Collective Reinforcement of First Impression Bias

Guillaume Deffuant and Sylvie Huet

Cemagref-LISC
BP 50085, F 63172 Aubière Cedex, France
Guillaume.Deffuant@Cemagref.fr Sylvie.Huet@Cemagref.fr

Abstract: We propose a simple model of attitude dynamics in which an agent tends to ignore the features which contradict its views. For instance, having received a first very negative feature, the agent may stop to consider any moderately positive feature. We call this phenomenon "first impression bias" (FIB). We consider a population of agents which are all in contact with a media, communicating randomly chosen features of an object. In some cases, we observe on simulations that FIB is significantly more frequent when the agents interact with each other than when they are only in contact with the media. We design an analytical aggregated model of the global agent-based model behaviour which helps to explain the higher number of FIB due to the interactions.

1 Introduction

"One never gets a second chance to make a first impression." This saying illustrates the likely irreversibility of a judgement process. We have all witnessed or experienced that a strong first impression prevents from further considering the features which could change it. Closely related are more general phenomena of information filtering which have been identified long ago, for instance in rumour diffusion (Allport et al. 1947). To simplify, we shall call "First Impression Bias" (FIB) these phenomena of irreversibility or information filtering.

The idea of impression is close to the concept of attitude as defined in Eagly and Chaiken (1998): "Attitude is a psychological tendency that is expressed by evaluating a particular entity with some degree of favour or disfavour". Many researchers assume that attitudes are formed and modified as people gain information. Moreover, attitudes have been postulated to motivate behaviour and to exert selective effects at various stages of information processing (Eagly and Chaiken 1998): information may be filtered by the individuals, i.e. they ignore it. Festinger in his theory of cognitive dissonance (1957) proposes some mechanisms for this selection: people seek out information that supports their attitudes (or decisions) and avoid information that challenges their attitudes (or decisions), in order to minimise their cognitive dissonance. Following this theory, even if they assimilate information which contradicts their global attitude, people are reluctant to talk about it, because they avoid expressing express their dissonance.

Several researches in social modelling include some effect of attitude on information transmission and vice-versa. Rumour diffusion models take into account information relevance or availability (Allport and Postman 1947; Lawson and Butts 2004; Galam 2003). Variants of epidemics models can be interpreted as representing rumour or information diffusion (Huet and Deffuant 2006; Tsimring and Huerta 2003). The bounded confidence models (see Urbig 2003 for a review) implement reception and emissions filters on attitudes or opinions. In the model of innovation diffusion of Deffuant et al. (2005), this opinion dynamics is coupled with an information propagation. However, none of these models focuses specifically on FIB.

We propose a simple agent based model (ABM) of FIB, which abstracts from the cited researches in social psychology. We consider an object defined by several features (for instance price, functionality, impact on health…). Each feature has a utility (negative or positive) and we suppose that agents are told about a randomly chosen object's features by a media, at a given frequency. Thus some agents get the positive features first, others, the negative ones. The agents tend to ignore incongruent features (i.e. features which contradict the agent global impression). Moreover, agents talk to their neighbours about congruent features they retained.

We particularly focus on the following question: do the interactions between agents modify the likelihood of FIB in the population ? With the simple model we consider, the answer is clearly positive. In some cases, the number of agents which show FIB is significantly higher when agents interact. In order to explain this observation, we propose an analytical aggregated model of the agent-based model global behaviour. We follow a general approach of "double-modelling" (Deffuant G., 2004) which aims at providing explanations of the collective effects observed in ABM simulations, using aggregated models of the ABM behaviour. This is the major novelty of this paper, compared to earlier versions (Huet and Deffuant 2006, Deffuant and Huet 2006).

The paper is organised as follows. First we show that, in some cases, the FIB is reinforced by the interactions among agents. Then we design an aggregated model which fits these results, and gives some explanation of the FIB reinforcement due to the interactions. The final section is a brief discussion about the model and these results.

2 The collective reinforcement of FIB in agent-based simulations

First, we describe the model: the considered object, and then the process for a single agent. The analysis of this models allows us to derive the results in the case of a population of isolated agents, and we compare them with the results obtained with interacting agents, with various network topologies.

