



**HAL**  
open science

# The Leviathan model without gossips and vanity: the richness of influence based on perceived hierarchy

Sylvie Huet, G. Deffuant

► **To cite this version:**

Sylvie Huet, G. Deffuant. The Leviathan model without gossips and vanity: the richness of influence based on perceived hierarchy. Social Simulation Conference 2015, Sep 2015, Groningen, Netherlands. 13 p. hal-01229403

**HAL Id: hal-01229403**

**<https://hal.science/hal-01229403>**

Submitted on 16 Nov 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## The Leviathan model without gossips and vanity: the richness of influence based on perceived hierarchy

SYLVIE HUET, GUILLAUME DEFFUANT

Irstea, UR LISC Laboratoire d'ingénierie des systèmes complexes,  
9 avenue Blaise Pascal, F-63178 Aubière, France

`sylvie.huet@irstea.fr, guillaume.deffuant@irstea.fr`

**Abstract.** This paper studies a simplified version of the Leviathan model, without gossip and vanity processes. During dyadic meetings, the agents change their opinion about themselves and about their interlocutor. Speakers highly valued by their listeners are more influential. Only due to the asymmetry of the influence function, the model produces several patterns depending on the parameter values, some of which not yet identified in the first Leviathan model study. In particular, in some cases the leaders have a high opinion of other leaders whereas in other cases they have a low one.

The recently proposed Leviathan model [1] considers a population of agents, each characterized by its opinion (a continuous number between -1 and +1) about each of the agents (including itself), and dynamics of these opinions through processes of opinion propagation and vanity, taking place during random dyadic encounters. This model is very rich in terms of emerging behaviors, and [1] focuses only on a limited set of parameter values. To go further, we propose here a more complete study of one of the model's basic processes which is the direct opinion propagation (without gossiping). Whereas in many opinion models, the influence increases with the similarity between the agents (homophily [2] [3-5] [6]), in the Leviathan model, the influence increases with the superiority of the speaker as perceived by the listener (difference between listener's opinion about the speaker and listener's self-opinion). This modeling choice can be grounded in the research in social psychology considering for example [7] or [8] who has experimentally shown that people feeling high in power tends to discount other's advice on the contrary those feeling low.

Indeed, complementary to the classical approaches based on homophily, a large body of work considers the influence in terms of credibility of the source [9-11]. The influence function of the Leviathan model is a sigmoid logistic model. This model is more classically used as a probabilistic version of a threshold model [12] [13].

Surprisingly, and even in the absence of vanity and gossip, the model generates a rich variety of patterns and this paper aims at identifying and characterizing them. It is organised as follows. We start by a description of the dynamics. Then, we describe the observed patterns and offer some generalised definitions of their characteristics as well as how they appear during trajectories of the model. We especially point out the link between the influence function parameters, the level of agreement of agents about

their respective order and the average sign of the opinion of the population. Finally, we discuss our results and identify some complementary studies to carry out.

## 1 Opinion propagation depending on perceived hierarchy

We consider a set of  $N$  agents, each agent  $i$  is characterised by her list of opinions about the other agents and about herself:  $(a_{i,j})_{1 \leq i,j \leq N}$ . We assume  $a_{i,j}$  lies between -1 and +1, or it is undefined (equal to nil) if the agent  $i$  never met  $j$  and nobody has talked to  $i$  about  $j$  yet. At initialisation, we suppose that the agents never met, therefore all their opinions are undefined. When opinions change, we always keep them between -1 and +1, by truncating them to -1 if their value is below -1 after the interaction, or to +1 if their value is above +1. The individuals interact in uniformly and randomly drawn pairs  $(i, j)$  and at each encounter they try to influence each other on their respective values. We define one iteration, i.e. one time step  $t \rightarrow t + 1$ , as  $N/2$  random pair interactions (each individual interacts 1 time on average during one iteration). To be more precise, one iteration involves the following steps:

Repeat  $N/2$  times:  
 Choose randomly a couple  $(i,j)$   
 Save the opinions which are going to change in temporary variables to ensure the update during the  $i$  and  $j$  meeting is synchronous  
 Influence( $i,j$ )  
 Influence( $j,i$ )

The influence( $i,j$ ) process is:

