OPINION DYNAMICS AND BOUNDED CONFIDENCE MODELS, ANALYSIS, AND SIMULATION[*]

Rainer Hegselmann
Department of Philosophy, University Bayreuth

Ulrich Krause
Department of Mathematics, University Bremen

Abstract When does opinion formation within an interacting group lead to consensus, polarization or fragmentation? The article investigates various models for the dynamics of continuous opinions by analytical methods as well as by computer simulations. Section 2 develops within a unified framework the classical model of consensus formation, the variant of this model due to Friedkin and Johnsen, a time-dependent version and a nonlinear version with bounded confidence of the agents. Section 3 presents for all these models major analytical results. Section 4 gives an extensive exploration of the nonlinear model with bounded confidence by a series of computer simulations. An appendix supplies needed mathematical definitions, tools, and theorems.

Keywords: opinion dynamics, consensus/dissent, bounded confidence, nonlinear dynamical systems.

1. INTRODUCTION

Consider a group of interacting agents among whom some process of opinion formation takes place. The group may be a small one, e.g., a group of experts asked by the UN to merge their different assessments on, say, the magnitude of the world population in the year 2020, into one single judgement. On the other extreme, the group may be an entire society in which the individuals ruled by various networks of social influences develop a wide spectrum of opinions. In each case there is a process of opinion dynamics that may lead to a consensus among the agents or to a polarization between the agents or, more general, that results in a certain fragmentation of the patterns of opinions. An understanding if not an analysis of the involved process of opinion formation is hardly possible without an explicit formulation of a mathematical model.

An early formulation of such a model was given by J.R.P. French in 1956 in order to understand complex phenomena found empirically about groups (French 1956). See also the related mathematical analysis by F. Harary in 1959 (Harary 1959). Another source of opinion dynamics is the work by M.H. De Groot in 1974 (De Groot 1974) and by K. Lehrer in 1975 (Lehrer 1975) and subsequent work by C. G. Wagner, S. Chatterjee, E. Seneta, J. Cohen among others. (See (Wagner 1978), (Chatterjee 1975), (Chatterjee and Seneta 1977), (Cohen et al. 1986) and the references given therein. There is an interesting connection to the Delphi technique for pooling opinions of experts, (cf. (De Groot 1974)). The focus of these contributions is on consensus and how to reach it. This is true in particular for the encompassing study of Lehrer and Wagner (Lehrer and Wagner

1991) which emphasizes rational consensus as a fundamental feature from justice to epistemology. The importance of modelling disagreement beside consensus was pointed out especially by R. P. Abelson (Abelson 1964) and by N.E. Friedkin and E.C. Johnsen (Friedkin and Johnsen 1999). For example, there is the following remark in (Abelson 1964, p. 153): "Since universal ultimate agreement is an ubiquitous outcome of a very broad class of mathematical models, we are naturally led to inquire what on earth one must assume in order to generate the bimodal outcome of community cleavage studies. Offhand, there are at least three potential ways of generating such a bimodal outcome."

From French to Friedkin/Johnsen, as difficult as these models might be in their details, they are all comparatively simple in the sense that they are *all linear models*. This means in particular that the analysis needed can be carried out by powerful linear techniques such as matrix theory, Markov chains and graph theory. The first *nonlinear model* within this line of thinking was formulated and analyzed in (Krause 1997) and (Krause 2000). For this model see also (Beckmann 1997), (Dittmer 2000), (Dittmer 2001) and (Hegselmann and Flache 1998). A model similar in spirit has recently been investigated in (Weisbuch et al. 2001) and (Deffuant et al. 2000). Of course, having an explicit mathematical model does not mean at all that one has explicit mathematical answers. Already R. Abelson observed that a "More drastic revision of the model can be introduced by making the system nonlinear". (Abelson 1964, p. 150). Though elementary, the model is nonlinear in that the structure of the model changes with the states of the model given by the opinions of the agents (see Section 2). Not only that helpful mathematical tools like Markov chains are no longer applicable, it turns out, moreover, that rigorous analytical results are difficult to obtain. For that reason we carry out the analysis of the above nonlinear model to a large extent by simulations on the computer. Indeed, though we are proud to present analytical results, too, we want to emphasize the importance of careful computer simulations for social dynamics in general, and opinion dynamics in particular, whenever nonlinearities are involved. (See (Hegselmann and Flache 1998). For computer simulations on binary opinion dynamics see (Holyst et al 2001), (Latané and Nowak 1997), (Stauffer 2001), (Stocker et al. 2001).

In this paper we address the classical question of reaching a consensus as well as that of disagreement leading to polarization and, more general, the dynamics of *opinion fragmentation* of which *consensus* and *polarization* are just two interesting special cases. Section 2 of the paper develops and reviews within a unified framework four models of continuous dynamics: The classical model of consensus formation, a variation of this model due to Friedkin and Johnsen, a model with time-dependent (inhomogeneous) structure and a nonlinear model dealing with bounded confidence. Section 3 then presents for all these models major *analytical results* of which some are well-known and some are new. For the convenience of the reader we state the results with a minimum of mathematics. Precise mathematical definitions and theorems together with additional hints are delegated to an Appendix. Section 4 contains an extensive exploration of the nonlinear model with bounded confidence by *a series of computer simulations*. In the presentation we took care to explain step by step the computer experiments made.

2. MODELLING OPINION DYNAMICS

Consider a group of agents (or experts or individuals of some kind) among whom some process of opinion formation takes place. In general, an agent will neither simply share nor strictly disregard the opinion of any other agent, but will take into account the opinions of others to a certain extent in forming his own opinion. This can be modelled by different weights which any of the agents puts on the opinions of all the other agents. This process of forming the actual opinion by taking an average over opinions can be repeated again and leads, therefore, to a dynamical process in discrete time. Intuitively, one may expect that this process of repeatedly averaging opinions will bring newly formed opinions of different agents closer to each other until they flow into a consensus among all agents. In the next two sections we will show that the dynamics of opinion formation can be much

more complex than one would intuitively expect, even in the above simple model. The crucial point here is that the weights put on the opinions of others may change for reasons explained below. Only for the classical case of constant weights and enough confidence among agents the phenomenon of consensus is a typical one (cf., e.g., (De Groot 1974), (Lehrer 1975), and the pioneering but less formal (French 1956)).

Let n be the *number of agents* in the group under consideration. To model the repeated process of opinion formation we think of time as a number or rounds or of periods, that is a as *discrete time* $T = \{0,1,2,\cdots\}$. It will be assumed that the opinion of an agent can be expressed by a real number as, e.g., in the case of an expert who has to assess a certain magnitude. This assumption is made for simplicity, because already in that case the opinion dynamics considered can be quite intricate. This case is sometimes referred to as "continuous opinion dynamics" in contrast to the, even more restricted, case of "binary opinion dynamics" (cf. (Weisbuch et al. 2001)).

Later on we will also consider "higher dimensional" opinions (see the time-variant model in the next section). For a fixed agent, say i where $1 \leq i \leq n$, we denote the agents *opinion at time t* (in round t) by $x_i(t)$. Thus $x_i(t)$ is a real number and the vector $x(t) = (x_1(t), \cdots, x_n(t))$ in n-dimensional space represents the *opinion profile at time t*. Fixing an agent i, the *weight* given to any other agent, say j, we denote by a_{ij}. To keep things simple we introduce a_{ii} such that $a_{i1} + a_{i2} + \cdots + a_{in} = 1$. Furthermore, let $a_{ij} \geq 0$ for all i, j. Having these notations, opinion formation of agent i can be described as averaging in the following way

$$x_i(t+1) = a_{i1} x_1(t) + a_{i2} x_2(t) + \cdots + a_{in} x_n(t). \tag{2.1}$$

That is, agent i adjusts his opinion in period $t+1$ by taking a weighted average with weight a_{ij} for the opinion of agent j at time t. Of course, weights can be zero. For example, if agent i disregards all other opinions, this means $a_{ii} = 1$ and $a_{ij} = 0$ for $j \neq i$; or, if i follows the opinion of j then $a_{ij} = 1$ and $a_{ik} = 0$ for $k \neq j$. It is important to note that the *weights may change with time or with the opinion*, that is $a_{ij} = a_{ij}(t, x(t))$ can be a function of t and/or of the whole profile $x(t)$. By collecting the weights into a matrix, $A(t, x(t)) = (a_{ij}(t, x(t)))$, with n rows and n columns, we obtain a stochastic matrix, i.e., a nonnegative matrix with all its rows summing up to 1. Thus, using matrix notation, the general form of our model (GM) can be compactly written as

$$x(t+1) = A(t, x(t)) x(t) \text{ for } t \in T. \quad \text{(GM)}$$

The *main problem* we are dealing with in this paper is the following one: Given an *initial profile* (start profile) $x(0)$ and the dynamics specified by the weights – what can be said about the final behavior of the opinion profile, i.e., about $x(t)$ for t approaching infinity? In particular, when does the group of agents approach a *consensus c*, i.e., it holds $\lim_{t \to \infty} x_i(t) = c$ for *all* agents $i = 1, \cdots, n$?

In this generality, however, one cannot hope to get an answer, neither by mathematical analysis nor by computer simulations. Therefore, we will treat various interesting specializations of the above general model (GM). First, we begin with the *classical model* of fixed weights, i.e.,

$$x(t+1) = A x(t) \text{ for } t \in T, \quad \text{(CM)}$$

where A is a fixed stochastic matrix and $x(t)$ the column vector of opinions at time t. This model has been proposed and employed to assess opinion pooling by a dialogue among experts (De Groot 1974), (Lehrer 1975).

There is an interesting variation of this model developed by (Friedkin and Johnsen 1990, 1999). This model addresses opinion formation under social influence and assumes that agent i adheres to his initial opinion to a certain *degree* g_i and by a *susceptibility* of $1-g_i$ the agent is socially influenced by the other agents according to a classical model. By this variation the classical model becomes

$$x_i(t+1) = g_i x_i(0) + (1-g_i)\big(a_{i1} x_1(t) + \cdots a_{in} x_n(t)\big) \qquad (2.2)$$

or, in matrix notation,

$$x(t+1) = G x(0) + (I-G) A x(t) \quad \text{for } t \in T. \quad \text{(FJ)}$$

Here G is the diagonal matrix with the g_i, $0 \leq g_i \leq 1$, in the diagonal and I is the identity matrix. Obviously (CM) is a special case of (FJ), namely for $g_i = 0, 1 \leq i \leq n$. This model has been used to estimate from experiments the susceptibility of agents to interpersonal influence (French 1956).

A model similar to (CM) but in continuous time, that is a system of differential equations instead of difference equations, has been explored and applied rather early by (Abelson 1964). Though there are similarities in the search for consensus among agents, in the present article we will stick to the discrete time framework.

The models (CM) and (FJ), as well as Abelson's model, are both linear which makes them systematically tractable by analytical methods. The next type of model is still linear but time-variant (or, in the theory of Markov chains, inhomogeneous), that is

$$x(t+1) = A(t) x(t) \text{ for } t \in T, \qquad \text{(TV)}$$

where the entries of matrix $A(t)$, i.e., the weights, are dependent on time only. The time variant model (TV) portrays, e.g., the so called "hardening of positions" where agents put in the course of time more and more weight on their own opinion and less weight on the opinion of others. In the next section it will be shown that analytical results are still available, even for "higher dimensional" opinions, but the results are less sharp than in the time-invariant classical model.

The most difficult type of model occurs if the weights depend on opinions itself because then the model turns from a linear one to a *nonlinear* one. Thus, the model is of type (GM) where $A(x(t))$ does not explicitly depend on time. It is still completely hopeless to analyse the model in this generality. There is, however, a particular kind of nonlinearity which captures an important aspect of reality and seems at the same time tractable. But, compared with the other models, analytical insights are not so easy to obtain. For that reason we will investigate this model detailed and extensively by computer simulations in Section 4. The model we are going to exhibit portrays *bounded confidence* among the agents in the following sense. An agent i takes only those agents j into account whose opinions differ from his own not more than a certain *confidence level* ε_i. Fixing an agent i and an opinion profile $x = (x_1, \cdots, x_n)$ this set of agents is given by

$$I(i,x) = \big\{ 1 \leq j \leq n \,\big|\, |x_i - x_j| \leq \varepsilon_i \big\} \qquad (2.3)$$

where $|\cdot|$ denotes the absolute value of a real number. To make things not too complicated we assume that agent i puts an equal weight on all $j \in I(i,x)$. That is to say, in the light of the general model (GM) we let the weights given by $a_{ij}(x) = 0$ for $j \notin I(i,x)$ and $a_{ij}(x) = |I(i,x)|^{-1}$ for $j \in I(i,x)$ ($|\cdot|$ for a finite set denotes the number of elements). Thus the *model with bounded confidence* is given by

$$x_i(t+1) = \left| I(i, x(t)) \right|^{-1} \sum_{j \in I(i, x(t))} x_j(t) \quad \text{for } t \in T. \quad \text{(BC)}$$

This model has been developed by (Krause 1997, 2000); see also (Beckmann 1997) and (Dittmer 2000, 2001). For another attempt to model a lack of confidence see (Deffuant et al. 2000), (Weisbuch et al. 2001). There, in a pairwise comparison between agents, "opinion adjustments only proceed when opinion difference is below a given threshold" ((Deffuant et al. 2000) , p. 2).

3. ANALYTICAL RESULTS FOR THE VARIOUS MODELS OF OPINION DYNAMICS

In the following we present several major results for the various models. Some are well known, others are less well known and some are new results.

A. The classical model

The classical model (CM) was given by

$$x(t+1) = A x(t) \quad \text{for } t \in T = \{0, 1, 2, \cdots\},$$

where A is a stochastic matrix. Obviously, $x(t) = A^t x(0)$ for all $t \in T$ and, hence, the analysis amounts to analyze the powers of a given matrix.

Result 1: On consensus

- If any two agents put jointly a positive weight on a third one then for every initial opinion profile a consensus will be approached (which consensus, of course, depends on the initial profile).

- More general, a consensus will be approached for every initial opinion profile if and only if finally for t big enough any two agents put jointly a positive weight on a third one.

For a formal statement of this result together with additional hints see Theorem 1 in the Appendix.

A special case of Result 1 is given if every agent puts a positive weight on any other agent. This corresponds to the intuition that the group should approach a consensus if every agent takes the opinion of any other agent into account. On the other extreme is the case where every agent sticks to his opinion without taking care of any other opinion. In this case A is the identity matrix and, of course, there is no consensus, except the special case of a consensus already in the initial opinion profile. Result 1 describes the confidence pattern between these extremes that allows for a consensus. An example for three agents is given by the stochastic matrix

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{4} & 0 & \frac{3}{4} \end{bmatrix}.$$

For this matrix A the power A^2 is strictly positive (does not contain any zero). A matrix for which some power is strictly positive is also called *primitive*. By Result 1 a primitive matrix A is sufficient for consensus. (The second part of Result 1 provides a condition that is not only sufficient but necessary, too.)

Moreover, by employing what is called the normal form of Gantmacher, a more refined result is possible (see Theorem 2 in the Appendix). Intuitively, the group of agents can be split up into subgroups such that the first g isolated subgroups consist only of *essential agents* in the sense that an agent puts weight only on agents in his group. Such a subgroup has *primitive structure* if its submatrix of weights is primitive.

Result 2: On opinion fragmentation

- For any given initial profile the opinion dynamics approaches a stable opinion pattern if and only if the subgroups of essential agents are all primitive. For the final stable opinion pattern only the initial opinions of essential agents play a role.

- In particular, the stable opinion pattern reduces to a consensus if and only if there exists just one subgroup of essential agents $(g=1)$.

(For more information see theorem 2 in the Appendix.)