2.1 A 3-feature neutral object with a highly negative feature

In general, we consider an object described by a set of features F. Each feature j is associated with utility u_j, which is supposed immediately perceived by agents.

In this paper, to simplify the study, we consider an object described by 3 features $F = (1,2,3)$:

- one feature with a highly negative utility: $u_1 = U$ (later on the major feature),
- two features with moderately positive utility: $u_2 = u_3 = u$ (later on the minor features).

We suppose that $U + 2u = 0$, therefore, the object is globally neutral. Our FIB model should yield a different global attitude to the object when the agent receives the negative feature first than when it receives the positive first. We choose the signs of the features to fix the ideas. The behaviour of the model would be equivalent with $U > 0$ and $u < 0$.

2.2 The agent model

An agent i is described by:

- An initial global attitude toward the object g;
- L_i : a subset of F containing the features currently retained by the agent; this list is initially empty.
- $G_i = g + \sum_{j \in L_i} u_j$: the global attitude about the object.

When $G_i . u_j \geq 0$, the feature j is said congruent for agent i, and incongruent otherwise.

The dynamics of filtering are determined by a positive number Θ, the incongruence threshold. We suppose that $|U| > \Theta$ and $u < \Theta$. Being told about feature j, an agent will react as follows :

- If j is congruent → "retain the feature".
- If j is incongruent: if $|u_j| > \Theta$ → "retain the feature", otherwise "ignore the feature" .

Here, "retain the feature" means that j is added to L_i (if L_i does not include j yet), and the global attitude is updated; "ignore the feature" means that j is not added to L_i , and the global attitude is unchanged.

Moreover, an agent diffuses the congruent features in L_i to its neighbours (see 2.4 for more details).

2.3 The FIB for isolated agents receiving features in random order

Consider now a single agent which receives messages about the features of our 3 feature object, in a random order. There are 3 possible cases: the agent is told about the major feature first (symbol: *Uuu*), in second position (symbol: *uUu*), in third position (symbol: *uuU*). A rational agent would yield the same final global attitude in the three cases, equal to his initial attitude g, because the object is neutral. Because $U > \Theta$ and $u < \Theta$, the agent can yield a different final attitude, as shown in table 1, which breaks down all possible cases.

Let us consider a few examples to understand how the results are obtained. Consider column $0 < g < u$, and suppose to fix ideas that for instance $g = 0.5u$:

- If the agent is told about the major feature first (*Uuu* order), the major feature is incongruent (it is negative and g is positive), but because $U > \Theta$, it is retained. The global attitude becomes $g – U = -1.5u$, which is negative (remember that $U = 2u$). Then the minor features become incongruent (the global attitude is negative and they are positive). Because $u < \Theta$, they are not retained, and the global attitude remains $g - U$, a negative value.

- Suppose now that the agent is told about the features in the uuU order. The two minor features are congruent, therefore they are retained. The global attitude becomes $g + 2u = 2.5u$. Then the major feature is received. It is incongruent but retained because $U > \Theta$, therefore the final global attitude is $g + 2u - U = g$, a positive value (in fact the rational result).

	$g < 0$		$0 < g < u$		$u < g < 2u$		$g > 2u$	
	Final att.	Sign	Final att.	Sign	Final att.	Sign	Final att.	Sign
Uuu	$g - U$	-	$g - U$	-	$g - U$	-	g	+
uUu	$g - U$	-	$g - u$	-	g	+	g	+
uuU	$g - U$	-	g	+	g	+	g	+

Table 1: Final attitude and its sign according to the position of the major feature in the FIB model, for the values of g where the order matters, when $U > \Theta$ and $u < \Theta$.