**Influence( $i,j$ )**  
 if  $a_{ii} = nil$ ,  $a_{ii} \leftarrow 0$   
 if  $a_{jj} = nil$ ,  $a_{jj} \leftarrow 0$   
 $a_{ii} \leftarrow a_{ii} + \rho p_{ij} (a_{jj} - a_{ii} + \text{Random}(-\delta, +\delta))$   
 $a_{jj} \leftarrow a_{jj} + \rho p_{ij} (a_{ii} - a_{jj} + \text{Random}(-\delta, +\delta)) + v_{ij}$

We recognise the equations of opinion influence (or propagation) in which opinions attract each other, but with two differences. The first difference is that the strength of the propagation of opinion is ruled by a parameter  $\rho$  multiplied by a function  $p_{i,j}$ . Function  $p_{i,j}$  implements the hypothesis that the more  $i$  perceives  $j$  as superior to itself, then the more  $j$  is influential on  $i$ . It is a logistic function (with parameter  $\sigma$ ) of the difference between the opinion of  $i$  about  $j$  ( $a_{i,j}$ ) and the opinion  $i$  about herself ( $a_{i,i}$ ):

$$p_{ij} = \frac{1}{1 + \exp\left(-\left(\frac{a_{ij} - a_{ii}}{\sigma}\right)\right)}$$

$p_{i,j}$  tends to 1 when  $a_{i,j} - a_{i,i}$  is close to 2 ( $i$  values  $j$  higher than herself), and tends to 0 when it is close to -2 ( $i$  values  $j$  lower than herself). When  $\sigma$  is small,  $p_{ij}$  rapidly changes from 0 to 1. When  $\sigma$  is large, this change is progressive.

If  $a_{i,j} = \text{nil}$  ( $j$  is unknown to  $i$ ), we assume that  $i$  has no opinion because he has not met or heard about  $j$ . At the first meeting, we suppose that the a priori about  $j$  is neutral and we set  $a_{i,j} \leftarrow 0$ . Let us also observe that, at the initialisation, an agent has no opinion about herself thus we also set  $a_{i,i} \leftarrow 0$  at the first discussion.

The second difference with simple attraction dynamics is the introduction of variable  $\delta$ . This variable models the idea that an agent  $i$  has no direct access to the opinions of another one ( $j$ ) and can misunderstand it. To take into account this difficulty, we consider the perception of the agent  $i$  as the value  $a_{i,j}$  more or less a uniform noise drawn between  $-\delta$  and  $+\delta$  ( $\delta$  is a model parameter). This random addition then corresponds to a systematic error the agents make about the others' opinions. It can be seen as a noise that distorts the perception that  $i$  has about  $j$ 's opinions. The parameter  $\delta$  rules the amplitude of this noise.

Note that the update is synchronous: every opinion changes occurring during a meeting are computed on the same value of opinions taken at the beginning of a pair meeting.

Finally, the model has 4 parameters:

- $N$ , the number of individuals;
- $\sigma$ , the reverse of the sigmoidal slope of the propagation coefficient;
- $\delta$ , maximum intensity of the noise when someone is alluded to;
- $\rho$ , the parameter controlling the intensity of the coefficient of the influence.

## 2 Exploring the emerging patterns in the parameter space

This section begins by presenting the emerging patterns we can observe from the variation of the parameter values. These patterns can be seen as emerging social orders from the agents' interaction. Then, their properties are studied further using a large experimental design showing they are coupled to form specific trajectories of simulation. A last part point out the relation between parameters of the influence function, the level of agreement of agents on how they order each other, and the sign of the average opinion in the population.

### 2.1 The emerging patterns

The following figures presenting the patterns use a representation in which the opinions of each agent is represented as the row of a  $N \times N$  square matrix. The element  $a_{i,j}$  from line  $i$  and column  $j$  is the opinion of agent  $i$  about agent  $j$ . We use colours to code for the opinions: blue for negative and red for positive opinions with light colours meaning that the absolute value is close to 0. In the tests, we consider  $N=20$  agents and  $\rho=0.8$ . We distinguish patterns with the following criteria:

- The presence of leaders: they are agents about which all agents have a very positive (close to +1) opinion; they are recognizable by the strength of the positive view they benefit from. In our matrix representation, they appear as a red column;
- The level of agreement about the agents: agents agree more or less about their peers (the average value of opinions about an agent is called its reputation for sake of simplicity). In our matrix representation, when all the values in a column are close (colors are homogeneous), the level of agreement about the agent of this column is high.
- The distribution of reputations represents a more or less skewed hierarchy.