An example of four agents is given by

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

There are two subgroups of essential agents formed by agents 1 and 2 and agents 3 and 4, respectively. The final stable opinion pattern consists of a partial consensus between agents 1 and 2 and a partial consensus between agents 3 and 4. In general one obtains no consensus between these subgroups.

B. The Friedkin-Johnsen model

This model (FJ) was given by

$$x(t+1)=G\,x(0)+(I-G)\,A\,x(t) \text{ for } t\in T.$$

If $G=0$ then (FJ) specializes to the classical model and it suffices to discuss $G\neq 0$.

Result 3: On opinion fragmentation with positive degrees

- If there is at least one agent with positive degree and there are "enough positive weights" then for any given initial profile the opinion dynamics approaches a stable opinion pattern that can be computed from the weights, the degrees, and the initial profile.

- The above stable pattern represents a consensus if and only if there prevails already a consensus among all agents with positive degree.

(For more information see Theorem 3 in the Appendix.)

Thus, in the case of positive degrees one can expect a consensus only for very special initial profiles. Assuming a primitive structure for the whole group this is very different from what holds for the classical model.

An example for four agents is given by

$$A = \begin{bmatrix} .220 & .120 & .359 & .300 \\ .147 & .215 & .344 & .294 \\ 0 & 0 & 1 & 0 \\ .089 & .178 & .446 & .286 \end{bmatrix}, G = \text{diagonal of } A$$

In this case all agents have positive degree and a consensus is possible only if there is a consensus at the beginning. For $x(0)=(25,25,75,85)$, e.g., one obtains as final opinion pattern (60, 60, 75, 75) (Friedkin and Johnsen 1999, p. 6). Since A has a strictly positive column, in the classical model one would obtain for $x(0)$ as above a consensus according to Result 1.

C. The time-variant model

The model with time-variance (TV) was given by

$$x(t+1) = A(t)x(t) \text{ for } t \in T,$$

where the weights collected in matrix $A(t)$ depend on time t. As one might expect from the classical model, the time variance will be no obstruction to a consensus as long as the weights remain sufficiently positive. If, however, the weights tend to zero very quickly then consensus disappears. This interesting phenomenon we illustrate by a simple example which works already for a group of two agents. Suppose agent 1 does never care about the opinion of agent 2. In a first scenario let agent 2 put at time t, for $t \geq 2$, the weight t^{-1} to agent 1 and the remaining weight $1 - t^{-1}$ to himself. Thus, the position of agent 2 is hardening by putting less and less weight to agent 1. It is not difficult to verify that $x_2(t) = (1-t^{-1})x_1(2) + t^{-1}x_2(2)$ and, because of $x_1(t) = x_1(0)$ for all t, that $x_2(t)$ tends to $x_1(0)$. Therefore, the two agents approach always a consensus in spite of hardening the positions. In a second scenario let agent 2 hardening his position faster, by putting at time t, for $t \geq 2$, the weight t^{-2} to agent 1 and $1 - t^{-2}$ to himself. A little computation shows that $x_2(t)$ tends to $\frac{1}{2}x_1\big((2) + x_2(2)\big)$ and, of course, $x_1(t) = x_1(0)$ for all t. Thus, for many initial profiles a consensus will not be approached. This little example shows that approaching a consensus does depend on the speed by which the weights are changing (cf. (Cohen 1986)). This vague insight is made more precise by the following result.

Result 4: One time-variance consensus

Consensus will be approached for every initial opinion profile, provided there exists a sequence of time points with the following property:

For the accumulated weights b_{ij} between time points it holds that for any two agents i and j there exists a third one k such that b_{ik} and b_{jk} are positive and the minima of both sum up to infinity for summing over all time points. Roughly speaking, the result says that for a consensus the weights cannot tend too fast to zero because their sum should be infinite. For a precise statement see Theorem 4 in the Appendix. The above result can be formulated and proved also for the case of multidimensional opinions, that is, $x_i(t)$ is not just a number but a vector in higher dimensions. The proof for this result is telling because it shows that the set in higher dimensions which is spanned by the agents' opinions must shrink to the point of consensus if time develops. (See the Lemma before Theorem 4 in the Appendix.)

D. Opinion dynamics with bounded confidence (BC)

The (BC) model was given by

$$x_i(t+1) = \big|I\big(i, x(t)\big)\big|^{-1} \sum_{j \in I(i, x(t))} x_j(t) \text{ for } t \in T,$$

where $I(i, x) = \Big\{ 1 \leq j \leq n \,\big|\, \big|x_i - x_j\big| \leq \varepsilon_i \Big\}$

and $\varepsilon_i > 0$ is the given confidence level of agent i. Though some properties hold still for different levels of confidence, in the following we will assume throughout a *uniform level of confidence*, i.e., $\varepsilon_i = \varepsilon$ for all agents i. Later on, in the next section when dealing with simulations, we will consider not only *symmetric confidence* intervals, i.e., $[-\varepsilon, +\varepsilon]$, but also *asymmetric confidence* intervals,

i.e., $\left[-\varepsilon_l, +\varepsilon_r\right]$ where $\varepsilon_l = \varepsilon_{left}$, $\varepsilon_r = \varepsilon_{right}$ are positive but can be different from each other. The set $I(i,x)$ in the asymmetric case is then given by

$$I(i,x) = \left\{ 1 \le j \le n \,\middle|\, -\varepsilon_l \le x_j - x_i \le \varepsilon_r \right\}.$$

In the asymmetric case $\varepsilon_l \neq \varepsilon_r$ we can have a *one sided split* between two agents i and j, namely $\varepsilon_l < x_j - x_i \le \varepsilon_r$ if $\varepsilon_l < \varepsilon_r$ and $\varepsilon_r < x_j - x_i \le \varepsilon_l$ if $\varepsilon_r < \varepsilon_l$. In a one sided split one agent (i if $\varepsilon_l < \varepsilon_r$ and j if $\varepsilon_r < \varepsilon_l$) takes the other agent (j and i, respectively) into account, but not vice versa.

The confidence levels $\varepsilon_l, \varepsilon_r$ serve as parameters of the model and the set of possible values of $(\varepsilon_l, \varepsilon_r)$ is called the *parameter space*.

A particular feature of this model, compared with the other models discussed, is that a consensus will be reached in finite time, if there is a consensus at all. For an opinion profile $x = (x_1, \cdots, x_n)$ we say that there is a *split* (or crack) between agents i and j if $\left| x_i - x_j \right| > \varepsilon$.

An opinion profile $x = (x_1, \cdots, x_n)$ we call an ε-*profile* if there exists an ordering $x_{i_1} \le x_{i_2} \le \cdots \le x_{i_n}$ of the opinions such that two adjacent opinions are within confidence, i.e.,

$$x_{i_{k+1}} - x_{i_k} \le \varepsilon \quad \text{for all} \ 1 \le k \le n-1$$

First we collect some fundamental properties of this model for the case $\varepsilon_l = \varepsilon_r$ (see (Krause 2000)).

Properties

I. The dynamics does not change the order of opinions, i.e., $x_i(t) \le x_j(t)$ for all $i \le j$ implies that $x_i(t+1) \le x_j(t+1)$ for all $i \le j$.

II. If a split between two agents occurs at some time it will remain a split forever.

III. If for an initial profile a consensus is approached then the opinion profile must be an ε-profile for all times.

IV. For $n = 2, 3, 4$ a consensus is approached if and only if the initial profile is an ε-profile.

The last property (IV) is no longer true for $n = 5$. That is, for five agents it can happen that they don't approach a consensus though their initial opinions are close in the sense of an ε-profile. If, however, the initial profile is an equidistant ε-profile, i.e., the distance between adjacent opinions is exactly ε, then for a number of five agents, too, a consensus will be approached. A careful consideration shows that for a number of six agents even in the case of an equidistant ε-profile a consensus cannot be reached. These remarks indicate that the number of agents involved can matter for the dynamics. For the simplest case of just two agents it is not difficult to give a complete analysis of the dynamics. For an arbitrary n, however, the mathematical analysis of the dynamics turns out to be rather difficult and up to now there are only a few general results available. For that reason we will present in the next section an extensive analysis of this model for higher values of n by computer simulation. Of the few general results we present two fundamental ones.

Result 5: On consensus for bounded confidence

Consensus will be approached for a given initial opinion profile, provided for an equidistant sequence of time points the following property holds:

For any two agents i and j there exists a third one k such that a chain of confidence leads from i to k as well as from j to k. In this case, the consensus is reached in finite time.

(See Theorem 5 in the Appendix for a precise formal statement.)

Result 6: On opinion fragmentation for bounded confidence

For any given initial profile there exists a finite $t^* \in T$ and a division of all agents into maximal subgroups such that within each of the subgroups there holds a consensus for all $t \geq t^*$. (Of course, those partial consenses will be different from each other in general.)

(See Theorem 6 in the Appendix for a precise formulation.)

4. SIMULATIONS

In this section we will explore the model with bounded confidence by means of simulations. We want to do that, firstly, in a systematic way, and , secondly, following the KISS-principle: *Keep it simple, stupid!* Under that principle it is fairly natural to start in a *first* step with homogenous and symmetric ε-intervals. *Homogeneity* means that the size and shape of the confidence interval is the same for *all* agents. *Symmetry* means, that the interval has the same size to the left *and* to the right, i.e. $\varepsilon_l = \varepsilon_r$. In a *second* step we will analyse different types of asymmetry.

4.1 SYMMETRIC CONFIDENCE

With continuous, one–dimensional, and normalized opinions x taken from the interval $x \in [0,1]$ only confidence intervals $0 \leq \varepsilon_l, \varepsilon_r \leq 1$ make sense. Thus, the parameter space is given by the unit square. The diagonal represents symmetric confidence. A systematic approach could be 'walking along the diagonal', starting the trip at the point $\langle 0,0 \rangle$.
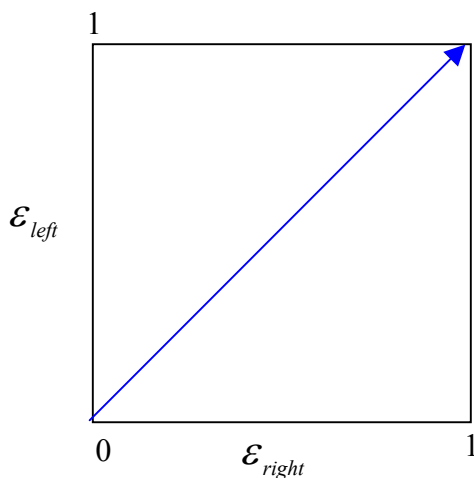

Figure 1: The parameter space, walking along the diagonal.

We will randomly generate a start distribution of 625 opinions. Updating is *simultaneous*. Figure 2 shows three stops on the tour along the diagonal. The stops are single runs. They all get going with the *same* start distribution. The ordinate indicates the opinions. Since it is useful for the following analysis the opinions are additionally encoded by colours ranging from red ($x = 0$) to magenta ($x = 1$). The abscissa represents time and shows the first 15 periods. Obviously it takes less than 15 periods to get a stable pattern: With the quite small confidence interval of $\varepsilon_l = \varepsilon_r = 0.01$ exactly 38 different opinions survive in the end. Under the much bigger confidence of $\varepsilon_l = \varepsilon_r = 0.15$ the agents end up in *two camps*, and with $\varepsilon_l = \varepsilon_r = 0.25$ the result of the dynamics is *consensus*. As it looks, the size of the confidence interval really matters.

Figure 2 shows only *single* runs. To get a better feeling of what is going on we run systematically simulations walking along the diagonal. The simulations start with $\varepsilon_l = \varepsilon_r = 0.01$, $\varepsilon_l = \varepsilon_r = 0.02$, …, $\varepsilon_l = \varepsilon_r = 0.4$. (For reasons that become obvious below, there is nothing new and interesting in the parameter space for $\varepsilon_l, \varepsilon_r > 0.4$.) For each of these 40 steps we repeat the simulation 50 times, always starting with a different random start distribution. Each run is continued until the dynamics becomes stable. *Figure 3* gives an overview.
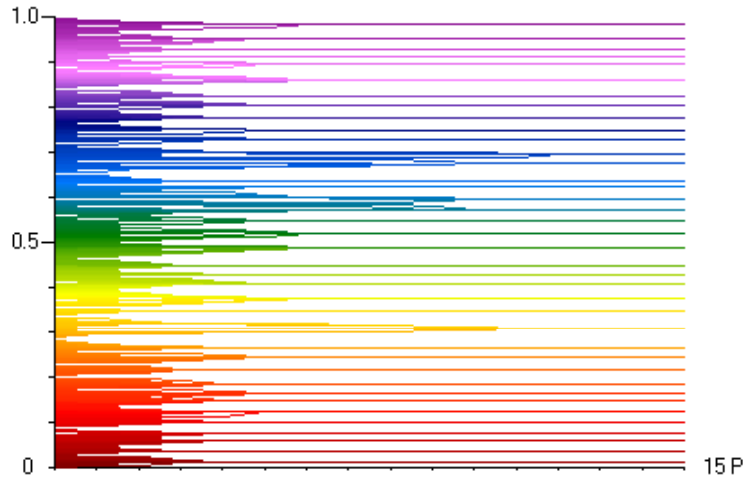
(a) $\varepsilon_l = \varepsilon_r = 0.01$
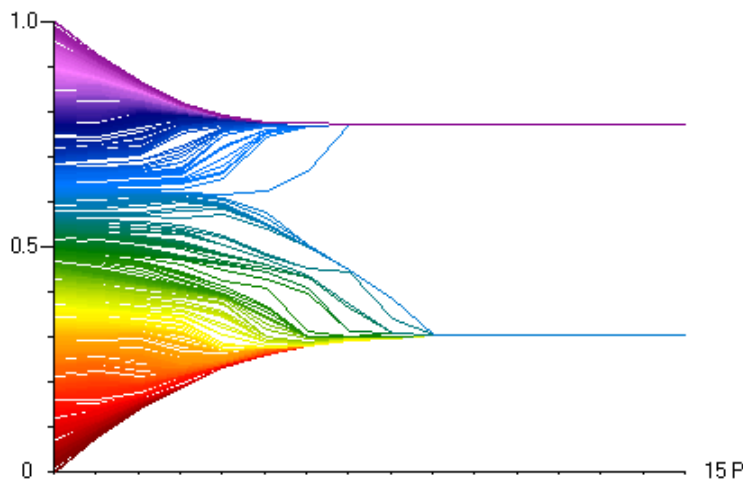
(b) $\varepsilon_l = \varepsilon_r = 0.15$
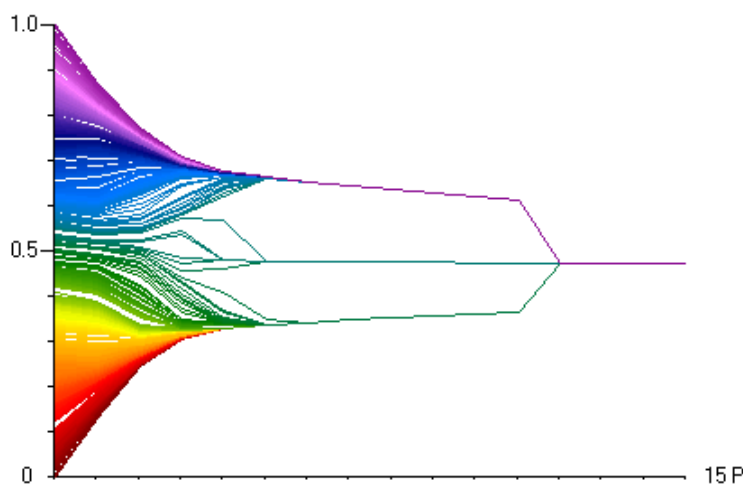
(c) $\varepsilon_l = \varepsilon_r = 0.25$

