



**HAL**  
open science

# A general model of synchronous updating with binary opinions

Alexis Poindron

► **To cite this version:**

Alexis Poindron. A general model of synchronous updating with binary opinions. 2019. halshs-02372486

**HAL Id: halshs-02372486**

**<https://shs.hal.science/halshs-02372486>**

Submitted on 20 Nov 2019

**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.

CES

Centre d'Économie de la Sorbonne  
UMR 8174

**A general model of synchronous updating  
with binary opinions**

Alexis POINDRON

**2019.24**



# A general model of synchronous updating with binary opinions

Alexis Poindron\*

October 21, 2019

## Abstract

We consider a society of agents making an iterated yes/no decision on some issue, where updating is done by mutual influence under a Markovian process. Agents update their opinions at the same time, independently of each other, in an entirely mechanical manner. They can have a favourable or an unfavourable perception of their neighbours. We study the qualitative patterns of this model, which captures several notions, including conformism, anti-conformism, communitarianism and leadership. We discuss under which conditions opinions are stable. Finally, we introduce a notion of entropy that we use to extract information on the society and to predict future opinions.

## 1 Introduction

Prior to votes, where preferences of agents are aggregated to produce a social preference (see, e.g., Arrow (1963), Suzumura (1983)), preferences of agents interact with each other during debates, discussions, advertisements, etc. This process, called opinion formation, is the one studied in the present paper. It is embedded into the literature of opinion formation, diffusion and dynamics in social networks. We consider binary opinions (yes or no, adopt or not adopt, be active or inactive, first or second candidate, etc.), and influences among agents can be positive, negative or null. Updating is synchronous.

One of the most prominent models of continuous opinion formation has been proposed by DeGroot (1974) and French (1956). Many linear models have been subsequently developed, among which Abelson (1964), Taylor (1968) and Friedkin and Johnsen (1990). One can mention also Buechel et al. (2015) who introduce conformism and anti-conformism as a dishonest report of one's true opinion. Models of binary opinions, which have originally been studied by physicians, are of a special interest for us since this is the framework of this article. The Ising

---

\*Université Paris 1 Panthéon Sorbonne. Article written within the Phd thesis at the Centre d'Economie de la Sorbonne, ED 465, under the supervision of Michel Grabisch and Agnieszka Rusinowska. Special thanks for my supervisors, who followed this article from the beginning to the end. Several important ideas for Theorem 2 as well as many significant improvements are by Michel Grabisch. The author is grateful to Katarzyna Sznajd-Weron for her insight about synchronous and asynchronous models, and Serge Galam for our several discussions, in particular, on the effects of anti-conformism and inflexibility. Andrea Galeotti is also thanked for his remarks on congestion and saturation.

model introduced by Lenz and Ising (1925) has been subsequently declined into numerous variants. The cellular automata were introduced by von Neumann (1966) and Ulam (1960), where each cell updates its state, black or white, depending on its current state and the states of its neighbours. Another significant model inspired from the Ising one is the voter model (Liggett (1985), Liggett (1999)), introduced by Clifford and Sudbury (1973) to model the competition of species over a spatial territory. The  $q$ -voter model (Castellano et al. (2009)) is a variant of the voter model, where agents adopt the opinion of  $q$  neighbours. In the model of Sznajd-Weron and Sznajd (2001), rather than being influenced by all neighbours, agents are influenced by their neighbours agreeing with each other. In all these models, opinions are reversible, in the sense that agents can revise their opinion and change back to their initial opinion. These models of opinion dynamics are particularly interesting to model social phenomena. For this reason, Galam (2008) coins the word ‘sociophysics’ to designate this hybrid field, at the intersection of physics and economics.

In the same period as DeGroot (1974), Granovetter (1978) introduces the threshold model, where agents become activated when the proportion of activated agents exceeds a given threshold. Watts (2002) proposes a similar model of cascades in random networks. These papers investigate the question: ‘Can an opinion initially shared by a small number of agents propagate to a large part of the society?’. Models of cascades have been studied also by Banerjee (1992) and Bikhchandani et al. (1992), as well as by Grabisch et al. (2019) whose model is generalised in this paper. The case of anonymous influence is meant to capture situations where the name of agents does not matter or is not known, like opinions on the internet about the quality of a product. It is studied by Förster et al. (2013).

Models with negative influence are of special interest for us, since this paper is an extension of some existing models (Granovetter (1978), Grabisch and Rusinowska (2013), Förster et al. (2013)) to negative influence, in particular, to anti-conformism. More precisely, we revisit the general model of binary opinion dynamics with aggregation functions (Grabisch, 2016) presented in Grabisch and Rusinowska (2013), where all agents are conformists. Galam (2004) and Galam and Jacobs (2007) encompass anti-conformist and inflexible agents to the initial model of presidential election by successive elections of representatives (Galam, 1986). In the same vein, Nyczka and Sznajd-Weron (2013) and Nyczka et al. (2012) propose  $q$ -voter models with anti-conformist and independent agents. An extension of Förster et al. (2013) to negative influence is proposed by Grabisch et al. (2019). An adaptation of the Granovetter model by including anti-conformist agents is proposed by Grabisch and Li (2019). We can also mention Jull and Porter (2019), where adoption is irreversible and agents can be either conformists or anti-conformists, with some probability. Touboul (2014) mixes conformist and anti-conformist agents, called hipsters, with information delays, and the literature on coordination and anti-coordination. Notable models are Morris (2000) who proposes a local game investigating contagion in a network where agents have an interest to coordinate. Bramoullé et al. (2004), Bramoullé (2007) and López-Pintado (2009) study models of anti-coordination, where agents are inclined to play different actions like in the chicken game.