Several features of table 1 are noticeable:

- If $g < 0$, in any order of reception, the final attitude is much more negative than the rational one (which yields g). This is a case of FIB, in which the first impression is the initial global attitude g. However, the sign of the final attitude is the same as the one obtained with a rational agent.
- If $0 < g < u$, if the major feature is received in first or second position, then the final attitude is negative. Therefore in this case, even the second impression can fix the global attitude. In this case the sign of the final global attitude depends on the reception order.
- If $u < g < 2u$, then the global attitude is finally negative only if the major feature is received first. In this case also, the sign of the final global attitude depends on the reception order.
- If $g < 2u$, then the final attitude is the rational one, whatever the reception order.

Consider now a population of N isolated agents with all the same initial attitude g, and which receive the features in a random order. Because the orders (Uuu, uUu, uuU) have the same probability (1/3), the proportion of agents with final negative attitudes are given by table 2. These results are confirmed by our experiments.

	$g < 0$	$0 < g < u$	$u < g < 2u$	$g > 2u$
Theoretical prop. of neg. final att.	1	2/3	1/3	0
Experimental prop. of neg. final att.	1.0 (0.0)	0.66 (0.05)	0.33 (0.04)	0.0 (0.0)

Table 2: Theoretical and experimental proportion of negative final attitudes for isolated agents. The rational result is 1 for $g < 0$, and 0 for $g > 0$ (because the object is neutral). The experiments were carried out with 1000 agents and 1000 replicas. The standard deviations are in parenthesis.

2.4 Interacting agents

Now, we suppose that the agents are connected by a given network: each agent has a set of neighbours to which it is linked. Each agent retains and transmits features following the rules previously defined (cf. 2.2). Again, all agents have the same initial global attitude g. The only difference with the "isolated" case is that agents can transmit features to their neighbours and can receive features from them. At each time step:

- We choose an agent at random, and with a probability f it receives a randomly chosen feature of the object (this represents the communication by the media),
- We choose an agent at random, and one of its neighbours at random, and then the first agent transmits to the other a randomly chosen retained congruent feature.

We focus on the cases where the FIB model provides final attitudes of signs that can be different from the rational result ($0 < g < u$ and $u < g < 2u$). The main experimental results are the following:

- For $u < g < 2u$, the number of final negative attitudes is similar to the one obtained with isolated agents.
- For $0 < g < u$, we observe a significantly higher number of final negative attitudes than the one obtained with isolated agents (see figure 1). The effect decreases when the number of links decreases, but it is still significant for networks with an 2 links per node on average (see figure 2). Here we use random networks with a given average of links per node (Erdos and Renyi 1960), random networks with a

constant number of links per node, or scale free networks (Barabasi and Reka 1999). We note that these networks give very similar results when they have the same average number of links. Moreover, when this average number reaches 6 or 8, the reinforcement of the FIB is very close to the one observed in a totally connected population.

These observations show that the reinforcement of FIB due to the interactions takes place because the major feature changes the attitude when it is in second position as well as when it is in first position. In order to better understand this phenomenon, we design aggregated models of the AB model.

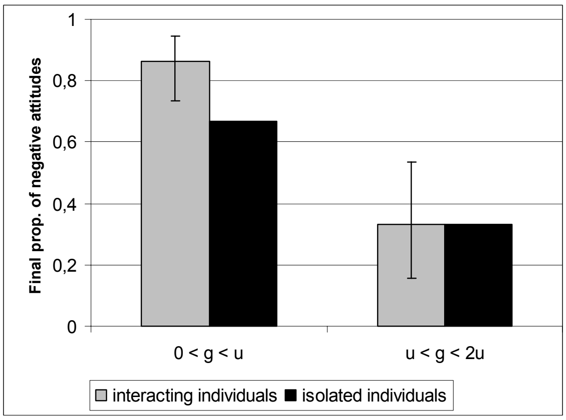

Figure 1: Comparison of the final proportion of negative attitudes between interacting agents (totally connected), and isolated agents. The data are mean standard deviations on 1000 replicas. When $0 < g < u$, there is a significantly higher proportion of final negative attitudes. The FIB is reinforced by the interactions.