Figure 2: Stops while walking along the diagonal
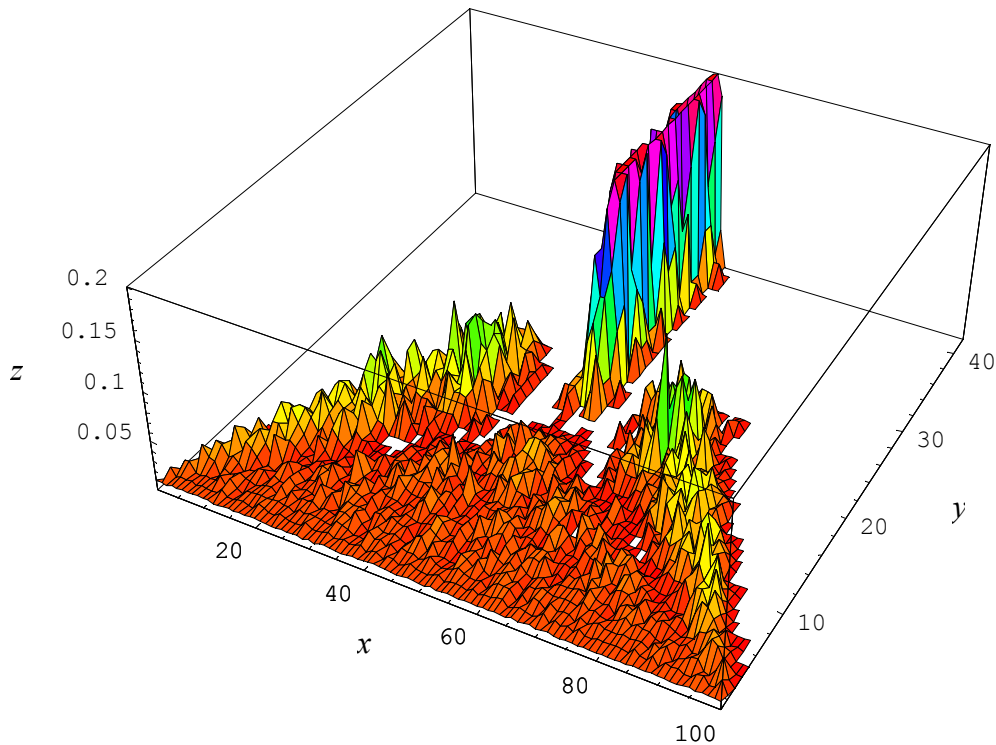
Figure 3: Walking along the diagonal – simulation results.

The *x-axis* in *Figure 3* represents the opinion space $[0,1]$ divided into 100 intervals. Our 40 step walk along the diagonal is represented by the *y-axis*. (Note: the steps are *not* time steps!) The *z-axis* represents the average (!) relative frequencies of opinions in the 100 opinion intervals of the opinion space after the dynamics has stabilised.

Figure 3 deserves careful inspection: At the beginning of our walk along the diagonal of *Figure 1*, i.e. the *y-axis* of *Figure 3*, there is only little confidence. For example, step 2 or 3 means that $\varepsilon_l = \varepsilon_r = 0.01$ or $\varepsilon_l = \varepsilon_r = 0.02$. In terms of *single* runs we are speaking about opinion dynamics like that in *Figure 2a*. The *z*-values show that as an average we find under that little confidence a small fraction of the opinions in *all* intervals of the opinion space. (Note: That does *not* imply that in a *single* run all intervals are occupied after stabilisation; see *Figure 2a* and *Figures 12a, 12b*.) *No part* of the opinion space seems to have especially high or especially low frequencies. That changes as we step further along the diagonal. Look, for instance, at step 15, i.e. $\varepsilon_l = \varepsilon_r = 0.15$. A *single* run example for a dynamics based on confidence intervals of this size is given by *Figure 2b*, where we end with two camps holding different opinions. *Figure 3* makes it clear that this is a fairly typical result: As we step forward along the diagonal the average distribution of stabilised relative frequencies becomes less and less uniform. To the left and to the right of the centre mountains of increasing height emerge. As we continue our walk on the diagonal the landscape changes dramatically again: At about step 25, i.e. $\varepsilon_l = \varepsilon_r = 0.25$ the *'Lefty'* and the *'Righty Mountains'* come to a sudden and steep end. At the same time a new and steep centre mountain emerges. An example for a *single run* in this region of the parameter space is given by *Figure 2c*. *Figure 3* shows that this single run with a total confidence interval including half of the opinion space is a typical run, leading to a consensus including all or almost all agents. Except in the centre almost all opinion intervals are empty and the corresponding opinions are eliminated.

Summing up and generalising what we see in *Figure 3* one might say: As the homogeneous and symmetric confidence interval increases we transit from phase to phase. More exactly, *we step from fragmentation (plurality) over polarisation (polarity) to consensus (conformity)* .

That a dynamics, governed by averaging among those opinions which are within a certain fairly small confidence interval, leads to an evenly distributed variety of opinions (plurality) will probably not surprise that much (see Section 3, part D, Result 6). That fairly large confidence intervals drive our dynamics towards conformity is probably even less surprising (see Section 3, part D, Result 5). Big confidence intervals should drive all opinions direction centre and there they converge. (For all values $\varepsilon_l, \varepsilon_r > 0.4$ the result is always conformity, and that is the reason why we stop the walk on the diagonal at the point $\varepsilon_l = \varepsilon_r = 0.4$.) Thus, the real task of understanding is the phase in between: *What* mechanisms drive our dynamics under 'middle sized' confidence into *polarisation*?
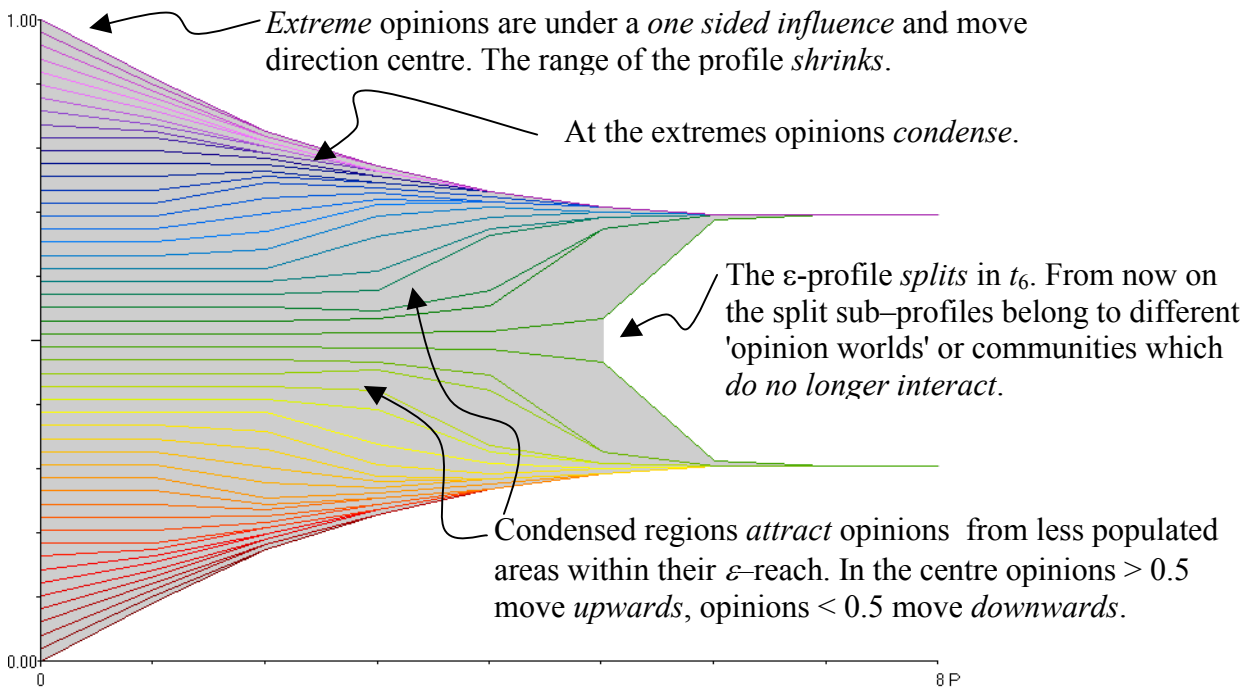
Figure 4: Regular start profile, $\varepsilon_l = \varepsilon_r = 0.2$.

Figure 4 gives decisive hints to understand polarity. The dynamics starts with a *regular* profile, i.e. a profile for which the distance between any two *neighbouring* opinions is the *same*. A grey area between two neighbouring opinions indicates that the distance between the two is not greater than $\varepsilon_l, \varepsilon_r$. Careful inspection shows:

- *Extreme* opinions are under a *one sided influence* and move direction centre. As a consequence the range of the profile *shrinks*.
- At the extremes opinions *condense*.
- Condensed regions *attract* opinions from less populated areas within their ε–reach.
- In the centre opinions > 0.5 move *upwards*, opinions < 0.5 move *downwards*.
- The ε–profile *splits* in period t_6. From now on the split sub–profiles belong to different 'opinion worlds' or communities which *do no longer interact*.

Figure 5 shows one aspect of the dynamics in *Figure 4* more in detail. The *x*–axis gives the periods, the *y*–axis of *Figure 5* shows the change from one period to the next, i.e.

$$\Delta_i (t-1) = x_i (t) - x_i (t-1) \quad t = 1, 2, \cdots \qquad (4.1)$$

for all opinions from *Figure* 4. The opinions are indicated by their colour.
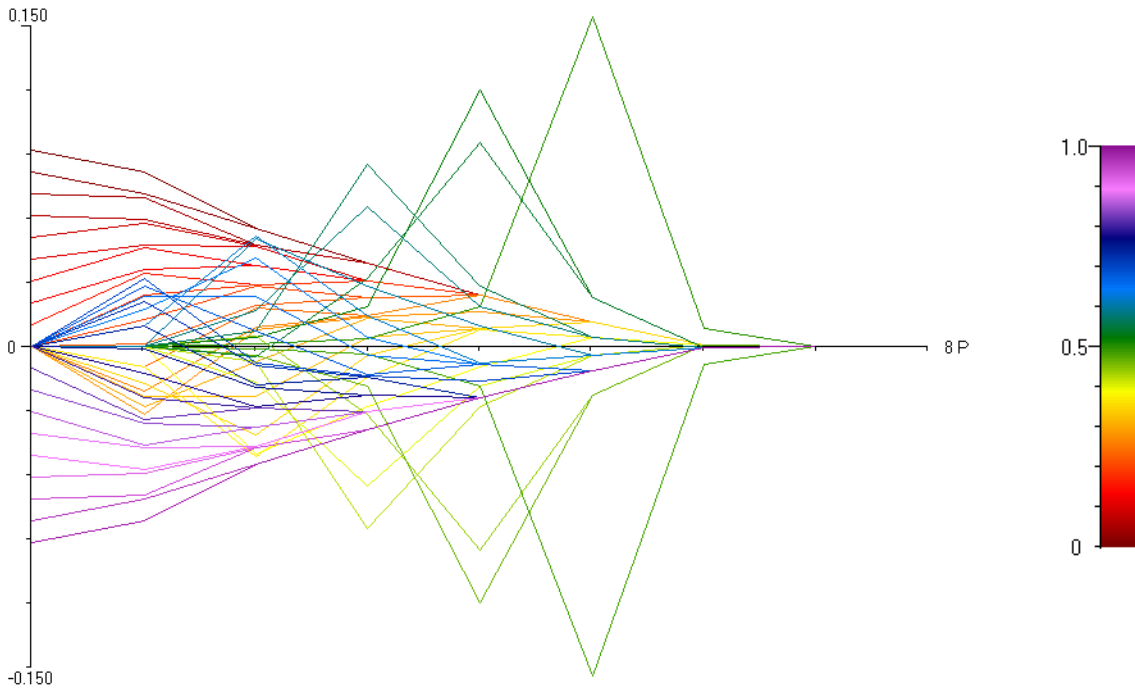
Figure 5: Opinion changes from period to period (50 opinions, regular start profile,
$\varepsilon_l = \varepsilon_r = 0.2$ *).*

Figure 5 makes clear that opinion changes start at the extremes and are there (at the start!) most extreme for the most extreme opinions. Opinions next to 1 (magenta) move downwards (negative Δ_i), opinions next to 0 (red) move upwards (positive Δ_i). From period 0 to 1 nothing changes more in the middle of the opinion space. But as times goes by changes work through the opinion space direction centre. The opinions directly in the centre are the last to be affected. They make the biggest 'jumps'. From period 7 to 8 onwards nothing changes anymore, i.e. $\Delta_i = 0$ for all opinions x_i.

The effects described above become in some respect more obvious with more opinions. *Figure 6* shows for 500 opinions what *Figure 5* showed for 50. The opinions of the regular profile are again indicated by their colours. Since colouring of the profile is done sequentially, an earlier coloured opinion may later become hidden by other opinions. Figure 6, top shows a colouring of the profile in an ascending order (0 to 1) while Figure 6 follows a descending order (1 to 0). By visual inspection it becomes immediately clear that changes start at the extremes and reach the centre only with some delay.

The decisive key for an understanding of polarity are obviously the splits in the ε-profile, induced by a one sided–influence at the extremes, a shrinking range of opinions combined with an increasing frequency of opinions in certain areas of the opinion space which attract opinions in less 'populated' areas within their reach. Under simultaneous updating cracked profiles can never get connected again. Agents/opinions outside ones own sub profile are 'out of range/reach' in a quite severe sense: They are not only outside ones own confidence interval, but also outside the confidence intervals of all others one takes seriously, outside the confidence intervals of all others which those others take seriously etc.
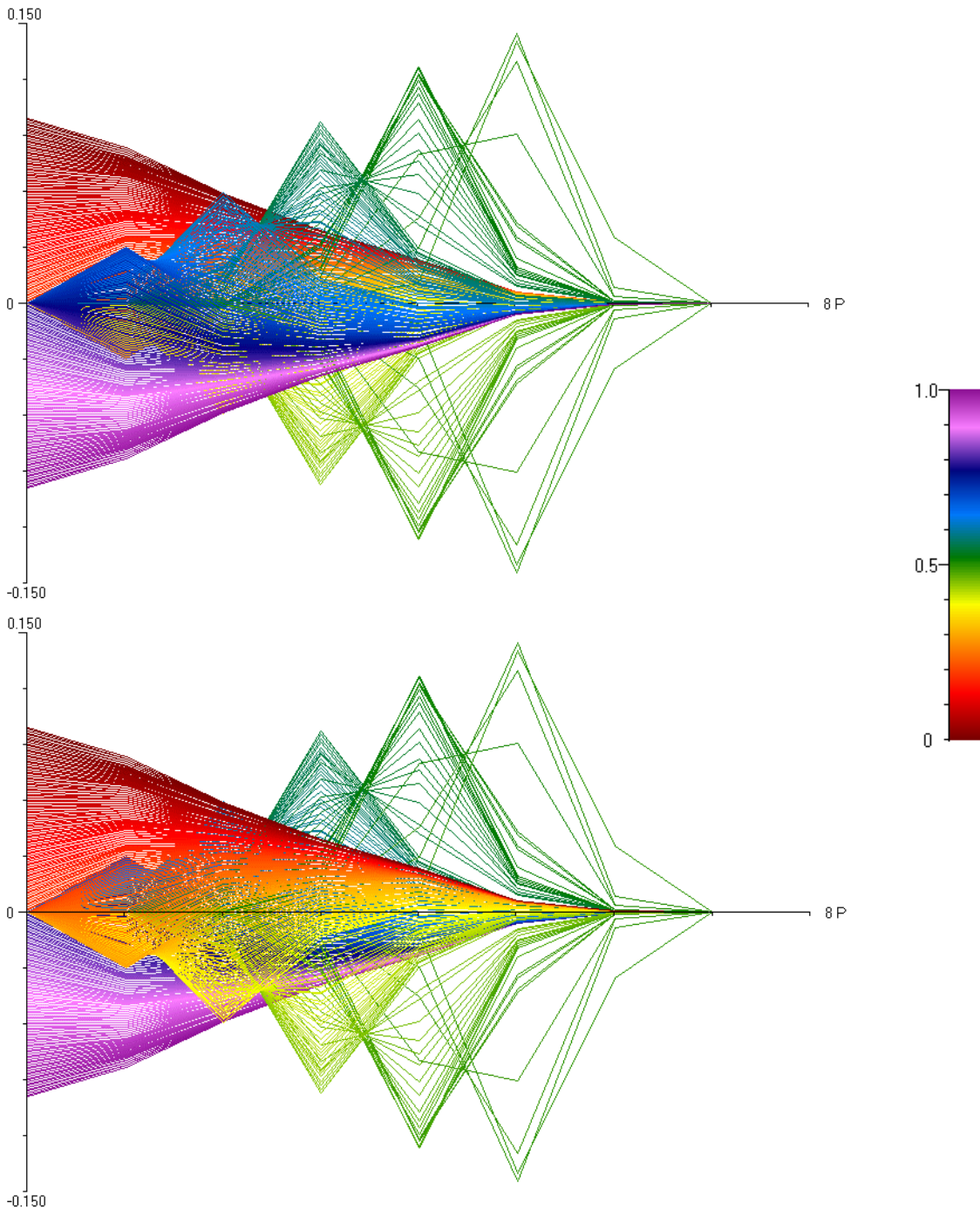
Figure 6: Opinion changes from period to period (500 opinions, regular start profile, $\varepsilon_l = \varepsilon_r = 0.2$). *Top: Colouring in an ascending order. Bottom: Colouring in an descending order.*