The aim of this paper is to propose a general model of binary opinion updating, where any kind of positive and negative influences are allowed. The model that we introduce is based on aggregation functions, in the vein of Grabisch and Rusinowska (2013), of which this paper is an extension, by including negative influence. Any kind of opinion updating is a particular case of our model, provided that: (i) alternatives are binary; (ii) decision making is reversible; (iii) updating is synchronous and (iv) the process is Markovian; (v) aggregation functions are fixed.

The paper is structured as follows. Section 2 exposes the model. It first recalls the main tools introduced in Grabisch and Rusinowska (2013). Section 2.2 introduces several tools which

permit to treat negative influence. We show that negative influence can be treated in the same manner as the positive one, by choosing an appropriate pre-order on states. Section 2.3 proposes a coalitional analysis, where the yes and no-influential coalitions introduced in Grabisch and Rusinowska (2013) are generalised to negative influence. These coalitions are shown to exhaustively describe the qualitative patterns of the model. Section 3 introduces a non-topological notion of groups. Contrary to the literature on networks which considers degrees, cliques, conventions (Jackson and Storms (2018)), etc., our notion of group is purely based on the signs of influence: two agents belong to the same group if they receive the same influences. The definition allows for a wide variety of topologies among a given group. This tractability allows to consider groups from a social point of view, that is, large networks of agents that do not necessarily know each other but share the same beliefs and the same influences in the large sense. Moreover, this section investigates relations between four classes of models, namely, conformism, leadership, communitarian (anti-coordination) and mixed model (with anti-conformist agents). We show that simple relations exist between these models, but we can split them between two categories associated to different dynamics. The conformist and the communitarian societies on the one hand, and the leadership and the mixed societies on the other hand, produce the same dynamics. In Section 4 we give several sufficient conditions to ensure stability of opinions. Section 5 introduces an index of entropy, which captures a certain notion of disorder by counting what we call ‘abnormalities’. We use entropy to define the notion of natural dynamics and to explain why some societies are more prone than others to opinion fluctuations. We introduce in Section 5.3 the minimal entropy principle, which is used in a similar manner as the maximum likelihood principle. Based on our notion of groups, an observer first assumes a certain type of society of which it recovers the parameters by minimising the number of abnormalities. Observing some time series of opinions, the observer can use the minimal entropy principle to recover information of the society and predict future opinions.

## 2 Description of the model

### 2.1 Generalities

We consider a set  $N := \{1, \dots, n\}$  of agents making an iterated yes-no decision on some issue. Each agent starts from an initial opinion and updates it due to mutual influence. Decisions are taken simultaneously by agents on the basis of the current *state of the world*  $S$ , simply called *state*, being the set of agents whose opinion is yes. We use the following convention: ‘Yes’ is coded by 1 and ‘No’ by 0. We write by  $1_S$  the indicator function of  $S$ , i.e.,  $1_S(i) = 1$  if  $i \in S$  and 0 otherwise. Given the state  $S$  of the society, the probability that the next state is  $T$  is written by  $b_{S,T}$ . Assuming that the process iterates, we obtain a stochastic process, called an *influence process*. The process is assumed to be: (i) Markovian, i.e.,  $b_{S,T}$  only depends on  $S$  and  $T$ , and not on the whole history of the process; (ii) stationary, i.e., time does not appear in the computation of  $b_{S,T}$ . We compute  $b_{S,T}$  from the probabilities  $p_i(S)$  of each agent  $i$  to say yes when the current state is  $S$ . If these probabilities are independent among agents, as it is assumed in this paper, then we have:

$$b_{S,T} = \prod_{i \in T} p_i(S) \prod_{i \notin T} (1 - p_i(S)). \quad (1)$$

We propose a qualitative description of the process described by the  $2^n \times 2^n$  row-stochastic transition matrix  $\mathbf{B} := [b_{S,T}]_{S,T \subseteq N}$ , that is, we examine the absorbing classes and their type

(for example, whether they are periodic or not). For this reason, we define the *reduced matrix*  $\mathbf{B}$ , where:

$$\tilde{b}_{S,T} = \begin{cases} 1 & \text{if } b_{S,T} > 0 \\ 0 & \text{if } b_{S,T} = 0 \end{cases} \quad (2)$$

When  $b_{S,T} > 0$ , we also write  $S \rightarrow T$ . If  $b_{S,T} = 1$ , we write  $S \xrightarrow{1} T$ . The reduced matrix can be equivalently represented by a *transition graph*  $\tilde{\Gamma} = (2^N, E)$ , where  $E$  is the set of directed arcs  $(S, T)$ , and an arc exists if and only if  $\tilde{b}_{S,T} = 1$ . We will also use the *weighted transition graph*  $\Gamma = (2^N, E, \mathbf{B})$ , where each arc  $(S, T)$  has weight  $b_{S,T}$ .

Opinion updating is based on aggregation functions (Grabisch, 2016), which aggregate in a single number the opinions of all agents.

**Definition 1.** The *aggregation function*  $A_i$  of agent  $i$  is a mapping from  $\{0, 1\}^n$  to  $[0, 1]$ . We denote by  $\mathbf{A} := (A_i)_{i=1, \dots, n}$  the vector of aggregation functions for all agents  $1, \dots, n$ .