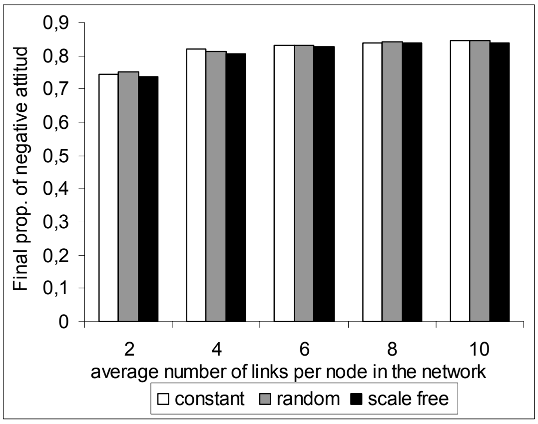

Figure 2: Final proportion of negative attitudes for $0 < g < u$, and various network topologies. The reinforcement of the FIB is significant for all the topologies, although it decreases when the network is less connected. Each result is the mean of 100 replicas on a population of 1000 agents.

3 Analysis of the result using aggregated models

3.1 Groups and transition graphs

The general idea to build the aggregated model is to consider groups of agents in the population and to define the transfer equations which rule the flows of probability densities between the groups. First, we need to determine the groups, and the possible flows between them. The natural approach to define the groups is to consider the list of retained features.

Figure 3 shows the possible states of this list and the possible transitions between them for $0 < g < u$, and figure 4 for $u < g < 2u$. On the right of each figure, the graph is simplified by aggregating the groups from which there is no more bifurcations, or in which all groups have the same sign of attitude. Indeed, our objective is to model the evolution of the number of final negative attitudes, therefore to distinguish between groups in which the attitude is of the same sign is not necessary.

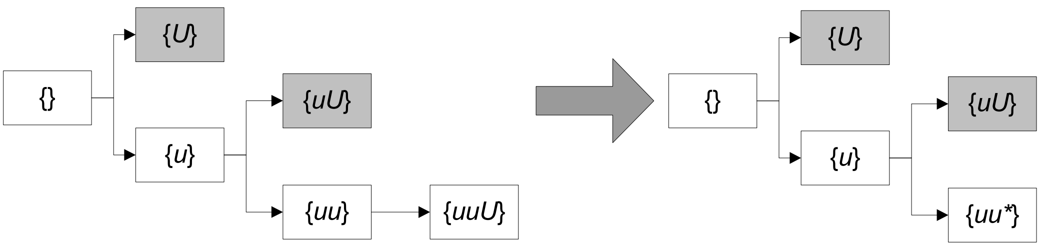

Figure 3: The graph of transitions between the groups for $0 < g <$ u, defined by the set of retained features. On the right, the simplified version where the groups from which there is no bifurcation are merged. Groups in grey have a negative attitude.

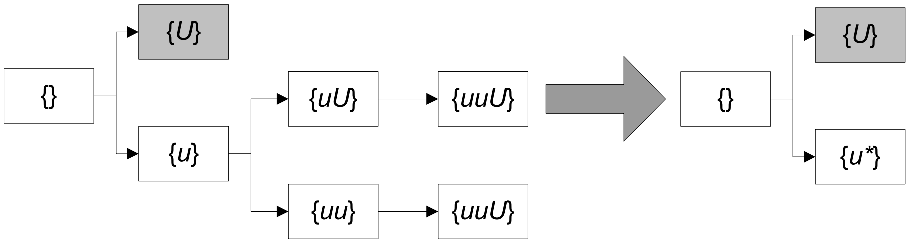

Figure 4: The graph of transitions between the groups for $u < g < 2u$, defined by the set of retained features. On the right, the simplified version, where the groups from which there is no bifurcation or give the same sign of final attitude are merged. The groups with a negative attitude are in grey.

3.2 Transfer equations

The second stage of the process is to determine the flows in each transition. This requires to evaluate the probability that the agents in each group retains a feature which makes them change their group. This probability is directly related to the features which are sent by each group. This is broken down in table 3 and 4.

Group	Media	$\{U\}$	$\{uU\}$	$\{u\}$	$\{uu*\}$
Communicated features	U, u	U	U	u	U

Table 3: communicated features for each group in the case $0 < g < u$.