The splits do not only explain polarity, they explain the stabilisation of our dynamics in general. For fairly small confidence intervals the stabilisation leads to a fairly high number of surviving opinions, i.e. *plurality. Figure* 7 shows a regular start profile with 100 opinions and $\varepsilon_l = \varepsilon_r = 0.05$. The profile splits 8 times. Again a grey area between two neighbouring opinions indicates that the distance between the two is not greater than $\varepsilon_l, \varepsilon_r$. For 'middle sized' confidence intervals we get only a small number of surviving opinions, i.e. *polarisation; Figure 4* gives an example. Under large confidence intervals the profile splits never or the splits leave alone extreme and quite small minorities while an overwhelming majority converges in the centre of the opinion space. *Figure 8* gives an

example of total *consensus*. Note that even the four opinions in the centre of the opinion space move for a short while out of the centre. But they merge there again much earlier than the upper and lower part of the profile arrives in the centre as well.
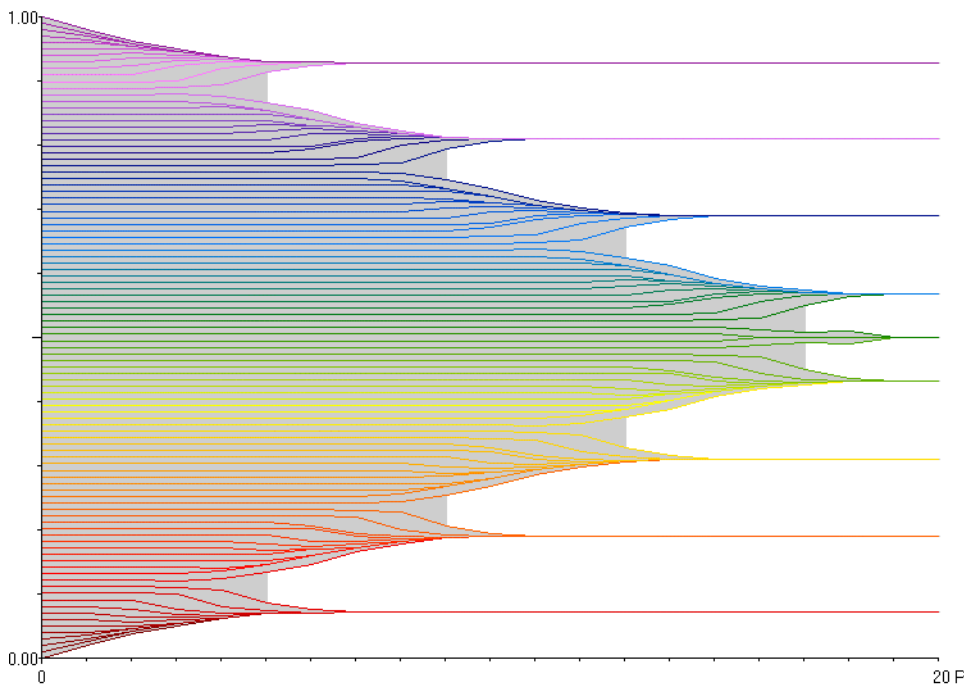

Figure 7: 100 opinions, $\varepsilon_l = \varepsilon_r = 0.05$, 8 splits.
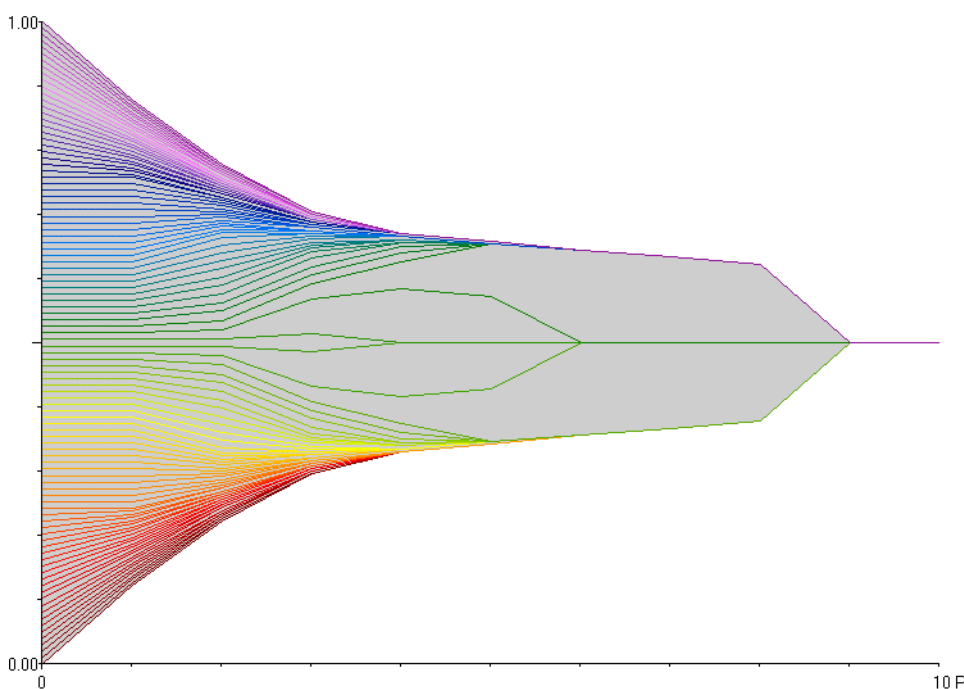

Figure 8: 100 opinions $\varepsilon_l = \varepsilon_r = 0.25$, no split, total consensus.

The simulation results from *Figure 3* are based on *randomly* generated start profiles (uniform distribution), not on regular start profiles. But the effects described so far do not essentially depend on the regularity of the start profile. What irregularity adds are *density fluctuations* in the initial distribution of opinions. They are additional causes for splits and induce opinion changes deep inside the start–profile without any delay right at the beginning of the process.

4.2 ASYMMETRIC CONFIDENCE

Up till now we considered walks along the diagonal of the parameter space (*Figure* 1), i.e. symmetry in the sense that $\varepsilon_l = \varepsilon_r$. But confidence may be *asymmetric*. One can think of several types of asymmetry. In the following we will analyse two cases. In the *first* case (4.2.1) the asymmetry is *independent* of the opinion an agent holds. Whatever their opinion might be, all agents have the same confidence intervals $\varepsilon_l, \varepsilon_r$ with $\varepsilon_l \neq \varepsilon_r$. In the *second* case (4.2.2) the asymmetry is *dependent* upon the opinion the agent holds: An agent with an opinion more to the right [left] has more confidence in the right [left] direction.

4.2.1 OPINION INDEPENDENT ASYMMETRY

How to get an overview about what is going on under opinion independent asymmetries? Again our approach is a systematic walk through the parameter space. But instead of taking the route along the diagonal we now walk on straight lines *below* (or *above*) the diagonal as indicated by *Figure 9*. For this type of asymmetry it does not matter whether confidence is biased to the right *or* biased to the left. Therefore it is only *one* of the triangles, either the one below or the one above the diagonal, that has to be analysed. We confine ourselves to the triangle *below*, i.e. a bias to the right. For all effects we find in that area of the parameter space there exist corresponding effects in the triangle above (bias to the left).
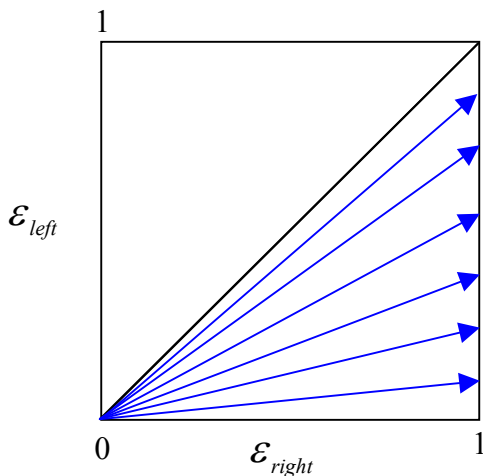

Figure 9: Research strategy for opinion independent asymmetries.

To get a first feeling we look at the three *single* run examples of asymmetric confidence in *Figure 10*. We find phenomena we are already familiar with, for instance fast stabilisation, plurality, polarity, and conformity. But obviously the asymmetric confidence drives the dynamics somehow into the direction favoured by the asymmetry, i.e. to the right.
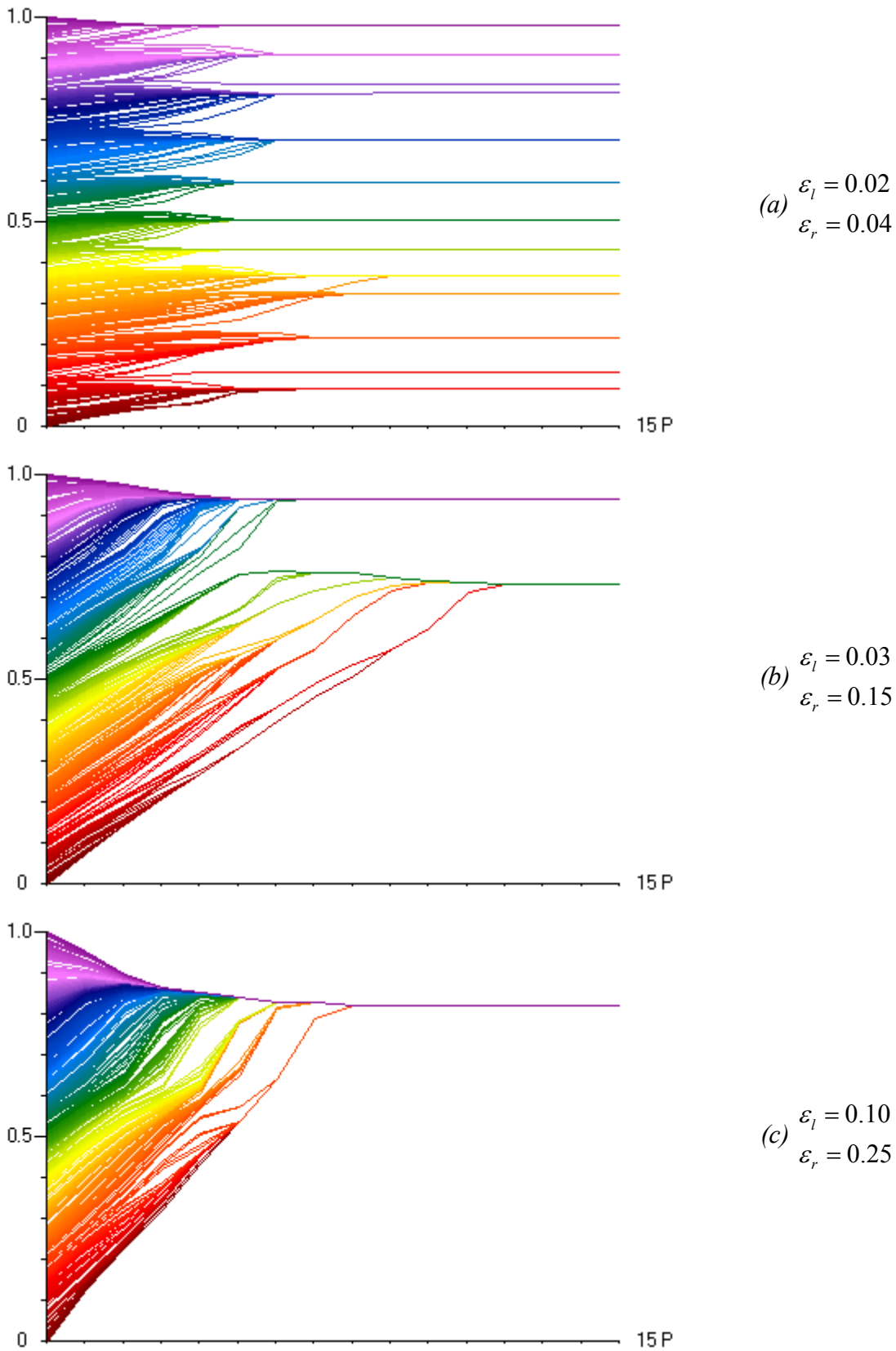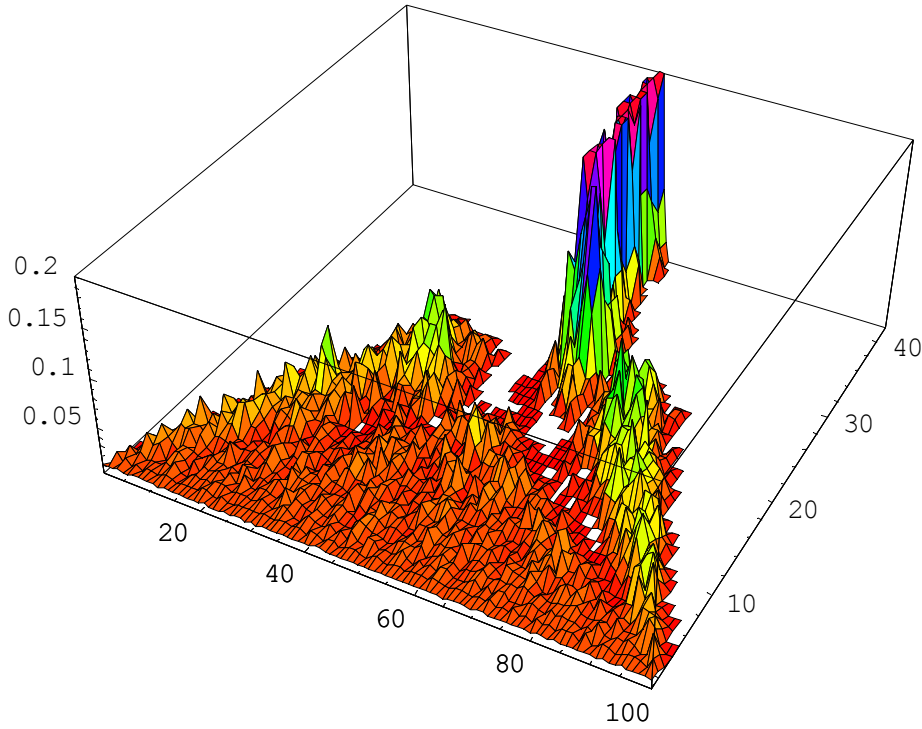
Figure 10: Single runs, 625 opinions, random start profile.

