the issue is minority induced since initially a minority $p_S=0.25$ is in favor. Contrarily, in B the extremism is against the issue but majority induced since the initial majority of $1-p_S=0.75$ is against it.

Now consider the above classes A and B, but as subclasses of the same class within one unique area. Two situations may occur as illustrated in Fig. 6 where A individuals are represented by circles and B ones by squares. They are white when in favor and black if against. In the first situation (upper part of the figure) people from each subclass do not mix together while updating their individual opinions. They are segregated from each other yet sharing the same class

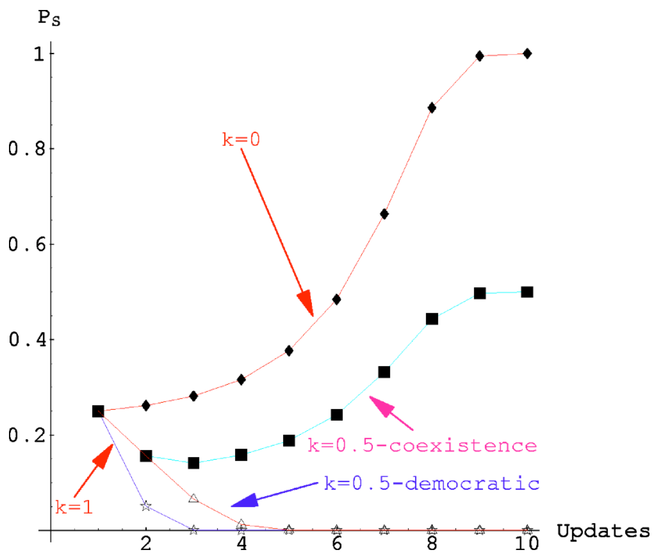

FIG. 5. The dynamics of public opinions for two populations on different areas with respectively $k=0$ and 1 at an initial support of $p_S=0.25$ is shown on the upper and lower curves. In between the two populations are in the same area as subclasses of one unique class. The upper one has segregation and yields a coexistence of opinions. The lower one has mixing and reveals a majority-induced extremism.

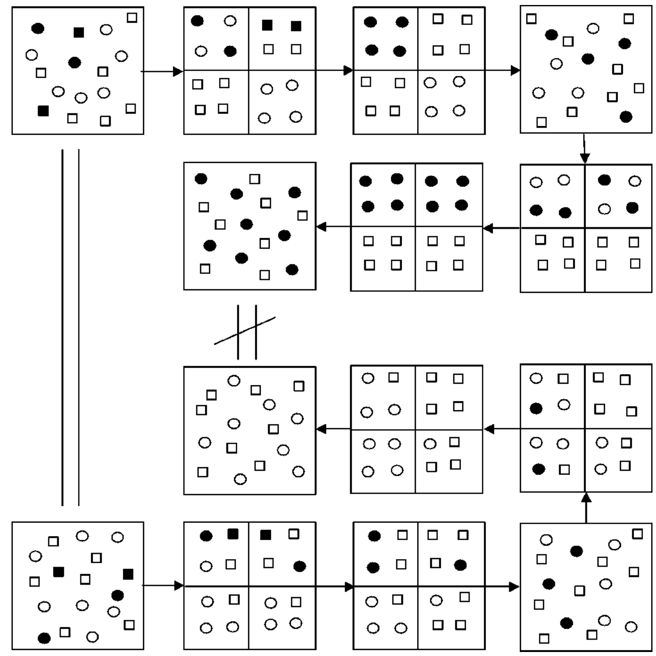

FIG. 6. A population composed from two subclasses sharing opposite beliefs. The circles have $k=0$ while the squares have $k=1$. The initial proportion in favor (white) is identical for both of them at $p_S=0.25$. Black color expresses an opinion against. In the upper series, people segregate while updating their opinions. The result is a perfect balance of the global public opinion with all circles against (blacks) and all squares in favor (whites). In contrast, in the lower series, circles and squares do mix together while updating. The result is a majority-induced extremism with all circles and squares in favor (white).

within the same geographical area. As a result, two opposite extremisms for each subclass are obtained as above. However, unlike the case of distinct geographical areas, here the resulting public opinion of the global class, which does include the whole population, is no longer exhibiting any extremism. The dynamics of segregated updates of opinions has produced a stable coexistence of both initial opinions with thus a collective balance remaining. A poll about the bias would reveal the average value $k=1/2$.