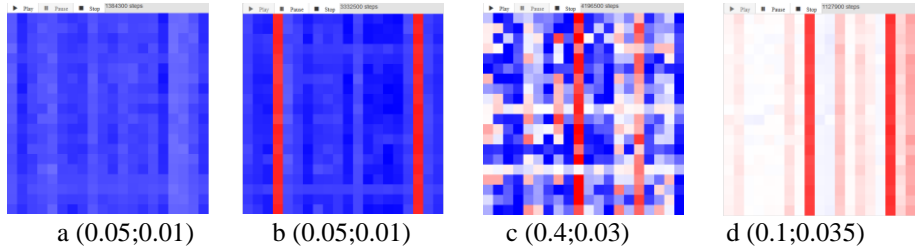
Moreover, we have characterized patterns using the sign of the average opinion of the population as well as the level of agreement on how they perceive each other (do they perceive each other in the same hierarchy), especially when they are leaders. In this latter case, we talk about disagreeing leaders or agreeing leaders.

Using these criteria, we have observed three main groups of patterns: dominance and crisis; disorder; and hierarchy. They are presented in the following.

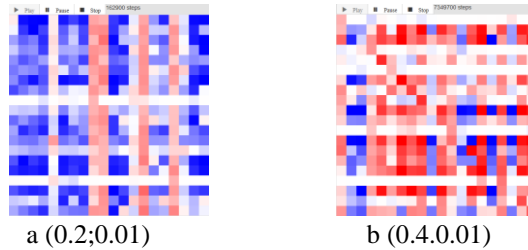
**Dominance and crisis.** Dominance is characterized by the presence of one or two agents with a reputation close to +1, that we call leaders (it can very temporarily increase up to 6 but it does not last) while every other agent has a significantly lower reputation. Fig. 1 shows the various types of dominance we observed plus the crisis in which there is no leader. Crisis (fig. 1a) never appears alone; it alternates with the dominance (fig. 1b). Moreover, various types of dominance can be observed, from negative (fig. 1 b and c) to positive (fig. 1d) as well as some in which leaders agree their respective positions (fig. 1b and d) or disagree (fig. 1c). Fig. 1c illustrates the disagreement between leaders: one can see the most positive leader (more dark red squares) in the center has a neutral opinion (white squares) about the second leader on the right; the second leader has also an opinion about the first leader which is close to 0 but it has a positive self-opinion.

**Disorder.** This pattern is characterized by a very weak agreement on each other values of opinion. Fig. 2 shows the two types we observe: (a) the negative disorder; and (b) the positive disorder. The positive disorder seems to exhibit a larger number of positive self-opinion (see the red square on the diagonal from the top left to the bottom right).

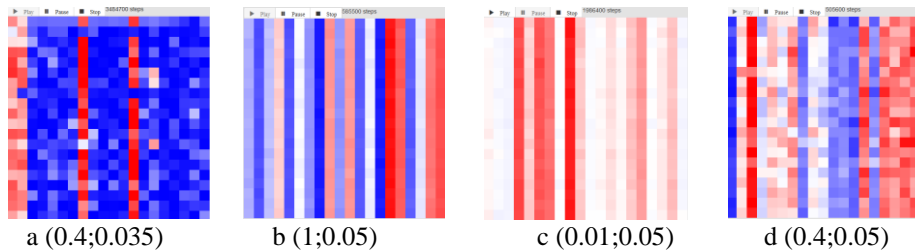
**Hierarchy** in which the distribution of reputations is less bimodal than in dominance patterns and the agreement on reputations is higher than in the disorder. However, as shown in fig. 3, different types can be identified: it can be more negative than positive as in (a) and (b), or more positive than negative as in (c) and (d); leaders can agree as in (b), (c) and (d), or totally disagree and have a negative opinion of each other as in (a). This latter case is not so easily differentiable from dominance with disagreeing leaders.



**Fig. 1. Dominance.** For various  $(\delta;\sigma)$  couples of value given below the graph and from left to right: Crisis in which no agent has any positive opinion; Negative dominance in which leaders agree; Negative dominance with disagreeing leaders about their respective leadership; Positive dominance with agreeing leaders. For low values of  $\delta$  and  $\sigma$ , periods of crisis alternate with periods of dominance, even if less frequent. For larger values of values of  $\delta$  and  $\sigma$ , crisis does not appear anymore and the level of agreement is lower due to the noise.  $N=20$ ,  $\rho=0.8$ .