Group	Media	$\{U\}$	$\{u*\}$
Communicated features	U, u	U	u

Table 4: communicated features for each group in the case $u < g < 2u$.

Let us call P_G the proportion of the population in group G, and $F_{G\to H}$ the flow from group G to group H by unit of time. The flows are written in table 5, when supposing that the probability to be in contact with a member of each group is directly proportional to its proportion in the population, in the case of total connection. Including different densities of networks would substantially modify the model, but it seems possible.

$0 < g < u$	$u < g < 2u$
(1.1) $F_{\{\ \}\to\{U\}} = P_{\{\ \}}\left(\frac{f}{3} + P_{\{U\}} + P_{\{uU\}}\right)$	(2.1) $F_{\{\ \}\to\{U\}} = P_{\{\ \}}\left(\frac{f}{3} + P_{\{U\}}\right)$
(1.2) $F_{\{\ \}\to\{u\}} = P_{\{\ \}}\left(\frac{2f}{3} + P_{\{u\}} + P_{\{uu*\}}\right)$	(2.2) $F_{\{\ \}\to\{u*\}} = P_{\{\ \}}\left(\frac{2f}{3} + P_{\{u*\}}\right)$
(1.3) $F_{\{u\}\to\{uU\}} = P_{\{u\}}\left(\frac{f}{3} + P_{\{U\}} + P_{\{uU\}}\right)$	
(1.4) $F_{\{u\}\to\{uu*\}} = P_{\{u\}}\left(\frac{f}{3} + \frac{1}{2}P_{\{u\}} + \frac{1}{2}P_{\{uu*\}}\right)$	

Table 5: Equations of the flows by unit of time between groups. Remember that f is the probability that agents receive a randomly chosen feature communicated by the media at each time step.

3.3 Comparing aggregated model with agent-based model

To check this qualitative reasoning, it is possible to simulate the evolution of the different group proportions from the equations of the aggregated model. To do so, we initialise the group proportions and iteratively modify them according to the flows, for a small time interval dt (for instance we chose $dt = 0.01$). More precisely:

- For $t = 0$, $P_{\{\ \}} = 1$, and all the other groups G, $P_G = 0$.
- Repeat until convergence:
    - For each group G initialise a variable $\Delta P_G = 0$,
    - For each flux $F_{G\to H}$, do: $\Delta P_G := \Delta P_G - F_{G\to H}.dt$ and $\Delta P_H := \Delta P_H + F_{G\to H}.dt$
    - For each group G do: $P_G := P_G + \Delta P_G$,

Therefore, at each time interval dt, we get new values for P_G for all groups G. We can compare these evolutions with the ones provided by the agent-based model.

We can observe on figure 4 that the aggregated model fits with good accuracy the agent-based model. In this case ($0 < g < u$, probability of media communication $f = 0.1$), in the end of the simulation, (not visible on the figure), the aggregated model predicts a proportion of 0.859 negative attitudes, to compare with 0.862, the mean result on 1000 replicates of the ABM. Moreover, figure 4 illustrates our qualitative analysis that groups $\{U\}$ and $\{uU\}$ feed each other, the last one at the expense of group $\{uu^*\}$. Both groups $\{U\}$ and $\{uU\}$ have proportions higher than 1/3, as expected.