The previous definition of an aggregation function does not impose any monotonicity conditions. Usually,  $A_i$  is assumed to be increasing (e.g, in the voter model or in Grabisch and Rusinowska (2013)). However, our model aims precisely at dealing also with negative influence. For this reason, we adopt a more general definition.

We identify the result of the aggregation function to the probability of saying yes:  $p_i(S) := A_i(1_S)$ . When the state of the society is  $S$ , the next state is given by the realisation of  $n$  Bernoulli random variables, like in Asavathiratham (2000) and Grabisch and Rusinowska (2013). Our model is therefore not deterministic (while, for example, the DeGroot model is). That is, agents aggregate first, and then toss a coin according to this probability, in order to take the decision to say yes or no.

*Remark 1.* If  $A_i(1_S) = 1 - A_i(1_{S^c})$  for all  $S$ , agent  $i$  treats in the same way the ‘yes’ and ‘no’ opinions, i.e., it is *unbiased* toward the word ‘yes’ or the word ‘no’ (otherwise, it is called ‘biased’). Assuming that aggregation functions are unbiased can be relevant to model elections for a candidate (but still, a biased aggregation function is some kind of political orientation), but not to model adoption of a new technology, in which case there is an investment cost. Aggregation functions representing majority influence, for example, are unbiased.

By the identification above, Equation (1) can be rewritten:

$$b_{S,T} = \prod_{i \in T} A_i(1_S) \prod_{i \notin T} (1 - A_i(1_S)). \quad (3)$$

Conversely, the weighted transition graph characterises  $\mathbf{A}$ , as stated by the following property:

**Property 1.**  $\mathbf{A}$  and  $\mathbf{B}$  are isomorphic.

*Proof.* Proofs are given in appendix. □

When needed, we will mention the vector of aggregation functions in superscript:  $\Gamma^{\mathbf{A}}$  or simply  $A_i$  when the object depends on the characteristics of a single agent.

A function taking values in  $\{0, 1\}$  is called *Boolean*.

We introduce some notations relative to sets. For two sets  $A \subseteq B$  in  $2^N$ , we write  $[A, B] := \{S \mid A \subseteq S \subseteq B\}$ . A collection of this form is called an *interval*. The cardinality of a set is denoted by the corresponding lower case, e.g.,  $s = |S|$ . The symmetric difference  $\Delta$  on sets  $S_1$  and  $S_2$  is defined by  $S_1 \Delta S_2 := (S_1 \cup S_2) \setminus (S_1 \cap S_2) = (S_1 \setminus S_2) \cup (S_2 \setminus S_1)$ . The complement

set of  $S$  is written  $S^c$ . If  $\mathcal{P} := \{S_1, \dots, S_k\}$  is a collection of sets, we write  $\mathcal{P}^{(c)} := \{S_1^c, \dots, S_k^c\}$  and  $\mathcal{P}^c := 2^N \setminus \mathcal{P}$ . We write also  $\mathcal{P}^{\Delta(Z)} := \{S_1 \Delta Z, \dots, S_k \Delta Z\}$ . The intersection (respectively union) of all the elements of a collection  $\mathcal{P}$  is written by  $\bigcap \mathcal{P}$  (respectively  $\bigcup \mathcal{P}$ ). For a mapping  $f$  defined on sets and  $\mathcal{P}$  a collection of sets, we write  $f(\mathcal{P}) = \{f(S) \mid S \in \mathcal{P}\}$ .

We recall some basic notions of Markov chains. Let  $\mathcal{C}$  be a nonempty collection of states.  $\mathcal{C}$  is *strongly connected* if either it is of the form  $\mathcal{C} = \{S\}$ , or for every distinct  $S, T \in \mathcal{C}$ , there is a path in  $\mathcal{C}$  from  $S$  to  $T$ . Moreover,  $\mathcal{C}$  is a *class* if it is strongly connected and maximal for this property, i.e., if no super-collection of  $\mathcal{C}$  is strongly connected. It is *absorbing* if there is no arc from some  $S \in \mathcal{C}$  to some  $T \notin \mathcal{C}$ . An *absorbing state*  $S$  is one such that  $S \xrightarrow{1} S$ . An absorbing class is *periodic* if for some  $k \geq 2$  there exists a partition  $\{P_1, \dots, P_k\}$  of  $\mathcal{C}$  such that when the process is at a state belonging to  $P_i$  at time  $t$ , then it will be in a state of  $P_{i+1}$  at time  $t + 1$ , with  $P_{k+1} := P_1$ . When each  $P_i$  reduces to a single state, we refer to the periodic class as a *cycle*.

When the sets  $S$  need to be explicit, we index agents by numbers. A common convention ignores the curly brackets and commas. For example,  $\{1, 2, 3\}$  can be written 123.

Theorem 1 below gives the shape of absorbing classes. This theorem does not involve any assumption on aggregation functions.

**Theorem 1.** (Grabisch and Rusinowska, 2013)

Consider an influence process based on aggregation functions **A**. Absorbing classes are:

- (i) either singletons (absorbing states)  $\{S\}$ ,  $S \in 2^N$ ,
- (ii) or cycles:  $S_1 \xrightarrow{1} S_2 \xrightarrow{1} \dots \xrightarrow{1} S_k \xrightarrow{1} S_1$ ,
- (iii) or collections  $\mathcal{C} = \mathcal{C}_1 \cup \dots \cup \mathcal{C}_p$ , where each collection  $\mathcal{C}_j$  is an interval  $[S_j, S_j \cup K_j]$  with at least one nonempty  $K_j$ .