**Fig. 2. Disorder:** negative on the left; positive on the right. They correspond to  $(\delta;\sigma)$  couples of values given below the graphs.  $N=20$ ,  $\rho=0.8$ .



**Fig. 3. Hierarchy** patterns from left to right: (a) with disagreeing leaders and a majorly negative population; (b) a majorly negative population with agreeing leaders; (c) a positive hierarchy with agreeing leaders; (d) a majorly positive hierarchy with agreeing leaders. They correspond to  $(\delta;\sigma)$  couples of value given below the graphs.  $N=20$ ,  $\rho=0.8$ .

Starting from the various characteristics and types of pattern we observe, we develop a procedure for automatically detecting the patterns and a large experimental design aiming at exploring the parameter space.

## 2.2 Systematic exploration of the parameter space

### 2.2.1 An algorithm detecting the pattern types

We build a pattern type detection algorithm based on a minimum number of indicators:

- the average opinion for the definition of the main sign of the pattern ;
- the bimodal properties of the distribution of opinions; the number of positive reputation to diagnose the absence of leader;
- the average distance between the maximum and the minimum opinion for an agent that we compare to  $2\delta$  in order to diagnose the disorder corresponding to a disagreement level higher than the noise  $\delta$  ;
- the global difference between the maximum and the minimum opinion to qualify the length of the opinion distribution, particularly to distinguish dominance from hierarchy.

The pattern identification algorithm is finally the following:

```

if (average opinion < 0) {
  if (average max-min distance > 2 $\delta$ ) pattern = disorder
  else {
    if (average opinion >= -0.5) {
      if (nb of positive reputation = 0) pattern = crisis
      else pattern = dominance } }
else {
  if (average max-min distance > 2 $\delta$ ) pattern = disorder
  else {
    if ((max opinion - min opinion) > 1) pattern hierarchy
    else dominance } }

```

Then we compute an experimental design running this diagnosis. The model includes 4 parameters. We fix  $N$ , the number of agents to 40, and  $\rho$  ruling the intensity of the opinion propagation coefficient to 0.8. This is in order to make tractable results of our study. We vary the other parameters as follows:

- $\delta$ , the intensity of noise disturbing the evaluation of other's opinions takes two different values: 0.01, 0.03, 0.05, 0.1, 0.2, 0.3, 0.4;
- $\sigma$ , ruling the slope of the logistic function determining the propagation coefficients takes the values 0.01, 0.03, 0.05, 0.07, 0.1, 0.2, 0.3;

For each set of parameter values, we run the model for 200,000,000 iterations (one iteration corresponding to  $N/2$  random pair interactions), and we repeat this for 10 replicas. We measure every 100,000 iterations a group of values allowing us to make conclusions about the properties of our patterns. The measured values over 10 replicas are averaged into indicator that is used in the next subsection to characterise our patterns.

### 2.2.2 The properties of patterns

We consider the complementary following indicators: the type of relations between leaders; the average agreement rate; the number of positive reputations; the number of agents thinking they are the best over the whole population. The type of relation between leaders is defined by a minimum disagreement about the leadership between the leaders. The disagreement occurs when they disagree on their respective order (who is the first, who is the second; in most cases the two think they are both the first). If more than 10% of the leaders disagree, we say this is a disagreement (D) type of leadership; on the contrary it is called (A) agreement. The average agreement rate on order corresponds to the average percentage of dyadic meetings over all the meetings during which the two agents disagree about how to order them. Table 1 presents the results we obtained over our experimental design. We globally identified again what we have observed in figures 1 to 3. Few configurations appear in only one execution of the model while we did 10 replicas of the 49 parameter sets. We decided to omit them in the table for sake of clarity.

**Table 1.** The colours of the patterns are defined by the sign of the average opinion of the population (blue for negative; red for positive). Quantities are averages over all times and replicas and parameter values for which the diagnosis of pattern and the related relation type between leaders has been identified as described in the table.