To get a more systematic overview we present the results of four stepwise walks below the diagonal. In the first walk we follow the straight line $\varepsilon_l = 0.9\varepsilon_r$. We start with $\varepsilon_r = 0.01$ and make 40

steps until we get to the point $\varepsilon_l = 0.36, \varepsilon_r = 0.4$. For each value of these 40 steps we repeat the simulation 50 times, always starting with a different random (uniform) start distribution. Each run is continued until the dynamics becomes stable. The other three walks follow in the same way the lines $\varepsilon_l = 0.75\varepsilon_r$, $\varepsilon_l = 0.5\varepsilon_r$, and $\varepsilon_l = 0.1\varepsilon_r$. We always stop when $\varepsilon_r = 0.4$ is reached. *Figure 11* gives an overview.
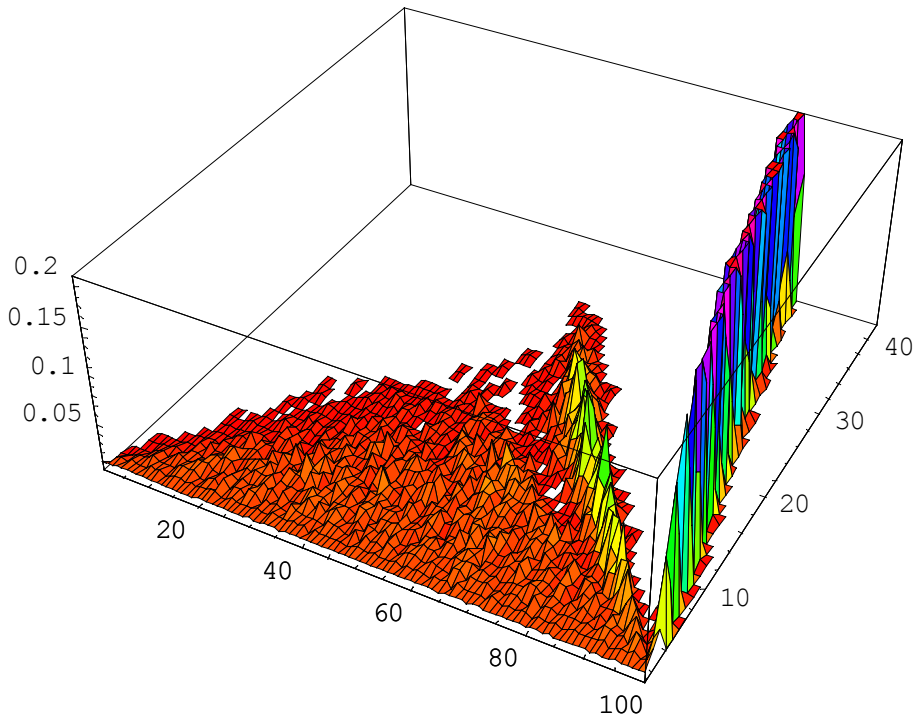

(a) $\varepsilon_l = 0.9\varepsilon_r$
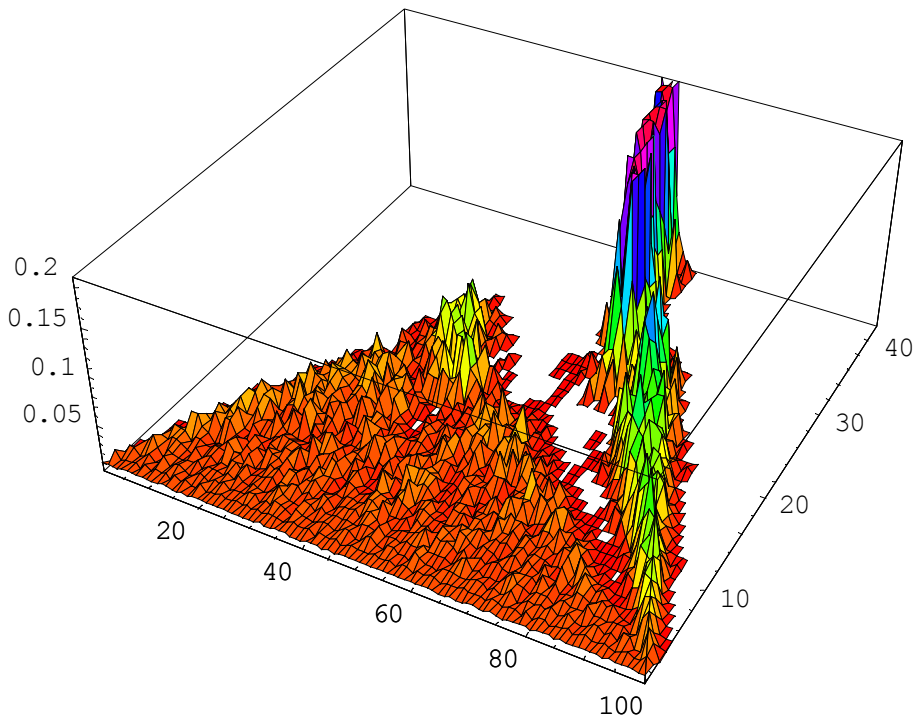
(b) $\varepsilon_l = 0.75\varepsilon_r$
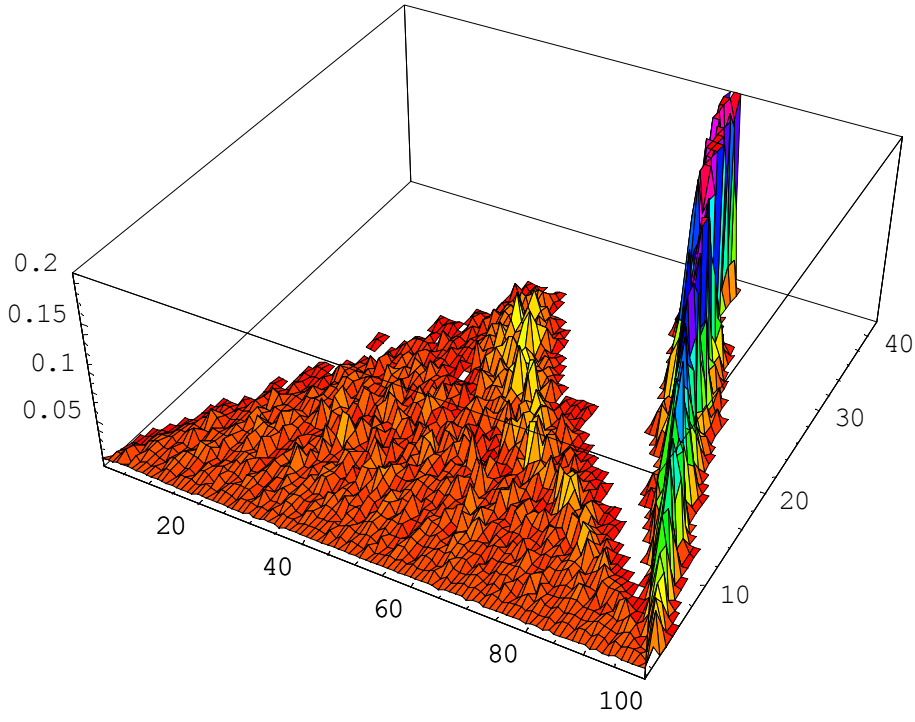
(c) $\varepsilon_l = 0.5\varepsilon_r$

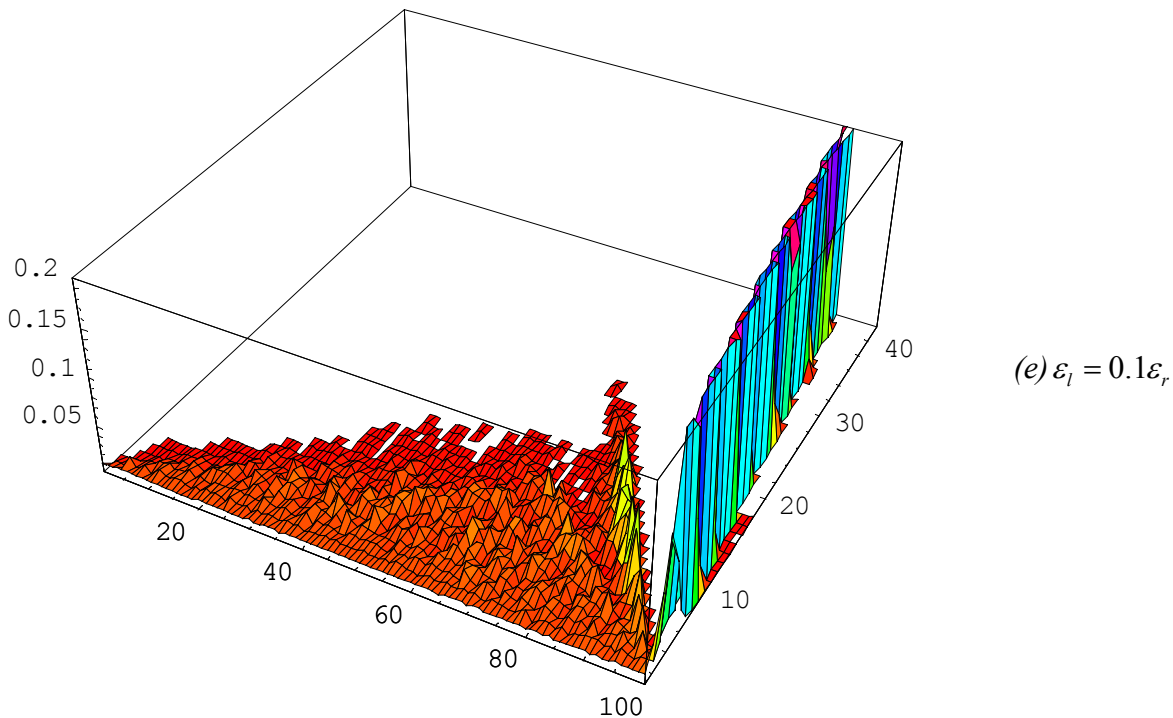(d) $\varepsilon_l = 0.25\varepsilon_r$

(e) $\varepsilon_l = 0.1\varepsilon_r$

Figure 11: Walking below the diagonal – simulation results.

Inspection of *Figure 11a–e* supports the following observations:

- As ε_r (and thereby ε_l) increases all four walks finally lead again into a region of the parameter space where *consensus* prevails. But as ε_l compared to ε_r becomes smaller and smaller the resulting consensus moves into the favoured, i.e. here into the right direction.

- For a very small ε_r and ε_l the dynamics stabilises with a lot of surviving opinions. Thus again we have a phase one might coin *plurality*.

- As ε_r and ε_l increases polarisation emerges. As long as ε_l is only a little bit smaller than ε_r (cf. *Figure 11a*, $\varepsilon_l = 0.9\varepsilon_r$) it is the type of polarisation we know from the symmetric case: in the left and in the right of the of the opinion space extreme opinion camps emerge, grow, and get closer to each other with an increasing ε_r (and thereby increasing ε_l). It is a somehow a *'symmetric'* polarisation: The camps have about the same size and the same distance from the centre (or the borders, respectively) of the opinion space.

- As ε_r becomes significantly greater than ε_l we observe a new type of *asymmetric* polarisation. The most obvious effect is that a big opinion camp emerges at the right border of the opinion space. This effect is extreme if ε_l is only a small fraction of ε_r (cf. *Figure 11e*). To the left of this main camp − in a certain distance, but still to the right of the centre of the opinion space– we observe smaller but nevertheless outstanding frequencies. This reflects the fact, that asymmetric confidence tends to produce in a certain region of the parameter space two or few opinion camps of different size: The bigger one normally more to the border of the opinion space.

Figures 11a – 11e (and similarly *Figure 3*) show how many opinions on an average over 50 simulation runs end up (after stabilisation) in each of the 100 intervals in which the opinion space was divided. Thus, one *does not generally* see, how many opinions on the average survive at all. This information is given by *Figures 12a* and *12b*. Both figures show the average number of different

opinions that survive after the dynamics has stabilised. The results are based on 25 simulation runs for each pair $\langle e_l, \varepsilon_r \rangle, \varepsilon_l, \varepsilon_r = 0, 0.02, 0.04, 0.06, ..., 0.4$. All simulations start with 625 randomly generated opinions. *Figure 12a* shows the number *n* of surviving opinions with $n \leq 10$, while *Figure 12b* is a detail of *Figure 12a* with $n \leq 2$ as the upper limit of the *z*–axis. Both figures clarify a bit more our speaking of plurality, polarization, and consensus as three different phases: For very small confidence symmetric or asymmetric intervals lots of opinions survive (*plurality*). But as either ε_l or ε_r or both increase we observe a sharp decline of surviving opinions. Soon one gets to the green–yellow base where for the most part only 2 opinions survive (*polarization*). *Figure 12b*– in principle a magnification of *Figure 12a*– shows that with an further increase in ε_l or ε_r this polarization turns into *consensus*.

Figure 12c shows the *final average* opinion in the whole population for all points $\langle \varepsilon_l, \varepsilon_r \rangle$ with $\varepsilon_l, \varepsilon_r \leq 0.4$, again based on 25 simulation runs with 625 random opinions at the beginning. It is no surprise that for symmetric confidence this final average opinion is about 0.5. With *asymmetric* confidence the *mean* opinion moves into the direction favoured by the asymmetry. This effect is extreme if there is only little confidence in the non favoured direction. *Figure 12d* shows that the effect becomes milder as the confidence in the non favoured direction increases as well.