In the second situation (lower part of the figure) people from each subclass do mix together while updating their individual opinions. As a result, at a tie with mixed individuals, A people adopt the opinion in favor while B people go against. With two subpopulations with more or less the same large enough size this process is equivalent on average to having a bias $k=1/2$. The resulting extremism is majority induced since it is along the initial global majority among the whole population. Mixing or segregation within the same situation may thus lead to drastically different public opinions as illustrated in Fig. 6.

We now extend the above cases to the general case of two biases k_1 and k_2. In case of segregation we have two different separators at $p_{c1,4}$ and $p_{c2,4}$ which are obtained from Eq. (4). Three different situations may occur depending on the respective ordering between $p_{c1,4}$, $p_{c2,4}$, and the initial $p_S(t)$. Assuming $p_{c1,4}<p_{c2,4}$ we have

(1) $p_S(t)<p_{c1,4}<p_{c2,4}\rightarrow p_S(t+n)=0$, i.e., global extremism against;

(2) $p_{c1,4}<p_S(t)<p_{c2,4}\rightarrow p_S(t+n)=\frac{1}{2}$, i.e., global coexistence;

(3) $p_{c1,4}<p_{c2,4}<p_S(t)\rightarrow p_S(t+n)=1$, i.e., global extremism in favor.

Going to the case of mixing is more complex since we need to define the way the various tie situations are resolved. At this stage it is of importance to stress that defining the rule (b) implies the existence of an interaction among the four agents since it excluded the possibility of conserving the tie. They are forced to agree on a common opinion. Accordingly while mixing agents from the two classes we have four different types of configuration with, respectively, four, three, two, one, and zero members of class A.

Each time there exists a majority, i.e., four or three members of one subclass, they adopt a common opinion according to their common beliefs. When one member of the other class is present it will adopt the opinion of the local majority of the three other subclass members. It gives for the various tie configurations the probabilities

$$O_AO_AS_AS_A, O_AO_AS_AS_B \rightarrow OOOO \quad \text{with } k_1,$$

$$SSSS \quad \text{with } (1-k_1),$$

$$O_AO_BS_BS_B, O_BO_BS_BS_B \rightarrow OOOO \quad \text{with } k_2,$$

$$SSSS \quad \text{with } (1-k_2).$$

The case of two members of each class requires an averaging of the respective common beliefs with

$$O_AO_AS_BS_B \rightarrow OOOO \quad \text{with } \frac{k_1+k_2}{2},$$

$$SSSS \quad \text{with} \left(1 - \frac{k_1 + k_2}{2}\right).$$

For two subclasses with proportions α and $(1-\alpha)$, adding all the above configurations results on average in the rules

$$OOSS \rightarrow OOOO \quad \text{with probability } K,$$

$$SSSS \quad \text{with } (1-K),$$

where

$$K = [\alpha^4 + 4\alpha^3(1-\alpha)]k_1 + 6\alpha^2(1-\alpha)^2\frac{k_1+k_2}{2} + [4\alpha(1-\alpha)^3 + (1-\alpha)^4]k_2. \quad (7)$$

Inserting $\alpha=\frac{1}{2}$, $k_1=0$, and $k_2=1$ yields $K=\frac{1}{2}$, i.e., a perfect coexistence of both opinions as argued above.

V. CONCLUSION

To conclude, we have presented a simple model of opinion dynamics which is able to reproduce some complexities of the social reality. It suggests that the direction of the inherent polarization effect in the formation of public opinion driven by a democratic debate is biased by the existence of common beliefs within a population. Since the resulting collective bias may vary from one population to another, different areas were found to hold identical extremism while very similar ones may exhibit opposite extremisms. The effect of mixing or segregating within different populations living in the same area was also studied. Homogeneous versus heterogeneous situations were shown to result in different qualitative outcomes.