Pattern	Relation between leaders	Average number of positive reputations	Average agreement rate on order	Average number of agents thinking being the best	Frequency
Crisis	-	0,0	0,62	1,3	1,77%
Dominance	A	2,2	0,63	1,3	10,73%
	D	1,4	0,60	1,3	4,48%
Hierarchy	A	8,6	0,85	1,2	5,37%
	D	1,0	0,84	1,1	0,05%
Disorder	A	2,5	0,72	1,2	2,40%
	D	7,9	0,66	6,6	7,65%
Dominance	A	28,8	0,93	1,0	14,23%
Hierarchy	D	33,0	0,93	1,1	51,78%
Disorder	A	24,2	0,47	23,2	0,03%
	D	20,8	0,55	20,4	1,52%

The table 1 confirms that the pattern dominance has one or two leaders even if a negative Hierarchy with disagreeing leaders also appears with 1.0 leader but it is probably a false diagnosis and it appears rarely (0.05%).

The disorder in its positive form is very specific in terms of agreement on order but also in terms of number of agents thinking that they are the best; the two subcatego-



ries “Agreeing” and “Disagreeing” leaders are not really relevant since there is no leaders; they do not vary from each other and the “A” modality is very rare, probably due to the threshold chosen for the agreement. The disorder in its negative form is softer; it differs less from the other negative patterns.

Crisis always corresponds to disagreeing leaders (even if they are not positive in this case as for the other patterns).

Hierarchy, whatever its signs is characterized by a higher level of agreement rate on order during meetings. It is more characterized by disagreement between its leaders when the average opinion is positive while it is the contrary for the negative average opinion. Overall, and except the case cited earlier, it shows more leaders than the dominance.

The “disagreement” between leaders corresponds to a smaller number of leaders in case of dominance and negative hierarchy. The negative patterns, crisis, dominance and disorder, whatever the type of relation between leaders, show a lower average agreement rate (between 0.6 and 0.7) than the positive ones.

**Table 2.** Trajectories (set of patterns appearing during one execution of the model). Regarding the trajectory sign, we put altogether “positive” and “positive and negative” (\*pAndN) when they represent less than 5% of the runs.

<b>Set of patterns appearing in one trajectory</b>	<b>Trajectory sign</b>	<b>Relation between leaders</b>	<b>% of runs</b>
Disorder	Negative		3,88%
	pAndN* and positive		2,24%
<b>dominance disorder crisis</b>	<b>Negative</b>	<b>Agree and disagree</b>	<b>12,04%</b>
dominance disorder	Negative	Agree and disagree	1,84%
dominance disorder crisis hierarchy	Negative	Agree and disagree	3,06%
dominance crisis hierarchy	Negative	Agree and disagree	0,82%
dominance disorder hierarchy	pAndN and negative	Agree and disagree	2,24%
Dominance	Positive	Agree	3,88%
<b>dominance hierarchy</b>	Negative	Agree and disagree	0,41%
	<b>pAndN</b>	<b>Agree</b>	<b>7,55%</b>
	<b>Positive</b>	<b>Agree</b>	<b>37,14%</b>
dominance hierarchy disorder	Positive	Agree	0,82%
<b>Hierarchy</b>	<b>positive or pAndN</b>	<b>Agree</b>	<b>19,80%</b>

From our results we also confirm that some patterns are only observable coupled to others in trajectories. Table 2 breaks down these observations.

We need 4.9 % of the runs to represent a total parameter set. Let consider the pattern appearing more than 4.9 % and see the average parameter values required to reach them. We distinguish clearly in table 2 three areas: “dominance-disorder-crisis”, “dominance-hierarchy” and hierarchy corresponding to the most frequent trajectories.

### 2.2.3 Trajectories in the parameter space

In order to understand the richness of the model, we locate these three trajectories into the parameter space in table 3 with colour codes.

From the figures 2, we know that the empty bottom left cases of the table 3 corresponds to Disorder (negative for  $\sigma$  0.01 and  $\delta$  0.2; positive for  $\sigma$  0.01 and  $\delta$  0.4). Disorder has not been identified as a main trajectory in the previous phase due to its low frequency of appearance. However, it is probably due to the diagnosis which is very demanding on the level of disorder: that is probable this level changes a lot during a simulation and only a part of measures satisfied the threshold of  $2\delta$  as the minimum average difference of opinion about one agent.