A decisive step for an explanation of the phenomena stated above is an understanding of the *new type of* splits that occur under *asymmetric* confidence. Since $\varepsilon_l \neq \varepsilon_r$ it may be the case that $x_{i+1}(t) - x_i(t) \leq \varepsilon_r$, while $x_{i+1}(t) - x_i(t) > \varepsilon_l$, where x_i and x_{i+1} are two neighbouring opinions in our ε–profile. In such a situation the opinion $x_{i+1}(t)$ affects the opinion $x_i(t)$ since $x_{i+1}(t)$ is within the ε_r–reach of $x_i(t)$. But at the same $x_{i+1}(t)$ is not affected (any longer) by $x_i(t)$ since $x_i(t)$ is outside the ε_l–reach of $x_{i+1}(t)$. We thus have a *one–sided* split (see definition in Section 3, part D). In the opinion dynamics given by *Figure 13* several one–sided splits occur: Where ever a grey area between two neighbouring opinions has *only* white lines in the direction *top right,* there we have a *one–sided* split. In contrast to that a pattern resulting from both, white lines direction bottom right *and* white lines direction top right, indicates that the two neighbouring opinions are *mutually* within their relevant ε–reach: $x_i(t)$ in the ε_l–reach of $x_{i+1}(t)$, and $x_{i+1}(t)$ in the ε_r–reach of $x_i(t)$.
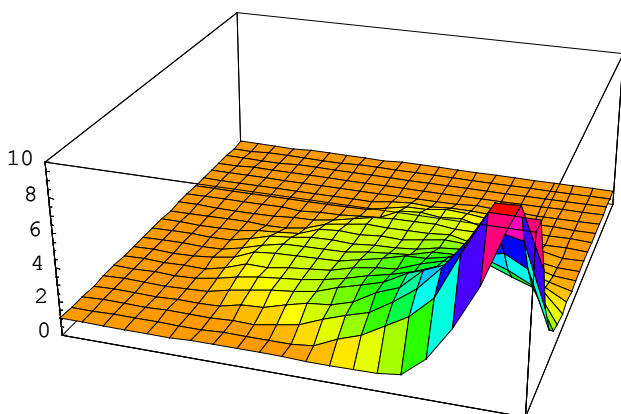

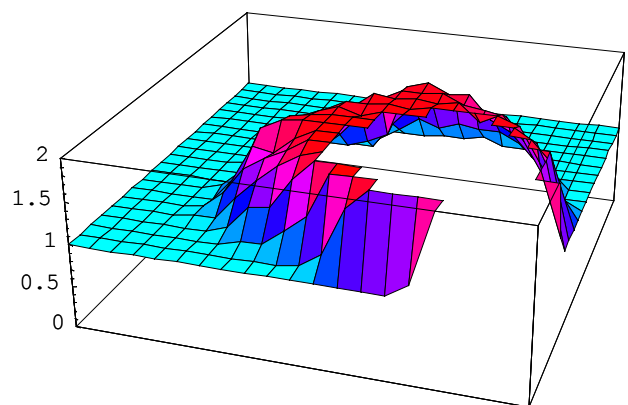*Figure 12a: Number of remaining opinions after stabilisation.*

Figure 12b: Number of remaining opinions after stabilisation, magnification of Figure 12a.
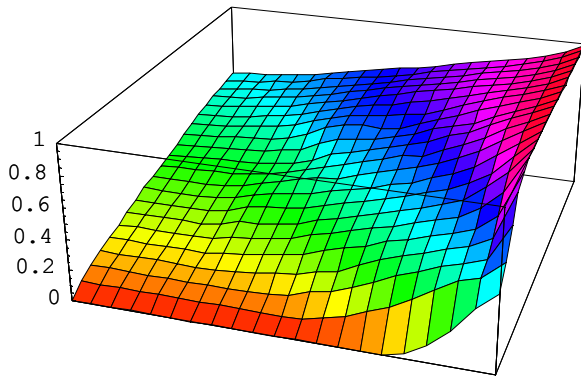
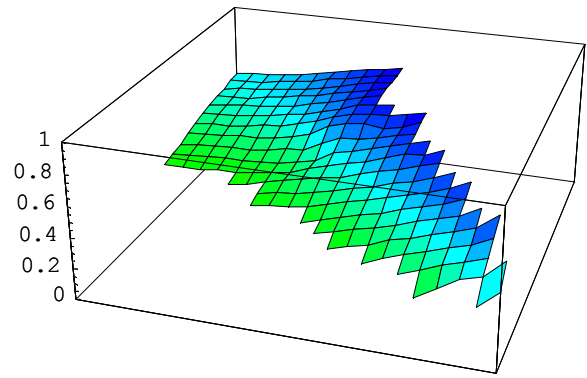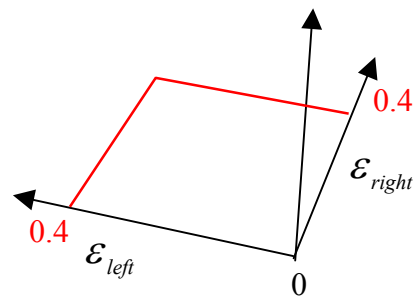*Figure 12c: Final average opinions $\bar{u}$ after stabilisation*

Figure 12d: Detail of Figure 12c: final average opinions $0.3 \leq \bar{u} \leq 0.7$ after stabilisation.

Legend for 12a–12d

Figure 12: Simulation results: Number of surviving opinions and final average.

Even without any split in the profile bare asymmetry of confidence drives the opinion dynamics in the favored direction. But one–sided splits amplify this effect dramatically. How the mechanism works can be seen in *Figure 13*: Opinions right *above* a split become the new extreme left opinions of a remaining ε–sub–profile that continues to converge. At least for a while this sub–profile is no longer influenced by more left opinions. The split off ε–sub–profile *below* the one–sided split converges and moves upwards. It is still under the influence of opinions above the one sided split. By that the whole dynamics is driven to the right. Note that, contrary to two sided–splits one–sided splits can close again.
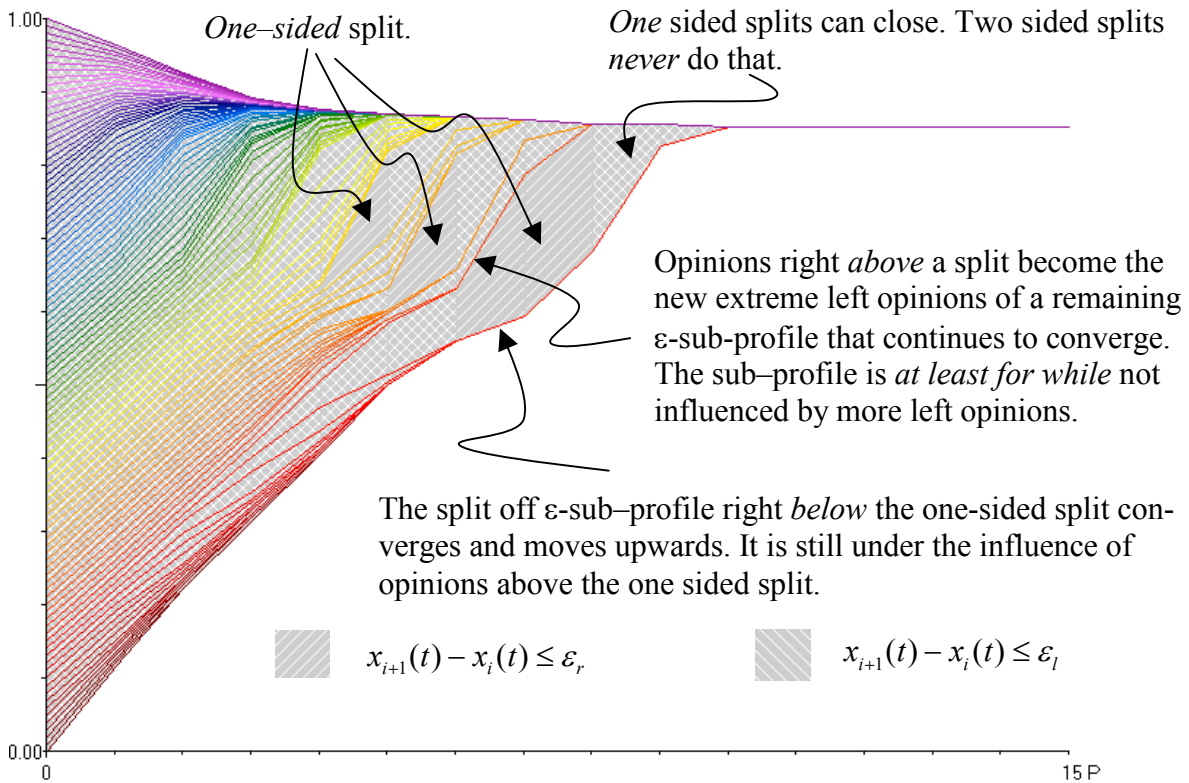
Figure 13: One sided splits (100 opinions, $\varepsilon_l = 0.8, \varepsilon_r = 0.24$).

4.2.2 OPINION DEPENDENT ASYMMETRY

The confidence intervals analysed in 4.2.1 are asymmetric. But the asymmetry is *independent of the opinion* itself. However, it is a quite common phenomenon that that those holding a more left (right) opinion often listen more to other left (right) opinions while being sceptical upon more right (left) views. So it seems natural to model an *opinion dependent asymmetry* in the following way: The more left (right) an opinion is, the more the confidence interval is biased direction left (right). For the special case of a 'centre' opinion, i.e. $x = 0.5$, the confidence interval should be symmetric.
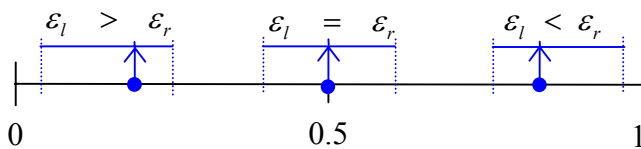

Figure 14: Opinion dependent confidence intervals.

Figure 14 illustrates how –following this idea– a confidence interval of a given *total* (!) size $\bar{\varepsilon}$ is shifted to the left (right) for an opinion to the left (right) of the centre. For the centre opinion the left and the right confidence interval are of equal size.

How to model opinion dependent asymmetries of confidence intervals? One possible strategy is to introduce an opinion dependent bias β_l to the left and a bias β_r to the right such that $\beta_l, \beta_r \geq 0$ and $\beta_l + \beta_r = 1$. Having done that we use β_l and β_r to divide any given confidence interval $\bar{\varepsilon}$ into a left

and a right part. Following our intuition stated above, the values β_r should be generated by a monotonically *in*creasing function f of the opinion x with $x \in [0,1]$. By $1 - f(x)$ we get the monotonically *de*creasing function that we need to generate the values β_l –again following the intuition stated above. Different slopes for the function f would then allow to model the *strength of the bias*.
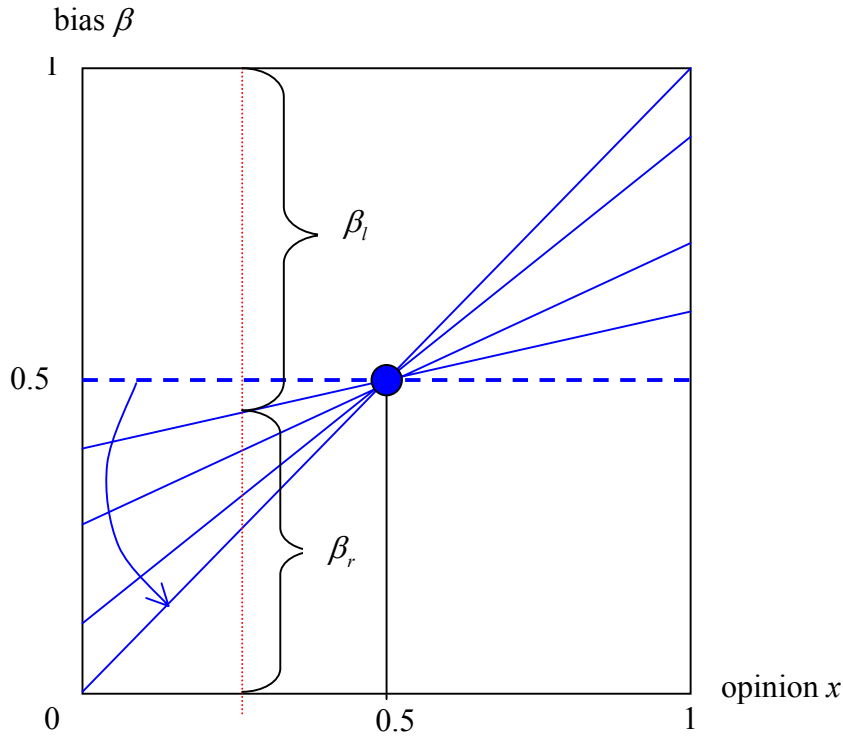

Figure 15: Parameter space for opinion dependent asymmetry.

Figure 15 illustrates an easy way to elaborate in detail this type of asymmetry. The *x*-axis indicates the opinion. The *y*-axis or, respectively, the blue lines indicate the opinion dependent *bias* $\beta_r(x)$ to the right and $\beta_l(x)$ to the left. The blue graphs are generated by rotations around the blue point $\langle 0.5, 0.5 \rangle$, i.e. according to the function $f(x) = mx + \dfrac{1-m}{2}$. It is $\beta_r(x) = mx + \dfrac{1-m}{2}$ and $\beta_l(x) = 1 - \beta_r(x)$. These biases are used to determine how a confidence interval of any given size $\overline{\varepsilon}$ is partitioned into an ε_l and ε_r. We define $\varepsilon_r(u) = \beta_r(u)\overline{\varepsilon}$ and $\varepsilon_l(u) = \beta_l(u)\overline{\varepsilon}$. Then it always holds that $\varepsilon_l(x) + \varepsilon_r(x) = \overline{\varepsilon}$. It is also guaranteed that $\varepsilon_l(0.5) = \varepsilon_r(0.5)$. In this setting it is the slope m in $f(x) = mx + \dfrac{1-m}{2}$ that controls the strength of the bias. For $m = 0$ we do not have any bias. Both parts of the confidence interval have, whatever the opinion might be, the same size. As m increases (by rotating the blue graph anti clockwise around $\langle 0.5, 0.5 \rangle$) the bias becomes stronger and stronger: People with a more left (right) opinion listen less and less to the right (left) side of the opinion space. We will confine ourselves to slopes within the range $0 \le m \le 1$. To give an *example*: For an opinion $x = 0.6$, a total confidence interval of $\overline{\varepsilon} = 0.4$, and $m = 0.5$ we get $\varepsilon_l = 0.18$ and $\varepsilon_r = 0.22$. For any positive m it holds that the more one's opinion is located to the left (right), the more one's confidence interval is shifted to the left (right).

This approach offers an easy way to analyse the effects of opinion dependent asymmetries of confidence intervals (*Figure 16*): For different absolute sizes of confidence intervals we start with symmetry. In each case we let *m* stepwise increase and study the resulting dynamics by means of simulation. The analysed area of the parameter space will be $0 \leq m \leq 1$.
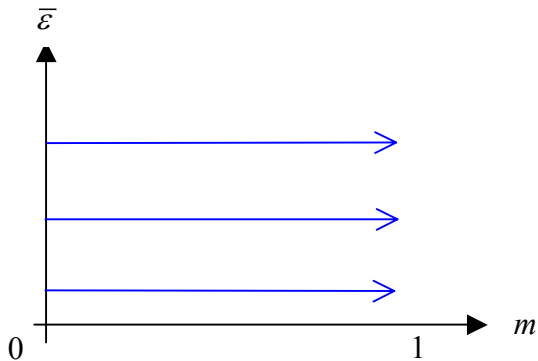

Figure 16: Opinion dependent asymmetries: Analysing the parameter space.