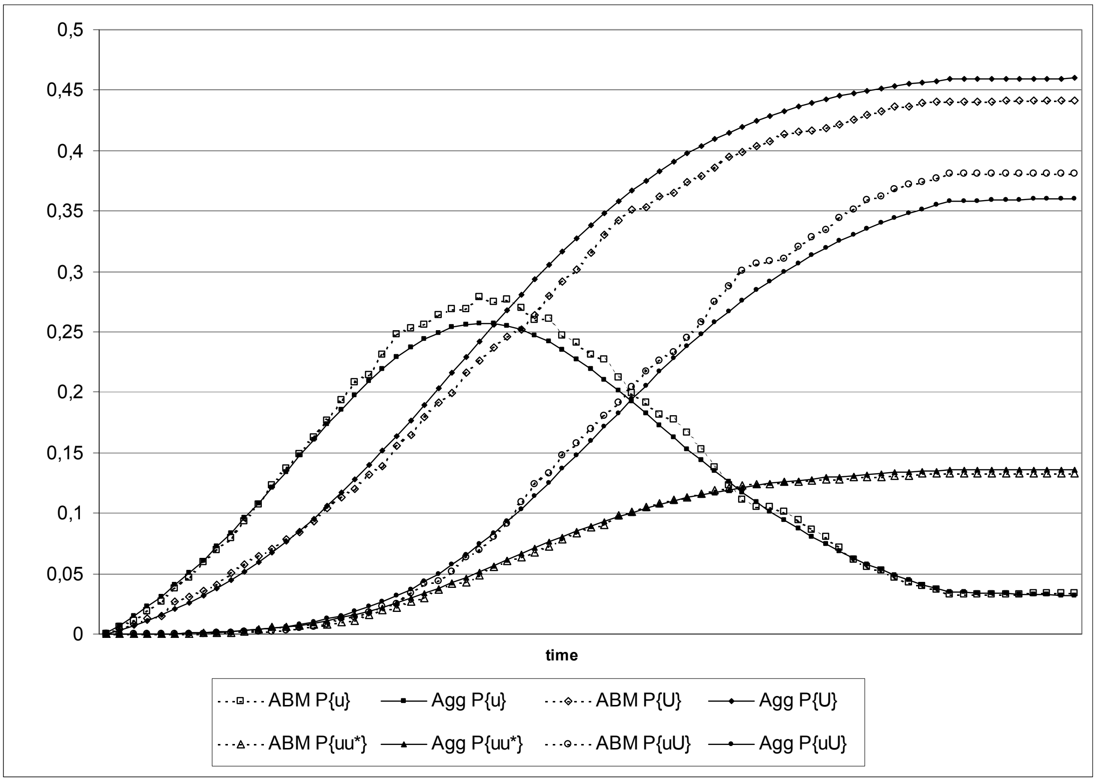

Figure 4. Comparison of evolutions of group proportions between one run of the agent based model (ABM) in total connection, and the aggregated model (Agg), for groups $\{u\}$, $\{U\}$, $\{uU\}$ and $\{uu^*\}$. The vertical axis is the proportion of the group. ABM curves are dotted lines, aggregated model curves are solid. $0 < g < u$. $f = 0.1$.

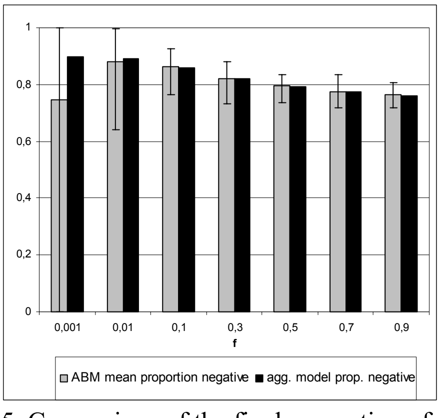

Figure 5. Comparison of the final proportion of negative attitudes for $0 < g < u$, between the ABM (grey) and the aggregated model for different values of f (black).

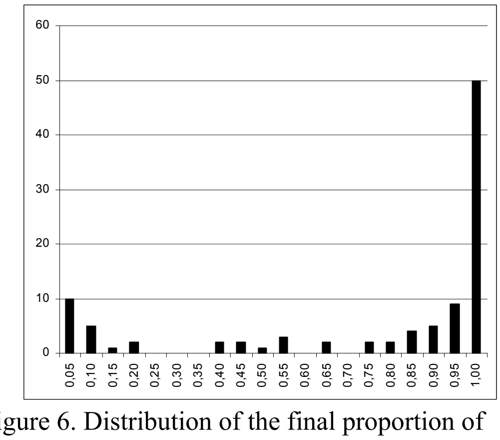

Figure 6. Distribution of the final proportion of negative global attitudes for $f = 0.001$

In order to evaluate the robustness of the approximation provided by the aggregated model, we tested it on several values of the frequency of media communication f. Figure 5 compares the final number of negative attitudes between aggregated and agent-based models when the probability of media communication f varies from 0.001 to 0.9, in total connection. The models significantly differ only in the case of $f = 0.001$. From previous work on the ABM (Deffuant and Huet 2006), we know that for weak frequency of diffusion (0.001 and less), the model tends to yield replicas in which either all final attitudes are positive, or all are negative.