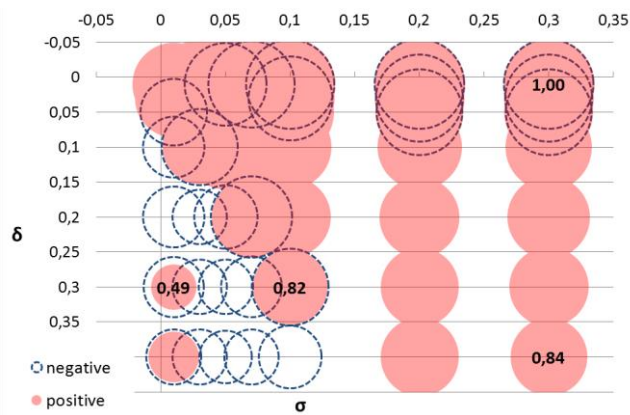
**Table 3.** Main trajectories in the parameter space: light orange is Dominance-Hierarchy positive; blue is Dominance-Disorder-Crisis; brown is Dominance-Hierarchy positive and « positive and negative »; rose is Hierarchy and Dominance-Hierarchy positive; red is Hierarchy. The white bottom left area corresponds to the pattern disorder in its negative shape then in its positive shape when  $\delta$  increases.

$\delta / \sigma$	0,01	0,03	0,05	0,07	0,1	0,2	0,3
0,01	DH+	DH+	DH+/DH+	DH+/DH+	DH+/DH+	DH+/DH+	DH+/DH+
0,03	DH+	DH+	DH+	DH+	DH+/DH+	DH+/DH+	DH+/DH+
0,05	DDC	DH+	DH+	DH+	DH+	DH+/DH+	DH+/DH+
0,1	DDC		DH+	DH+	H / DH+	H / DH+	DH+
0,2		DDC		H	H	H / DH+	H
0,3		DDC	DDC	DDC	H	H	H
0,4			DDC	DDC	DDC	H	H

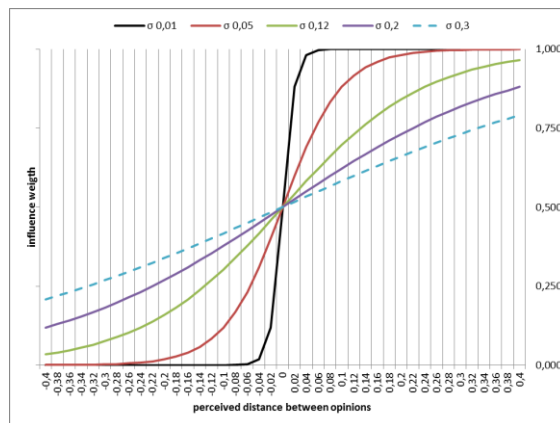
### 2.3 The link between agreement and sign

We observed in table 1 that the average rate of agreement is a good indicator to distinguish negative from positive patterns. Fig. 4 shows this indicator in the parameter space represented on abscissa for the positive (red) versus the negative diagnostics (blue). We observe that a strong change in the value of the indicator between positive and negative diagnosis is located in the space value of parameters as the frontier between the trajectory Dominance-Disorder-Crisis shown in table 3. We observe again the various areas observed in the table 3, especially regarding the sign (always posi-

tive or negative, or positive as well as negative and the level of agreement, weaker for low values of  $\sigma$  if  $\delta$  is large enough. This is then relevant to understand how two agents became disagreeing on their perceived hierarchy.



**Fig. 4.** Value of the average agreement rate order (size of the circle) for various  $\delta$  (ordinate) and  $\sigma$  (abscissa) and for globally negative patterns (blue) or positive patterns (pink)



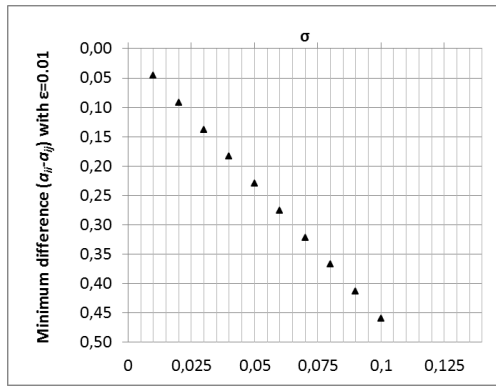
**Fig. 5.** Computation of the influence weight by the influence function for various values of  $\sigma$  and different distance between opinions (the peer one minus the self-opinion) for the listener