Figure 17 gives an overview. The graphics are of the same *type* as in *Figures 3* and *11*. The *x*-axis represents the opinion space $[0,1]$ divided into 100 intervals. The *z*-axis represents the average (!) relative frequencies of opinions in the 100 opinion intervals of the opinion space after the dynamics has stabilised. *In contrast* to the former figures the *y-axis does not* represent changes in ε; it now represents *changes of the parameter m* which controls the strength of the opinion dependent bias of $\bar{\varepsilon}$. Along the *y*-axis *m* is increasing from 0 to 1 in 26 steps of size 0.04 (while in the former graphics we saw 40 steps of an increasing ε_l or ε_r, respectively). Thus, each graphics represents a walk along one of the blue horizontal lines in *Figure 16*. *Figures 17a* to *c* show the simulation results based on $\bar{\varepsilon} = 0.2, 0.4, 0.6$.
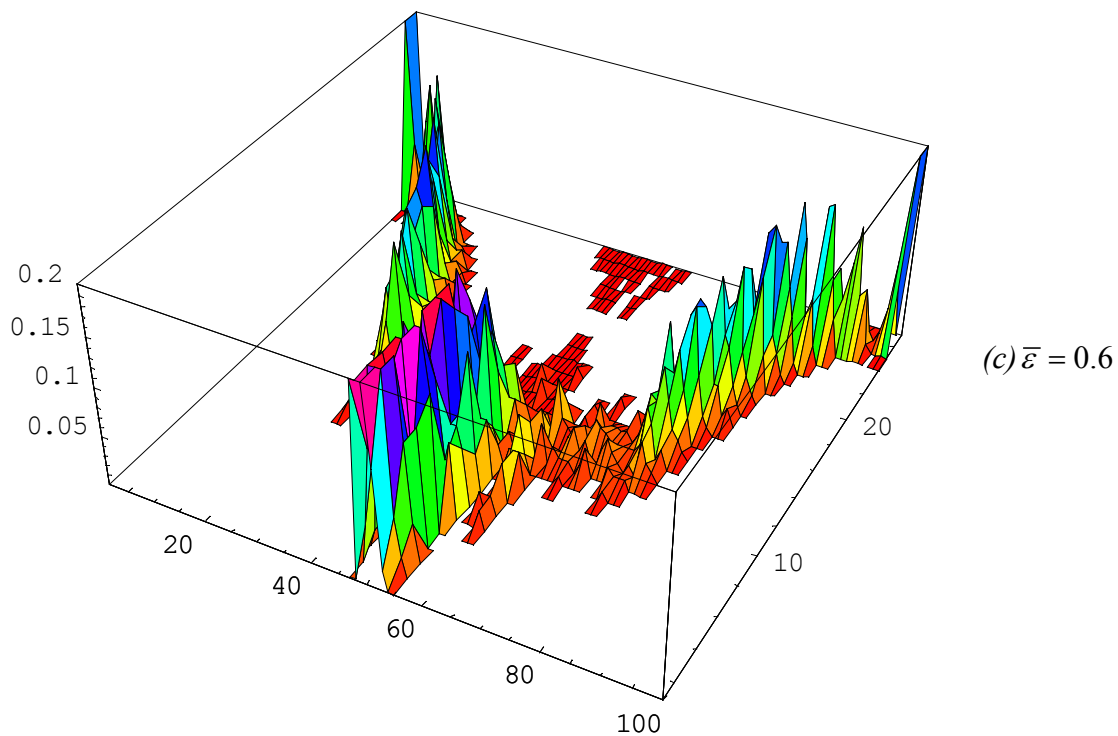

(a) $\bar{\varepsilon} = 0.2$

(b) $\overline{\varepsilon} = 0.4$

(c) $\overline{\varepsilon} = 0.6$
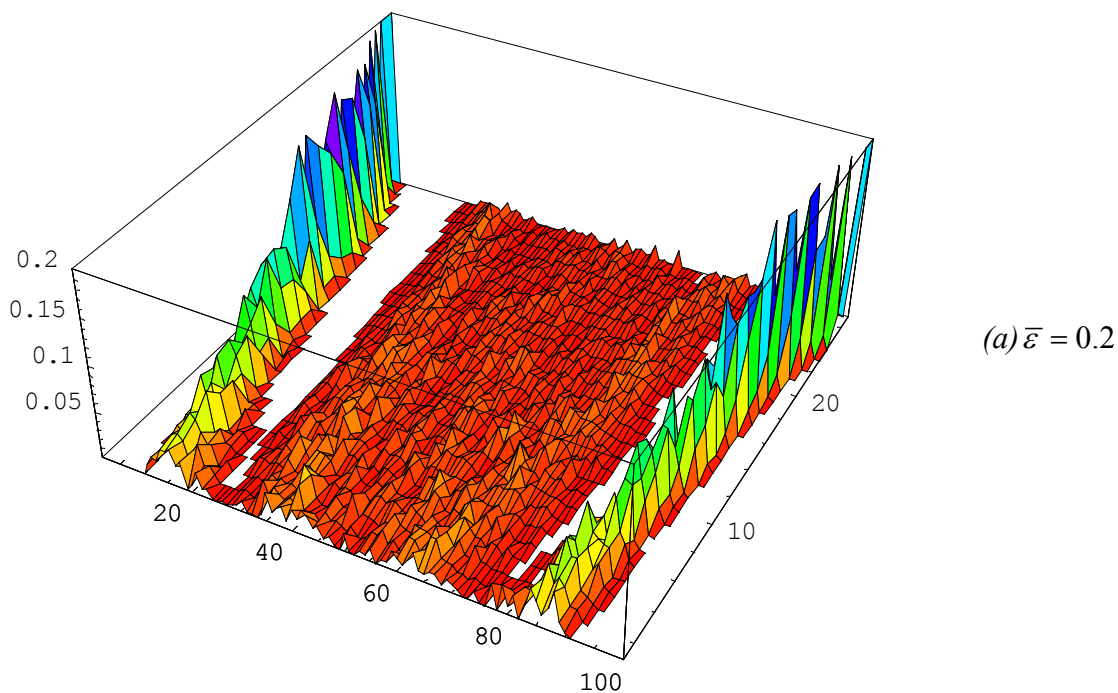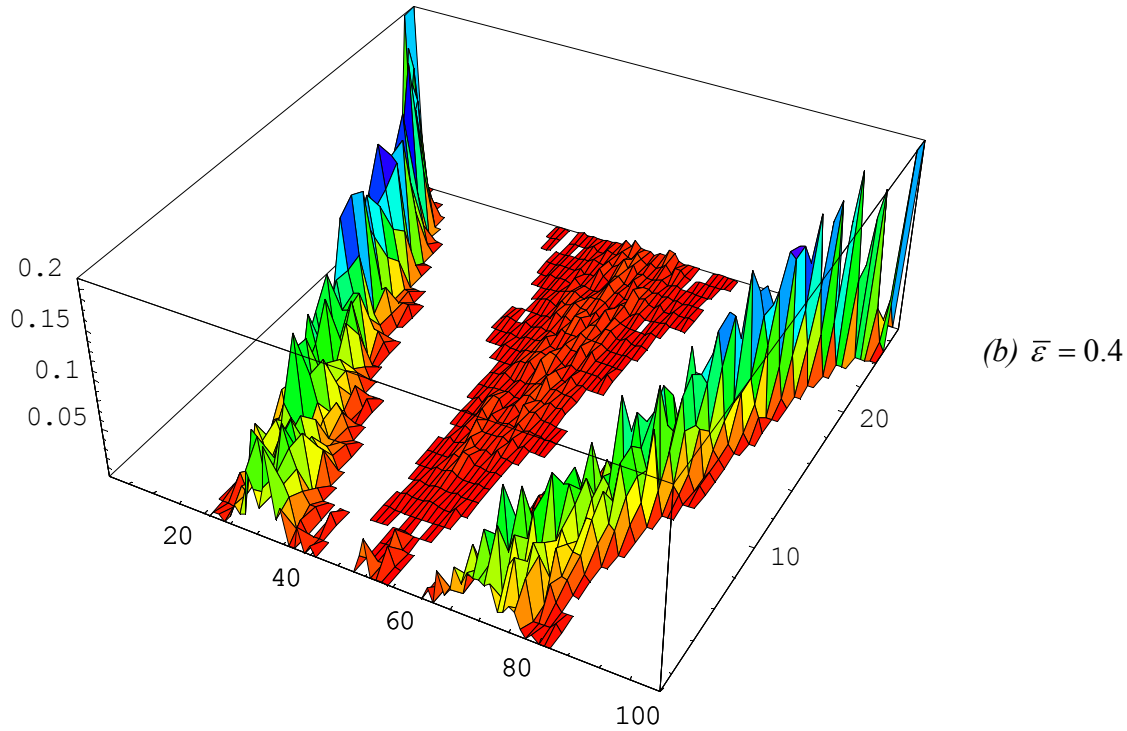
Figure 17: Increasing bias m (26 steps, $m = 0,...,1$) for three different confidence intervals.

The simulations in *Figure 17a* are based on $\overline{\varepsilon} = 0.2$. Along the *y*-axis we start (step 1) with $m = 0$, what implies $\varepsilon_l = \varepsilon_r = 0.1$. Thus, we now get going where we got by step 10 when walking along the diagonal in our first experiments with *symmetric* confidence intervals (*Figure* 2 and *Figure* 3). For a symmetric confidence interval of that size a very mild polarization starts to emerge: The *z*-values show that as an average we find (on the average!) small fractions of remaining opinions in *all* intervals of the slightly shrunk opinion space. At the extremes the frequencies are a bit higher, a

consequence of the one-sided influence which drives the opinions direction centre. But we are far away from a full fledged polarization as we will get for $\varepsilon_l = \varepsilon_r = 0.25$ (step 25 in *Figure* 3). *Figure 17a, firstly,* shows that a sufficiently high opinion dependent bias will produce a blatant polarization even under the condition of a comparatively small confidence interval. That polarization is, *secondly*, more severe in the sense, that the opinion distances of the two major camps are greater than the distances we observe in the symmetric cases. As *m* increases the distance between the two major camps becomes greater and greater. For an *m*=1 the two camps occupy the most extreme positions 0 and 1.

Figure 17b shows the simulation results for an increasing *m* based on $\overline{\varepsilon} = 0.4$. $m = 0$ corresponds step 20 of our walk along the diagonal in the first experiments with *symmetric* confidence intervals (*Figure* 2 and *Figure* 3). For a symmetric confidence interval of that size we got a clear polarization. With an increasing opinion dependent bias the polarization becomes even more dramatic under both perspectives, size of the camps and the distance between them.

The simulation results for a total confidence interval of $\overline{\varepsilon} = 0.6$ are shown in *Figure 17c*. $m = 0$ corresponds step 30 of our walk along the diagonal in the first experiments with *symmetric* confidence intervals (*Figure* 2 and *Figure* 3). An all including consensus is the result and that remains true for a mild opinion dependent bias *m*. But a certain point (about step 11, i.e. $m \approx 0.4$) that consensus breaks down. A sharper and sharper polarisation is the final result.

(a) m = 0

(b) m = 0.25

(c) m = 0.5

(d) m = 0.75

$$x_{i+1}(t) - x_i(t) \leq \varepsilon_{r,i}$$

$$x_{i+1}(t) - x_i(t) \leq \varepsilon_{l,i+1}$$
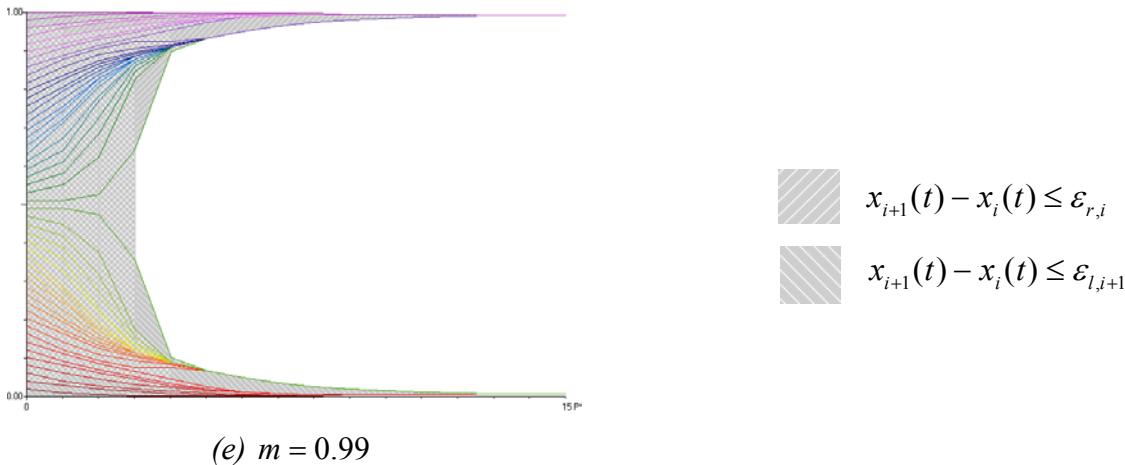
(e) m = 0.99
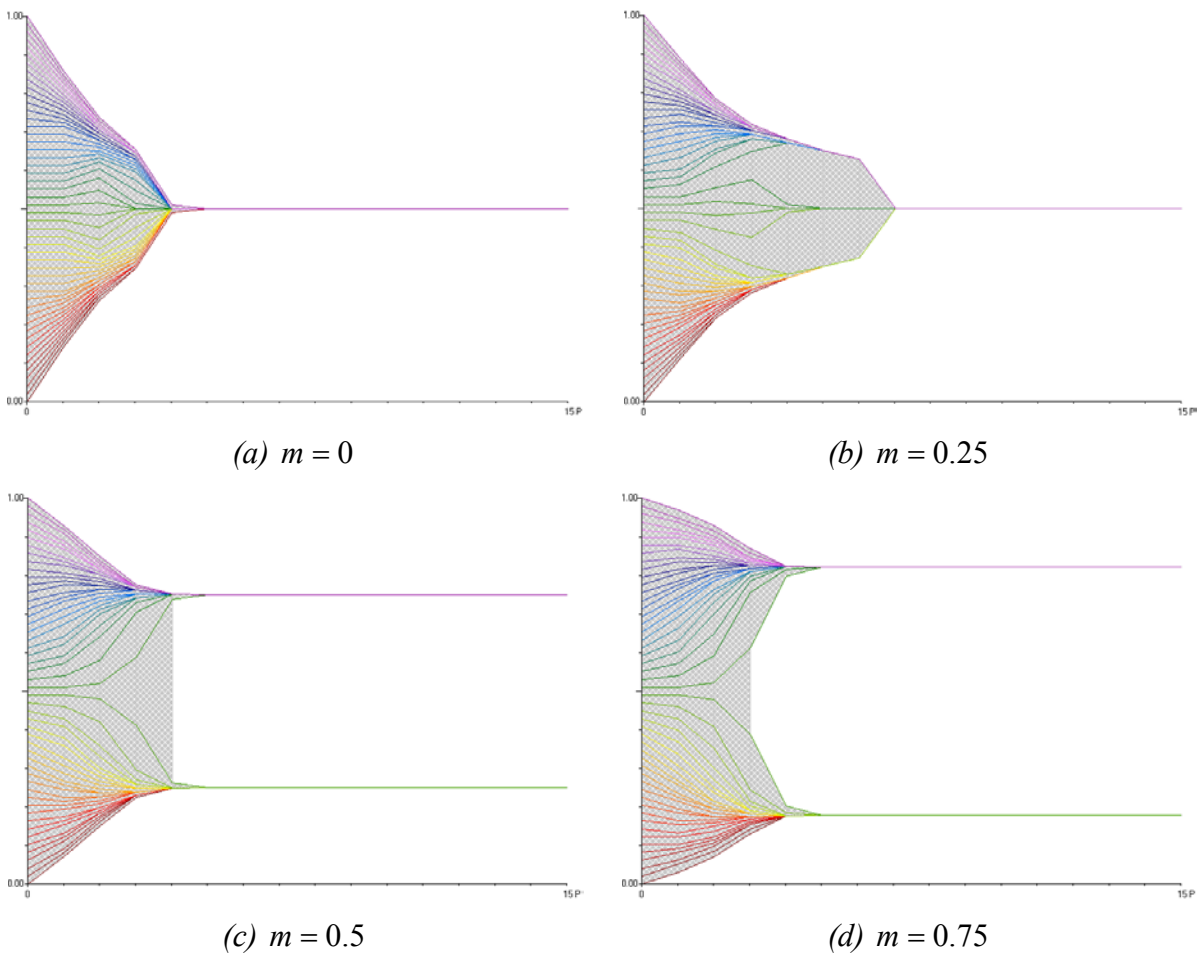
Figure 18: Five different biases for $\varepsilon = 0.6$

For a better understanding of the effects described so far it is helpful to look at *single* runs. *Figure 18* shows a sequence of *single* runs in which the opinion dependent bias gets stronger and stronger. All runs are based on the confidence interval $\overline{\varepsilon} = 0.6$ and an increasing *m*. To keep things simple we use a *regular* profile of 50 opinions. The runs show: As *m* increases it takes longer to get to a consensus. In *Figure 18c* the bias is sufficiently strong to cause a break down of the former consensus. In period 4 the profile splits finally and two polarised opinion camps remain. As *m* increases further, the distance between the two camps becomes greater. The principle cause for all these effects is that with an increasing *m* those at the extremes become *less and less* affected by opinions more in the centre of the opinion space. The decisive point becomes obvious by a comparison of *Figures 18a* and *18e*: Under *symmetric* confidence those at the extremes are under a one–sided influence that drives them direction centre, thereby causing a shrinking of the whole opinion space (*Figure 18a*). With an increasing opinion based bias the drive direction centre disappears or – depending upon the size of the confidence interval– is significantly weakened *at the extremes*. In *Figure 18e* – based on an heavy bias of *m* = 0.99 – the opinions at the extremes stay where they are. At the same time they attract step by step opinions in whose (at least) *one*– sided ε –reach *they* are. Thus, instead of a drive direction centre the opinion based bias generates a *drift to the extremes*.