Figure 6 shows a distribution of the results where this phenomenon is beginning to appear: the highest proportions of replicas are either all positive or all negative attitudes. Indeed, the agents which are the first in contact with the media have a determinant impact on the final result, because they have time to propagate the feature they received to almost the whole population. The mean-field type of approximation which is implicit in the aggregated model does not work anymore in such conditions. An other type of aggregated model, taking into account the probability of different replicas is necessary in this case.

3.4 Explaining from the aggregated model analyze the difference between isolated and interacting agents

The flow-equations provide valuable information about the behaviour of the system. Consider equations (2.1) and (2.2). At the first time step, among the set of agents leaving group {}, one third goes to group $\{U\}$, and two thirds to group $\{u*\}$. If this is true at the first time step, this remains true afterwards, because we then always have:

$$F_{\{\}\to\{U\}} = \frac{1}{2}F_{\{\}\to\{u*\}} \text{ which implies : } P_{\{U\}} = \frac{1}{3}\left(1 - P_{\{\}}\right) \text{ and } P_{\{u*\}} = \frac{2}{3}\left(1 - P_{\{\}}\right).$$

Because in the end $P_{\{\}} = 0$, we finally get:

$$P_{\{U\}} = \frac{1}{3} \text{ and } P_{\{u*\}} = \frac{2}{3}.$$

This corresponds to the results obtained by simulations.

In the case $0 < g < u$, suppose we had the following equations:

(1.1)’ $$F_{\{\}\to\{U\}} = P_{\{\}}\left(\frac{f}{3} + P_{\{U\}}\right)$$ instead of (1.1),

(1.3)’ $$F_{\{u\}\to\{uU\}} = P_{\{u\}}\left(\frac{f}{3} + P_{\{uU\}}\right)$$ instead of (1.3).

Then, applying the same reasoning as previously implies:

$$P_{\{U\}} = \frac{1}{3}\left(1 - P_{\{\}}\right),\ P_{\{u\}} = \frac{2}{3}\left(1 - P_{\{\}}\right) \text{ and } P_{\{uU\}} = \frac{1}{3}\left(1 - P_{\{\}} - P_{\{u\}}\right).$$

The additional terms $P_{\{uU\}}$ in (1.1) compared with (1.1)’, and $P_{\{U\}}$ in (1.3) compared with (1.3)’, imply:

$$P_{\{U\}} > \frac{1}{3}\left(1 - P_{\{\}}\right) \text{ and } P_{\{uU\}} > \frac{1}{3}\left(1 - P_{\{\}} - P_{\{u\}}\right).$$

Because in the end $P_{\{\}} = 0$ and $P_{\{u\}} = 0$, we get:

$$P_{\{U\}} > \frac{1}{3} \text{ and } P_{\{uU\}} > \frac{1}{3}.$$

The additional term $P_{\{uU\}}$ in (1.1) compared to (1.1)’ expresses that group $\{uU\}$ generates new members of group $\{U\}$, by sending feature U to members of group {}, and the additional term $P_{\{U\}}$ in (1.3) compared to (1.3)’ expresses that group $\{U\}$ generates new members of group $\{uU\}$, by sending feature U to members of group $\{u\}$. These new members are generated by the interactions among agents.

This qualitative analysis of the equations explains that in the case $0 < g < u$, the interactions between agents increase the propagation of the major feature into groups {} and $\{u\}$ compared to the simulations of isolated agents, or with the case $u < g < 2u$.