That can be understood from the shape the influence function  $p_{ij}$ . Fig. 5 reminds how it is for various values of  $\sigma$ . We observe for low values of  $\sigma$  (0.01, 0.05) that the weight becomes very quickly (ie for a small distance) 0 or 1. Then, when  $\delta$  is close or higher than the distance making the weight close to or equal to 0 or 1, this means that an agent is susceptible to consider her peer from influential to not influential (or almost) in just one meeting and just due to the noise varying from  $-\delta$  to  $+\delta$ . Since an agent became not influential (or almost) for another one, she is not listened by her anymore (or almost). Then, it cannot change its image in the other's mind even if it

gains a very high reputation, higher than one of the agent which is not able to listen to it. In such a situation, these both agents think they are better than the other. We call the agents who are not anymore able to consider the opinion of a peer an agent blind to this peer.

Considering a it is possible to compute the smallest difference  $d$  between the self-opinion an agent and its highest opinion for one of its peer ( $a_{ii}-a_{ij}$ ) s for which the corresponding influence is a value  $\varepsilon$  which is small enough for neglecting the possible opinion changes. Derived from the influence function, the equation is:

$$a_{ii} - a_{ij} = \ln\left(\frac{1}{\varepsilon} - 1\right) \sigma \approx \sigma \ln \frac{1}{\varepsilon}$$



**Fig. 6.** Difference ( $a_{ii}-a_{ij}$ ) on ordinate between peer and the self in terms of opinion which implies that an agent are not influenced by others for various  $\sigma$  (abscissa) and  $\varepsilon=0.01$ .

We observe comparing the fig. 6 to the table 3 that the result is close to define the frontier in the parameter space between the trajectory Dominance-Disorder-Crisis and the other trajectories, if we consider that  $\delta$  is the typical distance between  $a_{ii}$  and the highest  $a_{ij}$ . This gives us some information about the frontier between negative and positive patterns.

For the parameters implying numerous disagreements on perceived hierarchies, the population tends to become negative and to show a dominance pattern. In the worst case, when  $\sigma$  is very very small for large values of  $\delta$ , a significant number of agents consider themselves as superior to all the others: that is the case in the disorder pattern. In case of low disagreement, the observed patterns become positive and have the characteristics of the hierarchy.

### 3 Discussion - conclusion

In this paper, we study a simplified version of the Leviathan model, without the processes of vanity and gossiping, keeping only the direct opinion propagation. Like in the complete Leviathan model, we observe patterns representing social orders but some of them are different as well as some new identified properties. This is very

surprising such a richness of social orders emerge of the simple fact that the influence function is asymmetric. This is for example the case of two different types of leadership depending if leaders agree on their respective position or not. Overall we diagnosed several patterns organized in trajectories: dominance in which one or two leaders have a low opinion about all the other agents who have negative opinion about each other except about the leaders ; it alternates during a simulation with a crisis pattern in which every agent has a negative opinion about all others, including itself. Disorder is a particular pattern in which the agents have different perceived hierarchies. During simulations several regular sequences of patterns appear. The main ones are: dominance-disorder-crisis, dominance-hierarchy, and hierarchy. Moreover these patterns/ sequences exhibit different average sign of opinions: most of them, except crisis can be of both signs but while the hierarchy is more often positive, dominance is more often negative. This work is a base for developing a new insight onto some behaviors of the Leviathan model by comparing the patterns emerging or not from the processes taking place in the model (ie opinion influence, vanity and gossip).

The patterns of the model can be discussed in light of some sociopsychological researches.

The charismatic leaders at the beginning in the model are the individuals who are viewed as positive by every other. Similarly, the charismatic leadership [14] in the socio-psychological literature is defined as the one who benefits from a high esteem of the larger number of the others. For these same authors, the leadership derives its effectiveness from its influence on follower self-concept which is in turn a moderator of the leadership effectiveness. Identically, for [15] and [16], "Leadership is a relational term. It identifies a relationship in which some people are able to persuade others to adopt new values, attitudes and goals ...". We have seen in the model that overall leaders have a greater influence and tend to determine the value of everyone. However the criteria to define a leader, as well as the process of their emergence, are not so clear. Our modelling study has shown that in some cases leaders have the same perceived hierarchies whereas in other cases their perceived hierarchies are different. So what is more important: the popularity and the capacity to be open to others; or on the contrary, the stubbornness and strong resistance to influence? These results can feed the debate in "leadership" research.