At this stage we did not address the difficult question of how to remedy this phenomenon of reversal of opinion with the natural establishment of minority-induced extremism. The first hint could be in avoiding the activation of a common general background in the social representation of reality. However, direct and immediate votes could also be rather misleading. Holding an immediate vote without a debate as soon as a new issue arises has other drawbacks. At this stage, collaboration with psychosociologists as well as political scientists would be welcome.

In addition, in real life not every person is open minded and changes opinion. Therefore it would be interesting to introduce stubborn agents in the model. The model may generalize to a large spectrum of social, economical, and political phenomena that involve propagation effects. In particular it could shed light on both the processes of fear propagation and rumor spreading.

[1] E. Ben-Naim, e-print cond-mat/0411427.

[2] F. Wu and B. A. Huberman, e-print cond-mat/0407252.

[3] G. Weisbuch, G. Deffuant, and F. Amblard, e-print cond-mat/0410200; G. Weisbuch, G. Deffuant, F. Amblard, and J. P. Nadal, Complexity **7**, 55 (2002).

[4] C. J. Tessone, R. Toral, P. Amengual, H. S. Wio, and M. San Miguel, Eur. Phys. J. B **39**, 535 (2004).

[5] M. C. Gonzalez, A. O. Sousa, and H. J. Herrmann, Int. J. Mod. Phys. C **15**, 45 (2004).

[6] D. Stauffer and H. Meyer-Ortmanns, Int. J. Mod. Phys. C **15**, 241 (2004).

[7] F. Slanina and H. Lavicka, Eur. Phys. J. B **35**, 279 (2003).

[8] S. Galam, Physica A **320**, 571 (2003).

[9] D. Stauffer and K. Kulakowski, TASK Q. **2**, 257 (2003).

[10] P. L. Krapivsky and S. Redner, Phys. Rev. Lett. **90**, 238701 (2003).

[11] M. Mobilia and S. Redner, Phys. Rev. E **68**, 046106 (2003).

[12] S. Galam, Eur. Phys. J. B **25**, 403 (2002).

[13] F. Schweitzer, J. Zimmermann, and H. Mühlenbein, Physica A **303**, 189 (2002).

[14] K. Sznajd-Weron and J. Sznajd-Weron, Int. J. Mod. Phys. C **11**,1157 (2002).

[15] D. Helbing, I. Farkas, and T. Vicsek, Nature (London) **407**, 487 (2000).

[16] S. Solomon, G. Weisbuch, L. de Arcangelis, N. Jan, and D. Stauffer, Physica A **277**, 239 (2000).

[17] S. Galam, *Le Monde*, 28 March, 2000, pp. 18–19.

[18] S. Galam, B. Chopard, A. Masselot, and M. Droz, Eur. Phys. J. B **4**, 529 (1998).

[19] S. Galam, Y. Gefen, and Y. Shapir, J. Math. Sociol. **9** 1 (1982).

[20] S. Galam, J. Math. Psychol. **30**, 426 (1986); J. Stat. Phys. **61**, 943 (1990); Physica A **285**, 66 (2000).

[21] S. Galam and S. Moscovici, Eur. J. Social Psychol. **21**, 49 (1991).

[22] S. Galam, Physica A **336**, 49 (2004).

[23] R. D. Friedman and M. Friedman, *The Tyranny of the Status Quo* (Harcourt Brace Jovanovitch, San Deigo, 1984).

[24] S. Moscovici, in *Communication Yearbook*, edited by J. A. Anderson (Sage Publications, Newbury Park, 1990), Vol. 14.

[25] S. Galam and J. P. Radomski, Phys. Rev. E **63**, 051907 (2001).

[26] T. C. Schelling, *Micromotives and Macrobehavior* (Norton and Co., New York, 1978).

[27] M. Granovetter, Am. J. Sociol. **83**, 1420 (1978).