In particular, if all aggregation functions are Boolean, absorbing classes are absorbing states or cycles. Collections  $\mathcal{C}^i$  in absorbing classes of the third kind are not necessarily pairwise disjoint, but they must be so if the transitions are sure from one  $\mathcal{C}^i$  to another (periodic classes). Absorbing classes of the third kind are either aperiodic classes, or periodic classes with at least one interval. In the sequel, we will refer to the *dynamics of a society* as being the kind of absorbing classes reached: (i), (ii) or (iii).

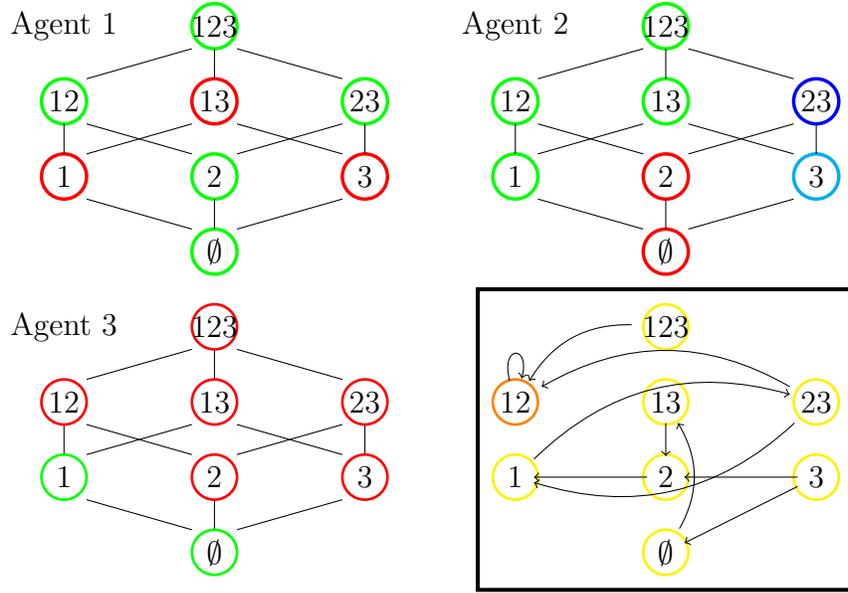
Let us see an example involving the notions introduced so far.

**Example 1.** Let  $N = 123$ . The aggregation functions characterised by the values taken on the  $2^n$  states are represented with coloured lattices. The resulting transition graph  $\tilde{\Gamma}$  is represented in yellow, at the bottom-right of Figure 1. We observe that the state 12 is absorbing ( $12 \xrightarrow{1} 12$ ) and that it is the unique absorbing class. The probabilities of transitions above the arrows of the transition graph  $\tilde{\Gamma}$  can be computed from (1). For example:  $b_{23,12} = 1 \cdot \frac{2}{3} \cdot (1 - 0) = \frac{2}{3}$ .

Another example is given in Figure 9, where a cycle appears:  $1 \xrightarrow{1} 23 \xrightarrow{1} 1$ .

## 2.2 Modelling positive and negative influence

So far we have not imposed any conditions on aggregation functions. In particular, aggregation functions are not necessarily monotonous in each entry. In order to introduce a new kind of monotonicity, we introduce the following  $\preceq_B$  order, which is the specific tool of this paper.



Three aggregation functions represented with coloured lattices  
 $A_i(1_S) = 1$  at green states.  $A_i(1_S) = \frac{2}{3}$  at blue states.  
 $A_i(1_S) = \frac{1}{3}$  at cyan states.  $A_i(1_S) = 0$  at red states.

Figure 1: From the aggregation functions to the transition graph (boxed)

### 2.2.1 The partial order $\preceq_B$

**Definition 2.** (Partial order  $\preceq_B$ )

Let  $S, S', B \in 2^N$ . We define the  $\preceq_B$  partial order (or simply the  $\preceq_B$  order) on sets by:

$$S_1 \preceq_B S_2 \Leftrightarrow \begin{cases} S_1 \cap B \subseteq S_2 \cap B \\ S_1 \cap B^c \supseteq S_2 \cap B^c \end{cases} \quad (4)$$

and  $S_1 \prec_B S_2$  if at least one of these two inclusions is strict.

That is,  $S_1 \preceq_B S_2$  if  $S_1$  is a subset of  $S_2$  within  $B$ , and a superset of  $S_2$  outside  $B$  ( $S_2$  is more engaged into  $B$  than  $S_1$ ).  $\preceq_N$  is the usual set order  $\subseteq$ , while  $\preceq_\emptyset$  is the reverse set order  $\supseteq$ .  $S_1 \preceq_B S_2$  is illustrated below. We note that  $S_1 \preceq_B S_2 \Leftrightarrow S_2^c \preceq_B S_1^c$  and that  $S_1 \preceq_B S_2 \Leftrightarrow B \preceq_{S_1} S_2$ .

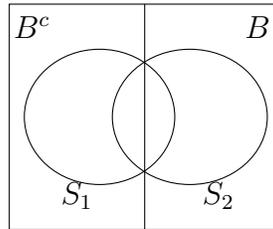


Figure 2: Illustration of  $S_1 \preceq_B S_2$ :  $S_2$  is more engaged into  $B$  than  $S_1$ .