4 Discussion

The analytical aggregated model explains why the collective effect of FIB reinforcement takes place only when $0 < g < u$. The equations show that, in this case, the communication of U by agents of group $\{U\}$ to agents of group $\{u\}$, and by agents of group $\{uU\}$ to agents of group $\{\}$, increases the growth of groups $\{uU\}$ and$\{U\}$, compared to the case of isolated individuals. The simulations of the aggregated model show a quite good accuracy in predicting the agent-based model in the case of a completely connected network, and a reasonable one for different network topologies. From this reasoning, one can expect that the reinforcement will be higher for neutral objects with a single initially incongruent major feature and more minor features, because we can have more contributing groups ($\{uU\}$, $\{uuU\}$, $\{uuuU\}$…). We observed that below a threshold of frequency of media communication, the "mean-field" type approximation done for the aggregated model is not valid. Other approaches representing different groups of replicates should be considered to predict the final distribution of results. Adapting the approach to networks of different densities is also an interesting challenge.

The model remains simplistic, because we privilege an incremental approach, in which the model properties are precisely studied and understood (Deffuant et al. 2003). Despite this simplicity, there is still some work to understand it fully, for instance in the case of objects including several major features.

Moreover, we shall try to improve the model by relating it more closely to a set of psycho-sociology theories and observations. We shall particularly consider the minority influence theory (Moscovici 1979, Mugny & Pérez 1998 ; Moscovici S., 2000), in which the assimilation of incongruent information seems without effect for a while, and suddenly triggers a complete change of attitude.

5 References

Anderson N.H., 1971. Integration theory and attitude change. *Psychological Review*, 78, 171-206.

Allport G.W., Postman L., 1947. The psychology of rumor. Russel & Russell, Inc. (1965).

Barabasi A.L. and Reka A., 1999. Emergence of Scaling in Random Network. *Science*, vol. 286, October, 15th, 1999, pp. 509-512.

Eagly Alice H., Chaiken S., 1998. *The psychology of attitudes*. Thomson/Wadsworth; 1st edition (January, 1993), 800 pages

Deffuant G., Weisbuch G., Amblard F., Faure T., 2003. Simple is beautiful... and necessary. *Journal of Artificial Societies and Social Simulation* 6(1).

Deffuant G., 2004. Modéliser les Systèmes Complexes. Quelques pistes pour relever le défi. *Habilitation dissertation*.

Deffuant G., Huet S. and Amblard F., 2005. An individual-based model of innovation diffusion mixing social value and individual payoff dynamics. *American Journal of Sociology,* 110-4, January 2005, pp.1041-1069.

Deffuant G. and Huet S., 2006. Propagation effect of filtering incongruent information. *Journal of Business Research,* 9 p., forthcoming.

Erdös P. and Rényi A., 1960. Random Graphs. Publ. Math. Inst. Hung. Acad Sci. 5, 17, 1960; B. Bollobas, Academic Press, London, 1985.

Festinger L., 1957. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Galam S., 2003, Modelling rumors: the no plane Pentagon French hoax case. *Physica A*. Vol. 320. pp 571-580.

Huet S., Deffuant G., 2006. Effets d'un filtre cognitif sur la diffusion d'information. *MOSIM 2006*, Rabat.

Lawson G., and C. T. Butts, 2004, "Information Transmission Through Human Informants: Simulation", *CASOS'04*.

Mugny G., Pérez J.A., 1998. Recherches sur l'influence sociale. In Beauvois Jean-Léon, Joule Robert-Vincent, Monteil Jean-Marc, « 20 ans de psychologie sociale expérimentale francophone », *La psychologie sociale* n° 6, Presse Universitaire de Grenoble, 1998, p. 13-66.

Moscovici S., 1979. *Psychologie des minorités actives*, PUF, 1979/1996

Moscovici S., 2000. *Psychologie sociale des relations à autrui*. Nathan Université.

Tsimring L. S. and Huerta R., 2003. Modeling of Contact Tracing in Social Networks. *Physica A*, 325, pp. 33-39.

Urbig D., 2003. Attitude Dynamics With Limited Verbalisation Capabilities. *Journal of Artificial Societies and Social Simulation (JASSS)*. vol. 6 no. 1, 2003 <http://jasss.soc.surrey.ac.uk/6/1/2.html>