4.3 MAIN RESULTS

We can *summarise* our results, *firstly*, for the case of symmetric *and* opinion *independent* asymmetries:

- With an increasing symmetric or asymmetric confidence we step from plurality to polarisation and then to consensus.

- Under (a)symmetric confidence polarisation is (a)symmetric as well.

- In the symmetric case the major causes for polarisation are splits in the opinion profiles. They are caused by shrinking and condensing at the extremes, condensing induced by condensing, and condensing induced by density fluctuations in randomly generated start profiles.

- In the case of asymmetric confidence an asymmetric polarisation is caused and amplified especially by one–sided splits. At least temporarily one of the two resulting sub–profiles influences the other one, while not longer being influenced by the other one itself. This favours convergence into the direction favoured by the asymmetric confidence.

- With asymmetric confidence mean and median move into the favoured direction. This effect is extreme if there is only little confidence in the non favoured direction. The effect becomes milder as the confidence in the non favoured direction increases.

As to the effects of *opinion dependent asymmetries* we can, *secondly*, conclude:

- With an increasing opinion depending bias the drift direction centre is significantly weakened at the extremes and –depending upon the size of both, the bias and the confidence interval– may even totally disappear.

- For small confidence intervals which produce plurality in the symmetric case it holds: with an increasing bias at least a moderate polarisation starts earlier.

- With an increasing bias polarization is amplified: The two opinion camps at the extremes become bigger and their final position is more to the extremes.

- With an increasing bias reaching a consensus becomes more and more difficult. Polarisation is the result instead. If consensus is still feasible, it takes more time to get there.

One might ask whether the results depend upon simultaneous updating. The answer is: no. Random serial updating gives extreme opinions a slightly better chance to survive. But none of the results stated above depends crucially on simultaneous updating.

Future directions of research will, firstly, include the analysis of opinion spaces of *higher* dimensions. (For a first analytical result see theorem 4 in the Appendix). Secondly, we will analyse the effects of network structures in which interactions are restricted to *neighbouring* others, i.e. individuals living, for instance, within ones v. Neumann or Moore neighbourhood. First simulations show that this type of locality matters dramatically: If the neighbourhoods in which the agents interact are fairly small (though overlapping!), then the phase in between plurality and consensus, i.e. polarization, disappears. At least under bounded confidence locality of interactions may prevent societies from sharp polarisation.

APPENDIX: Theorems and Hints

A. The classical model (GM)

The model $x(t+1)=Ax(t)$, $t\in T$ has the *consensus property* if for every $x(0)\in\mathbb{R}^n$ there exists a $c\in\mathbb{R}$ such that $\lim_{t\to\infty}x_i(t)=c$ for all $i\in\{1,\cdots,n\}$, all $t\in T$.

Theorem 1

- If for any two $i,j\in\{1,\cdots,n\}$ there exists some $k\in\{1,\cdots,n\}$ such that $a_{ik}>0$ and $a_{jk}>0$ then the consensus property holds.

- The consensus property holds if and only if there exists some $t_0\in T$ such that the matrix power A^{t_0} contains at least one strictly positive column.

For the first part of Theorem 1 see (De Groot 1974), for the second part see (Berger 1981).

Theorem 2 Let A be in Gantmacher normal form with diagonal blocks A_i, $1\leq i\leq s$, $g\leq s$.

- $\lim_{t\to\infty}x(t)$ exists for every $x(0)\in\mathbb{R}^n$ if and only if the A_i are all primitive for $1\leq i\leq g$ or, equivalently, if 1 is the only root of A of modulus 1.

- The consensus property holds if an only if $g=1$ and A_1 is primitive or, equivalently, if 1 is the only root of A of modulus 1 and 1 is a simple root.

For a proof see (Gantmacher 1959).

B. The Friedkin-Johnsen model (FJ)

From the model (FJ), i.e.,

$$x(t+1)=Gx(0)+(I-G)Ax(t) \text{ for } t\in T$$

one obtains by induction

$$x(t)=V(t)x(0) \text{ for } t\in T, \text{ where}$$

$$V(t)=M^t+\left(\sum_{k=0}^{t-1}M^k\right)G \text{ with } M=(I-G)A.$$

For $G=0$, model (FJ) specializes to (CM) and $M=A$, $V(t)=A^t$.

Theorem 3

- Let $G\neq 0$ and suppose A is an irreducible matrix. For every $x(0)\in\mathbb{R}^n$ there exists $x(\infty)=\lim_{x\to\infty}x(t)$ and one has the formula $x(\infty)=(I-M)^{-1}Gx(0)$.

- Consensus $x(\infty)=(c,\cdots,c)$ holds if and only if $x_i(0)=c$ for all i with $g_i>0$.

See (Friedkin and Johnsen 1990, Appendix).

C. Time-variant model (TV)

Consider *multidimensional opinions,* i.e., the opinion of agent i at time t is given by $x^i(t)\in\mathbb{R}^m$, where $m\geq 1$ is the number of opinions considered. The (TV) model then reads

$$x^i(t+1)=\sum_{j=1}^n a_{ij}(t)x^j(t) \text{ for } 1\leq i\leq n, t\in T.$$

This shows that $x^i(t+1)$ is a convex combination of $x^1(t),\cdots,x^n(t)$ in $\mathbb{R}^m$. For points $z^1,\cdots,z^p\in\mathbb{R}^m$ the set of all *convex combinations* of these points is denoted by $\text{conv}\{z^1,\cdots,z^p\}$. For a subset $M\subset\mathbb{R}^m$ the *diameter of M* is

$$d(M)=\sup\{\|a-a'\|\,|\,a,a'\in M\}, \text{ where } \|\cdot\| \text{ is the Euclidean norm on } \mathbb{R}^m \text{ (but it could be any norm on } \mathbb{R}^m\text{).}$$

Lemma For $x^1,\cdots,x^n\in\mathbb{R}^m$ and $y^i=\sum_{k=1}^n a_{ik}x^k$, $1\leq i\leq n$ the following estimate holds

$$d\left(\text{conv}\{y^1,\cdots,y^n\}\right)\leq\left(1-\min_{1\leq i,j\leq n}\sum_{k=1}^n\min\{a_{ik},a_{jk}\}\right)d\left(\{x^1,\cdots,x^n\}\right).$$

This estimate shows that the diameter of the set spanned by the agents' opinions shrinks by a certain factor from one period to the next one.

This Lemma is the crucial step in proving the following theorem.

Let for $s,t\in T$ with $s<t$ the matrix $B(t,s)=(b_{ij}(t,s))$ denote the matrix product $A(t-1)A(t-2)\cdots A(s)$ which models the *accumulated weights* between periods s and t.

Theorem 4 Suppose there exist a sequence $0 = t_0 < t_1 < t_2 < \cdots$ in T and a sequence $\delta_1, \delta_2, \cdots$ in $[0,1]$ with $\sum_{m=1}^{\infty} \delta_m = \infty$ such that

$$\sum_{k=1}^{n} \min \left\{ b_{ik}(t_m, t_{m-1}), \ b_{jk}(t_m, t_{m-1}) \right\} \geq \delta_m$$

for all $m \geq 1$, all $1 \leq i, j \leq n$.

Then for any $x^1(0), \cdots, x^n(0)$ in $\mathbb{R}^m$ there exists a consensus

$x^* \in \operatorname{conv} \left\{ x^1(0), \cdots, x^n(0) \right\}$, i.e., $\lim_{t \to \infty} x^i(t) = x^*$ for all $1 \leq i \leq n$.

Furthermore, for $y^1(0), \cdots, y^n(0)$ in $\mathbb{R}^m$ with corresponding consensus y^* one has the sensitivity property that

$$\left\| x^* - y^* \right\| \leq \max_{1 \leq i, j \leq n} \left\| x^i(0) - y^j(0) \right\|.$$

For Theorem 4 in the special case of $m = 1$ see (Chatterjee 1975), (Chatterjee and Seneta 1977). For Theorem 4 in case of more general average procedures see (Krause 2000). Theorem 4 as above for multidimensional opinions is new.

D. Opinion dynamics with bounded confidence (BC)

For two agents $i, k \in \{1, \cdots, n\}$ and $s < t$, s and t in T, a sequence $(i_0, i_1, \cdots, i_{t-s})$ of agents is called a *confidence chain from* i *to* k for (s, t) if it holds that $i_0 = i$, $i_{t-s} = k$ and $i_j \in I(i_{j-1}, x(t-j))(x(0) \text{ given})$ for $j = 1, 2, \cdots, t-s$.

Theorem 5 Consensus will be approached (in finite time) for a given initial profile, provided there exist $h \geq 1$ such that for all m the following property holds: For any two agents i and j there exists a third one k such that a confidence chain exists from i to k and from j to k for $((m-1)h, mh)$.

The proof of Theorem 5 employs Theorem 4.

Theorem 6 For any given initial profile there exist $t^* \in T$, natural numbers $1 < n_1 < \cdots < n_k < n$ and nonnegative numbers c_j for $0 \leq j \leq k$ such that for every j $x_i(t) = c_j$ for all $n_j \leq i < n_{j+1} (n_0 = 1, n_{k+1} = n)$ for all $t \geq t^*$.

Theorem 6 can be derived by applying Theorem 5 to certain subgroups.

In a different manner, Theorem 6 was first proved by J.C. Dittmer (Dittmer 2000), (Dittmer 2001, p. 4618). There, instead of Theorem 5, the following result is used (Dittmer 2001, p. 4617):

If the opinion profile is an ε-chain for every time point then a consensus will be reached in finite time.

(This can be obtained also as a special consequence of Theorem 5.)

REFERENCES

ABELSON, R P (1964), "Mathematical models of the distribution of attitudes under controversy". In Frederiksen N and Gulliksen H (Eds.), *Contributions to Mathematical Psychology,* New York, NY: Holt, Rinehart, and Winston.

BECKMANN T (1997) Starke und schwache Ergodizität in nichtlinearen Konsensmodellen. *Diploma thesis Universität Bremen*.

BERGER R L (1981) A necessary and sufficient condition for reaching a consensus using De Groot's method. *J. Amer. Statist. Assoc*. 76. pp. 415 – 419.

CHATTERJEE S (1975) Reaching a consensus: Some limit theorems. *Proc. Int. Statist. Inst*. pp. 159 – 164.

CHATTERJEE S and Seneta E (1977) Toward consensus: some convergence theorems on repeated averaging. *J. Appl. Prob*. 14. pp. 89 – 97.

COHEN J, Hajnal J , and Newman C M (1986) Approaching consensus can be delicate when positions harden. *Stochastic Proc. and Appl*. 22. pp. 315 – 322.

DEFFUANT G, Neau D, Amblard F, and Weisbuch G (2000) Mixing beliefs among interacting agents. *Advances in Complex Systems* 3. pp. 87 – 98.

DE GROOT M H (1974) Reaching a consensus. J. *Amer. Statist. Assoc*. 69. pp. 118 – 121.

DITTMER J C (2000)   Diskrete nichtlineare Modelle der Konsensbildung. *Diploma thesis Universität Bremen*.

DITTMER J C (2001) Consensus formation under bounded confidence. *Nonlinear Analysis* 47. pp. 4615 – 4621.

FRIEDKIN N E and Johnsen E C (1990) Social influence and opinions. *J. Math. Soc*. 15. pp. 193 – 206.

FRIEDKIN N E and Johnsen E C (1999) Social influence networks and opinion change. *Advances in Group Processes* 16. pp. 1 – 29.

FRENCH J R P (1956) A formal theory of social power. *Psychological Review* 63. pp. 181 –194.

FUJIMOTO T (1999) A simple model of consensus formation. *Okayama Economic Review* 31. pp. 95 – 100.

GANTMACHER F R (1959) *Applications of the Theory of Matrices*. Interscience, New York.

HARARY F (1959) "A criterion for unanimity in French's theory of social power". In Cartwright D (Ed.), *Studies in Social Power*. Institute for Social Research, Ann Arbor.

HEGSELMANN R and Flache A (1998) Understanding complex social dynamics – a plea for cellular automata based modelling. *Journal of Artificial Societies and Social Simulation,* vol. 1 no. 3. <http://www.soc.surrey.ac.uk/JASSS/1/3/1.html>

HEGSELMANN R, Flache A and Möller V, "Cellular automata models of solidarity and opinion formation: sensitivity analysis". In Suleiman R, Troitzsch K G, Gilbert N and Müller U (Eds.), *Social Science Microsimulation: Tools for Modeling, Parameter Optimization, and Sensitivity Analysis*, Heidelberg: Springer. pp. 151 – 178.

HOYLST J A, Kacperski K and Schweitzer F (2001) Social impact models of opinion dynamics. *Ann. Rev. Comp. Physics* IX. pp.253 – 273.

KRAUSE U (1997), „Soziale Dynamiken mit vielen Interakteuren. Eine Problemskizze". In Krause U and Stöckler M (Eds.) *Modellierung und Simulation von Dynamiken mit vielen interagierenden Akteuren,* Universität Bremen. pp. 37 – 51.

KRAUSE U (2000), "A discrete nonlinear and non—autonomous model of consensus formation". In Elaydi S, Ladas G, Popenda J and Rakowski J (Eds.), *Communications in Difference Equations*, Amsterdam: Gordon and Breach Publ. pp. 227 – 236.

LATANÉ B and Nowak A (1997), "Self-organizing social systems, necessary and sufficient conditions for the emergence of clustering, consolidation, and continuing diversity. In Barnet G and Boster F (Eds), *Progress in Communication Science: Persuasion.* Norwood, NJ: Ablex. pp. 43 – 74.

LEHRER K (1975) Social consensus and rational agnoiology. *Synthese* 31. pp. 141 – 160.

LEHRER K and Wagner C G (1981) *Rational Consensus in Science and Society.* Dordrecht: D. Reidel Publ. Co.

STAUFFER D (2001) Monte Carlo simulations of Sznajd models. *Journal of Artificial Societies and Social Simulation*, vol. 5, no. 1. <http://www.soc.surrey.ac.uk/JASSS/5/1/4.html>

STOCKER R, Green D G and Newth D (2001) Consensus and cohesion in simulated social networks. *Journal of Artificial Societies and Social Simulation*, vol. 4, no. 4. <http://www.soc.surrey.ac.uk/JASSS/4/4/5.html>

WAGNER C G (1978) Consensus through respect: a model of rational group decision-making. *Philosophical Studies* 34. pp. 335 – 349.

WEISBUCH G, Deffuant G, Amblard F and Nadal J P (2001), Interacting agents and continuous opinion dynamics. <http://arXiv.org/pdf/cond-mat/0111494>