Let us introduce the mapping:

$$f_B : \begin{cases} 2^N \rightarrow 2^N \\ S \mapsto S \Delta B^c \end{cases} \quad (5)$$

It is a bijection from  $2^N$  onto itself. Note that  $f_N = \text{id}$  and that  $f_\emptyset$  transforms each set into its complement. Importantly,  $2^N$  endowed with the partial order  $\preceq_B$  is a lattice  $\mathcal{L}_B := (2^N, \preceq_B)$ , with top element and bottom elements  $B$  and  $B^c$ , and with infimum and supremum  $\wedge_B, \vee_B$  given for any  $S, T \in 2^N$  by:

$$S \wedge_B T := [(S \cap T) \cap B] \cup [(S \cup T) \cap B^c] \quad (6)$$

$$S \vee_B T := [(S \cup T) \cap B] \cup [(S \cap T) \cap B^c] \quad (7)$$

For  $\wedge_B$ , we take the intersection within  $B$  and the union outside  $B$ . For  $\vee_B$ , we take the union within  $B$  and the intersection outside  $B$  (see Figure 3).  $\wedge_B$  and  $\vee_B$  coincide with  $\cap$  and  $\cup$  when  $B = N$ . The *upset* and *downset* of a given set  $S$  under the partial order  $\preceq_B$ , respectively denoted by  $\{S' \mid S \preceq_B S'\}$  and  $\{S' \mid S \succeq_B S'\}$ , are shortly written  $\uparrow_B S$  and  $\downarrow_B S$ . If  $B = N$ , we simply denote them by  $\uparrow S (= [S, N])$  and  $\downarrow S (= [\emptyset, S])$ .

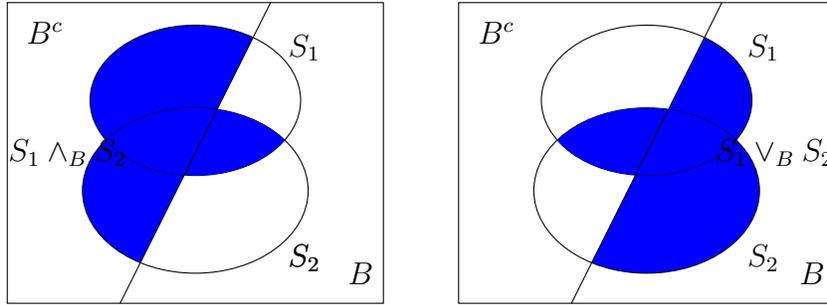


Figure 3: Infimum and supremum

$f_B$  is a lattice isomorphism from  $\mathcal{L}_N = (2^N, \subseteq)$  to  $\mathcal{L}_B = (2^N, \preceq_B)$ . Indeed, for any  $S, S' \in 2^N$ :

$$f_B(S \cap S') = f_B(S) \wedge_B f_B(S') \quad (8)$$

and

$$f_B(S \cup S') = f_B(S) \vee_B f_B(S') \quad (9)$$

Note that: (i)  $f_B(\emptyset) = B^c$ ; (ii)  $f_B(N) = B$  (bounded-lattice properties) and that we can write

$$S_1 \preceq_B S_2 \Leftrightarrow f_B^{-1}(S_1) \subseteq f_B^{-1}(S_2) \Leftrightarrow f_B(S_1) \subseteq f_B(S_2) \quad (10)$$

where the first equivalence comes from the fact that  $f_B$  is a lattice isomorphism and the second equivalence comes from the fact that  $f_B$  is an involution:  $f_B \circ f_B = \text{id}$ .

A *chain* is a sub-collection of  $2^N$ , totally ordered with respect to  $\preceq_B$ , while an *antichain* is a sub-collection of  $2^N$  whose elements are pairwise incomparable with respect to  $\preceq_B$ . In this paper, any chain (respectively antichain) is implicitly taken maximal, i.e., such that there exists no chain (respectively antichain) that strictly contains it.

## 2.2.2 Monotonicity collections and related notions

We now introduce another basic ingredient of our paper.

**Definition 3.** (Monotonicity collection  $\mathcal{M}_i$ ) Let  $A_i$  be the aggregation function of agent  $i$ . We define the *monotonicity collection*  $\mathcal{M}_i$  as follows:  $B \in \mathcal{M}_i$  if and only if  $S \preceq_B S' \Rightarrow A_i(1_S) \leq A_i(1_{S'})$ .

An agent  $i$  such that  $B \in \mathcal{M}_i$  says yes with a higher probability when more agents in  $B$  say yes, and less agents outside  $B$  say yes. Roughly speaking, agent  $i$  is positively influenced by agents in  $B$  and negatively by agents outside  $B$ . Let us see two important particular cases.

**Example 2.** (Conformism, Anti-conformism).

- (Conformism). An agent  $i$  is said to be *conformist* if  $N \in \mathcal{M}_i$ , that is, if for all  $S, S'$  with  $S \preceq_N S'$  (i.e.,  $S \subseteq S'$ ), it holds that  $A_i(1_S) \leq A_i(1_{S'})$  (i.e.,  $A_i$  is non decreasing). Conformist agents are those who say “yes” when more agents say “yes” (“more” in the inclusion sense).
- (Anti-conformism). An agent  $i$  is said to be *anti-conformist* if  $\emptyset \in \mathcal{M}_i$ , that is, if for all  $S, S'$  with  $S \preceq_{\emptyset} S'$  (i.e.,  $S \supseteq S'$ ), it holds that  $A_i(1_S) \leq A_i(1_{S'})$  (i.e.,  $A_i$  is non increasing). Anti-conformist agents are those who say “yes” when more agents say “no” (“more” in the inclusion sense).