The agents tend to have positive opinions on average when the influence function is smooth ( $\sigma$  large), meaning that the differences of influence according to the perceived hierarchy are small, for a not too large noise ( $\delta$ ). For a sharp influence function and a large noise, agents tend to put themselves high in their perceived hierarchy, and for extreme values, the highest. This is due to the asymmetrical property of the influence function and does not require vanity. This echoes the self-enhancement biases identified long ago in social psychology. [17] particularly stresses the "illusory of superiority" : when subjects estimate their relative position on a number of attributes, they typically report that they possess positive characteristics to a higher, and negative characteristics to a lower, degree than average or most others. This bias is often presented as an inner characteristic of the individual. The model suggests that it is socially built from noisy interactions between agents with strong perceived hierarchies. This explanation deserves some proper experiments to define its relevance.

## 4 References

1. Deffuant, G., Carletti, T., Huet, S.: The Leviathan model: Absolute dominance, generalised distrust and other patterns emerging from combining vanity with opinion propagation. *Journal of Artificial Societies and Social Simulation* 16, 23 (2013)
2. Castellano, C., Fortunato, S., Loreto, V.: Statistical physics of social dynamics. *Reviews of Modern Physics* 81, 591-646 (2009)
3. Byrne, D.: An Overview (and Underview) of Research and Theory within the Attraction Paradigm. *Journal of Social and Personal Relationships* 14, 417-431 (1997)
4. Mark, N.P.: Culture and Competition: Homophily and Distancing Explanations for Cultural Niches. *American Sociological Review* 68, 319-345 (2003)
5. Takács, K., Flache, A., Mäs, M.: Is There Negative Social Influence? Disentangling Effects of Dissimilarity and Disliking on Opinion Shifts (June 3, 2014). . Available at SSRN: <http://ssrn.com/abstract=2445649> or <http://dx.doi.org/10.2139/ssrn.2445649> (2014)
6. Axelrod, R.: The dissemination of culture a model with local convergence and global polarization. *Journal of Conflict Resolution* 41, 203-226 (1997)
7. Wood, J.V., Forest, A.L.: Seeking Pleasure and Avoiding Pain in Interpersonal Relationships. In: Alicke, M.D., Sedikides, C. (eds.) *Handbook of Self-Enhancement and Self-Protection*, pp. 258-278. The Guilford Press, New York, London (2011)
8. Tost, L.P., Gino, F., Larrick, R.P.: Power, competitiveness, and advice taking: Why the powerful don't listen. *Organizational behavior and human decision processes* 117, 53-65 (2012)
9. Crano, W.D., Cooper, R.E.: A preliminary investigation of a hyperbolic model of attitude change. Midwest Psychological Association. Michigan State University, Cleveland, Ohio, May 4-6, 1972 (1972)
10. Pornpitakpan, C.: The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology* 34, 243-281 (2004)
11. Tormala, Z.L., Rucker, D.D.: Attitude Certainty: A Review of Past Findings and Emerging Perspectives. *Social and Personality Psychology Compass* 1, 469-492 (2007)
12. Peyton Young, H.: The Diffusion of Innovations in Social Networks. In: Durlauf, L.E.B.a.S.N. (ed.) *The Economy as a Complex Evolving System*, vol. III, pp. 20. Oxford University Press (2003)
13. Huet, S., Edwards, M., Deffuant, G.: Taking into account the variations of social network in the mean-field approximation of the threshold behaviour diffusion model. *Journal of Artificial Societies and Social Simulation* 10, (2007)
14. van Knippenberg, D., van Knippenberg, B., De Cremer, D., Hogg, M.A.: Leadership, self, and identity: A review and research agenda. *The Leadership Quarterly* 15, 825-856 (2004)
15. Uhl-Bien, M.: Relational Leadership Theory: Exploring the social processes of leadership and organizing. *The Leadership Quarterly* 17, 654-676 (2006)
16. Hogg, M.A.: A Social Identity Theory of Leadership. *Personality and Social Psychology Review* 5, 184-200 (2001)
17. Hoorens, V.: Self-enhancement and superiority biases in social comparison. *European Review of Social Psychology* 4, 113-139 (1993)