Property 2 below gives the monotonicity collection of the aggregation function obtained by relabelling the states with the transformation  $f_B$ .

**Property 2.** (‘Rotation’ of the lattice) Assume that  $A_i$  is such that  $B \in \mathcal{M}_i$ . Then, the aggregation function  $A'_i$  with monotonicity collection  $\mathcal{M}'_i$ , obtained from  $A_i$  by  $A'_i(1_S) := A_i(1_{f_B(S)})$  for all  $S \in 2^N$ , is such that  $N \in \mathcal{M}'_i$ .

In what follows, we speak of *conformist polarisation* if all agents in the society are conformists (studied in Grabisch and Rusinowska (2013)), and we speak of *anti-conformist polarisation* if all agents of the society are anti-conformists. More generally, a polarisation is a collection of monotonicity collections for all agents.

**Definition 4.** The *polarisation* of  $\mathbf{A}$  (or the ‘polarisation of the society’) is the collection  $(\mathcal{M}_i)_{i \in N}$ .

Note that  $\mathcal{M}_i = \emptyset$  (see Example 3) is different from  $\mathcal{M}_i = \{\emptyset\}$  (which represents a particular case of anti-conformism).

The essential Property 3 below establishes the equivalence between the non-emptiness of  $\mathcal{M}_i$  and the independence of the signs of influence on the state of the world.  $\mathcal{M}_i \neq \emptyset$  means that there exists  $B$  such that  $S \preceq_B S' \Rightarrow A_i(1_S) \leq A_i(1_{S'})$ , depending on whether  $j \in B$  or  $j \notin B$ ,  $j$ ’s contribution to the opinion of agent  $i$  is either always positive or always negative. This property is important to make the subsequent Definition 5 meaningful.

**Property 3.** The following statements are equivalent:

- $\mathcal{M}_i \neq \emptyset$ .
- For any  $j \in N$ , either  $A_i(1_{S \cup j}) \geq A_i(1_S)$  for all  $S$ , or  $A_i(1_{S \cup j}) \leq A_i(1_S)$  for all  $S \in 2^N$ .

The following influence index is used to tell the nature of the influence (positive, negative or null) of a given agent on another one. Our analysis being exclusively qualitative, only the signs of influence matter for us. Property 3 justifies the absence of coefficients in front of the terms  $(A_i(1_{S \cup j}) - A_i(1_S))$ . This index really makes sense when  $\mathcal{M}_i \neq \emptyset$ , even though it would remain well defined otherwise.

**Definition 5.** (Influence index) Let  $i, j \in N$ . We define the *influence index* of  $j$  on  $i$  as follows:

$$\phi(j \rightarrow i) := \sum_{S \subseteq N \setminus j} (A_i(1_{S \cup j}) - A_i(1_S)). \quad (11)$$

If  $\phi(j \rightarrow i) > 0$  (respectively  $<, =$ ),  $j$  is said to have a *positive* (respectively *negative, null*) *influence* on  $i$ . We also say that an agent having null influence on  $i$  is *irrelevant* for this agent. In particular, under the assumption that  $\mathcal{M}_i \neq \emptyset$  for all  $i$ , we get from Property 3 that  $\phi(j \rightarrow i) = 0 \Leftrightarrow A_i(1_S) = A_i(1_{S \cup j})$  for all  $S \in 2^N$ .

The influence index reflects the qualitative influence relations between agents. We now define the graph of influence. There is an arrow from  $j$  to  $i$  if  $\phi(j \rightarrow i) \neq 0$ , either positive or negative<sup>1</sup>.

**Definition 6.** The *influence graph* is  $\mathcal{G} := (N, \mathcal{E})$  where the set of arcs  $\mathcal{E}$  is such that there is an arc from  $j$  to  $i$  if and only if  $\phi(j \rightarrow i) \neq 0$ .

Having established in Proposition 3 the equivalence between the non-emptiness of the monotonicity collection and the possibility to build an influence index which makes sense, let us see in Example 3 some examples of empty monotonicity collections.

**Example 3.** (Examples of empty  $\mathcal{M}_i$ : congestion, saturation)

Let  $A_i(1_S) = s(n - s)$  for all  $S$ . Then for any  $j$ ,  $A_i(1_{S \cup j}) \geq A_i(1_S)$  for  $s \leq \frac{n}{2} - 1$  and  $A_i(1_{S \cup j}) \leq A_i(1_S)$  for  $s \geq \frac{n}{2}$ . Because of the existence of  $j$  which  $i$  considers positively or negatively depending on the state of the world (only one suffices), agent  $i$  is such that  $\mathcal{M}_i = \emptyset$ . When all agents have this aggregation function, we get a model of saturation or congestion: agents are conformists or anti-conformists depending on the state of the world. This corresponds to situations like the frequentation of a restaurant. When there are not too many people in the restaurant, the dominant effect is the herd behaviour, since agents guess the quality of the restaurant from the number of clients. But when there are too many people, the congestion effect dominates. Since the restaurant gets noisy, the waiter is overwhelmed and service gets slow, etc., people do not choose the restaurant anymore.

Another example of agent  $i$  such that  $\mathcal{M}_i = \emptyset$  is an agent having a low self-esteem. It is positively influenced by itself when a lot of people share his opinion and negatively influenced by himself when few people share his opinion. One can think also of diplomatic or ethological situations, where peers can be friends in some states of the world, and enemies in other states of the world. The intricacies of friendships usually evolve depending on the presence of third parts. For example, mafias or nations which compete over small pieces of territory can unite against a common enemy. These situations will be excluded in this paper, as they would require a totally different mathematical treatment.

---

<sup>1</sup>Note that when negative influence is introduced, the topology of the influence graph and the influences do not coincide anymore. This implies that the common approach of centrality, betweenness, cliques, etc. are not applicable, or they must be taken at two separate levels. In our paper, we will not use any notion of centrality, though.

The previous examples are rather far-fetched. Hence, the following assumption is reasonable:

**Assumption.** From now on, we assume that  $\mathcal{M}_i \neq \emptyset$  for all  $i$ .

Theorem 2 below is used all through the article. It states that the collection  $\mathcal{M}_i$  is an interval which can be expressed with the influence index. Agents in  $\underline{B}_i$  exert a strict positive influence on  $i$ . Agents in  $(\overline{B}_i)^c$  exert a strict negative influence on  $i$ . Agents in  $\overline{B}_i \setminus \underline{B}_i$  are irrelevant. The baseline set of  $\mathcal{M}_i$  is  $\underline{B}_i$  and we can add some irrelevant agents to  $\underline{B}_i$ .

**Theorem 2.**  $\mathcal{M}_i = [\underline{B}_i, \overline{B}_i]$  where

$$\underline{B}_i := \bigcap_{B \in \mathcal{M}_i} B = \{j \in N \mid \phi(j \rightarrow i) > 0\} \quad (12)$$

$$\overline{B}_i := \bigcup_{B \in \mathcal{M}_i} B = \{j \in N \mid \phi(j \rightarrow i) \geq 0\} \quad (13)$$

$\mathcal{M}_i$  has only one element if and only if there exists no agent having a null influence on  $i$ , e.g., when it has a distance-based aggregation function (see Definition 10).

We can characterise conformism and anti-conformism (see Example 2) as follows:

$$i \text{ is a conformist agent} \Leftrightarrow N \in \mathcal{M}_i \Leftrightarrow \overline{B}_i = N \quad (14)$$

$$i \text{ is an anti-conformist agent} \Leftrightarrow \emptyset \in \mathcal{M}_i \Leftrightarrow \underline{B}_i = \emptyset \quad (15)$$

That is, an agent is conformist (respectively anti-conformist) if and only if no agent has a strictly negative (respectively strictly positive) influence on it.

Here are a few other examples of monotonicity collections.

**Example 4.** (Monotonicity collections) In Figure 1,  $\mathcal{M}_1 = \{2\}$ ,  $\mathcal{M}_2 = \{123\}$  (conformist agent) and  $\mathcal{M}_3 = \{\emptyset, 1\}$  (anti-conformist agent). Note that Agent 1 is not conformist. Indeed,  $N \notin \mathcal{M}_i$ , even though  $A_i(1_N) = 1$ . In Figure 5,  $\mathcal{M}_i = \{123\}$ . In Figure 9,  $\mathcal{M}_1 = \{123\}$ ,  $\mathcal{M}_2 = \mathcal{M}_3 = \{\emptyset\}$ .

Assumption BC naturally completes the assumption on the non-emptiness of monotonicity collections. It is not assumed by default in this paper. Note, in particular, that it is not an assumption of Theorem 2.

**Assumption BC.** For any  $i$  and  $B \in \mathcal{M}_i$ :

- $A_i(1_B) = 1$ ,
- $A_i(1_{B^c}) = 0$ .

Agents obeying to Assumption BC cannot be such that  $A_i(1_S) = c$  for all  $S$  (always tossing a coin). In Figure 1 with Agent 1, we have  $\{2\} \in \mathcal{M}_1$ ,  $A_1(1_2) = 1$  and  $A_1(1_{13}) = 0$ , and therefore Assumption BC holds. With Agent 2, we have  $N \in \mathcal{M}_2$ ,  $A_2(1_N) = 1$  and  $A_2(1_\emptyset) = 0$ , and therefore Assumption BC holds. With Agent 3, Assumption BC holds as well.

**Example 5.** [Anonymous influence] Let  $0 \leq s \leq n$  be the cardinality of the state  $S$  and  $p_i(s)$  the probability for agent  $i$  to say “yes” at the next period when the state is  $S$ . In Grabisch et al. (2019), two classes of aggregation functions are focused on:

$$\mathcal{Q}^c := \{p \mid p \text{ is nondecreasing and satisfies } p(0) = 0 \text{ and } p(n) = 1\} \quad (16)$$

$$\mathcal{Q}^a := \{p \mid p \text{ is nonincreasing and satisfies } p(0) = 1 \text{ and } p(n) = 0\} \quad (17)$$

Agents aggregating according to  $A_i \equiv p^c \in \mathcal{Q}^c$  are conformist, those aggregating according to  $A_i \equiv p^a \in \mathcal{Q}^a$  are anti-conformist (see Example 2); the coexistence of agents of these two groups corresponds to the mixed polarisation (or mixed society) in Section 3.2 (see Figure 9).<sup>2</sup>

### 2.2.3 The $\mathcal{P}_i^0$ and $\mathcal{P}_i^1$ collections

The collections  $\mathcal{P}_i^1$  and  $\mathcal{P}_i^0$  defined hereafter are fundamental objects for studying an influence process qualitatively. These are the collections such that the agent  $i$  says yes or no for sure.

**Definition 7.** Let  $i \in N$ . We introduce:

$$\mathcal{P}_i^1 := \{S \mid A_i(1_S) = 1\}. \quad (18)$$

$$\mathcal{P}_i^0 := \{S \mid A_i(1_S) = 0\}. \quad (19)$$

$$\mathcal{P}_i^* := 2^N \setminus (\mathcal{P}_i^0 \cup \mathcal{P}_i^1) = \{S \mid 0 < A_i(1_S) < 1\}. \quad (20)$$

$A_i$  is Boolean if and only if  $\mathcal{P}_i^* = \emptyset$ . Boolean aggregation functions are entirely defined with  $\mathcal{P}_i^1$  (or  $\mathcal{P}_i^0$ ). When Assumption BC holds,  $\mathcal{M}_i \subseteq \mathcal{P}_i^1$ . When aggregation functions are Boolean, equality holds if and only if  $\mathcal{P}_i^1$  is an interval.

**Example 6.** (Intervals) In Figure 1, the  $\mathcal{P}_i^1$  collections are represented in green, the  $\mathcal{P}_i^0$  collections are represented in red, and  $\mathcal{P}_i^*$  corresponds to the blue and cyan states. Agents 1 and 3 have Boolean aggregation functions.  $\mathcal{P}_3^1$  is an interval,  $\mathcal{P}_3^0$  is not. Neither  $\mathcal{P}_1^0$  nor  $\mathcal{P}_1^1$  is an interval.  $\mathcal{P}_2^1$  and  $\mathcal{P}_2^0$  are intervals.

When  $A_i$  is Boolean, intervals for  $\mathcal{P}_i^1$  and/or  $\mathcal{P}_i^0$  can be interpreted in terms of ‘boss sets’ (Hu and Shapley (2003a) and Hu and Shapley (2003b)).

**Definition 8.** (Boss sets) We say that  $K$  is a *boss set for  $i$*  if and only if  $i, K$  verify one of the following situations:

- (i)  $\mathcal{P}_i^1 = [K, N] = \uparrow K$ .  
 $i$  says yes for sure if and only if all agents of  $K$  say yes. In other words  $K \subseteq S \Leftrightarrow A_i(S) = 1$ .
- (ii)  $\mathcal{P}_i^1 = [\emptyset, K] = \downarrow K$ .  
 $i$  says yes for sure if and only if no agent outside  $K$  says yes. In other words,  $S \subseteq K \Leftrightarrow A_i(S) = 1$ .
- (iii)  $\mathcal{P}_i^0 = [K, N] = \uparrow K$ .  
 $i$  says no for sure if and only if all agents of  $K$  say yes. In other words,  $S \subseteq K \Leftrightarrow A_i(S) = 0$ .

---

<sup>2</sup>The reader is maybe expecting some considerations about nested or pairwise disjoint  $\mathcal{M}_i$ . While certainly essential in quantitative models, it seems that little information can be extracted in general from these considerations. Though disjoint  $\mathcal{M}_i$  have an interpretation in terms of groups (Section 3), and nested  $\mathcal{M}_i$  have an interpretation in terms of groups also and in terms of entropy (Section 5), such considerations are not essential in our model.

(iv)  $\mathcal{P}_i^0 = [\emptyset, K] = \downarrow K$ .

$i$  says no for sure if and only if no agent outside  $K$  says yes. In other words,  $K \subseteq S \Leftrightarrow A_i(S) = 0$ .

. If  $K = \{j\}$ , we say that  $j$  is a *boss* for  $i$ .

A *stubborn* agent can be defined as being a boss for himself: he only considers his own opinion.

The following property is straightforward.

**Property 4.**  $\tilde{\mathbf{B}}$  and  $(\mathcal{P}_i^0, \mathcal{P}_i^1)_{i \in N}$  are isomorphic.

It follows from Property 4 that knowing  $(\mathcal{P}_i^0, \mathcal{P}_i^1)_{i \in N}$  is sufficient to compute the absorbing classes. The converse, however, is wrong.

See Section 2.4.(i) to (viii) for a summary of the paper until now.

## 2.3 Coalitional analysis

In order to capture the qualitative patterns of the influence process, the ‘boundaries’ of  $\mathcal{P}_i^0$  and  $\mathcal{P}_i^1$ , Grabisch and Rusinowska (2013) propose the concepts of yes and no-influential coalitions in a conformist polarisation. (See also Grabisch and Rusinowska (2016), where the authors propose an algorithm to identify these coalitions).

**Definition.** (Conformist polarisation, Grabisch and Rusinowska (2013)) Assume that the polarisation of the society is conformist.

- A set  $S$  is a *yes-influential coalition* on  $i \in N$  if:  $A_i(1_S) > 0$  and  $\forall S' \subset S, A_i(1_{S'}) = 0$ .
- A set  $S$  is a *no-influential coalition* on  $i \in N$  if:  $A_i(1_{S^c}) < 1$  and  $\forall S' \subset S, A_i(1_{S'^c}) = 1$ .

The yes-influential coalitions describe the borders of  $\mathcal{P}_i^0$ . In the previous definition, a set  $S$  is a yes-influential coalition on  $i$  if the latter has some probability to say yes when agents in  $S$  say yes. This set is minimal in the sense that removing any agent from  $S$  leads agent  $i$  to say no for sure. The no-influential coalitions describe the borders of  $\mathcal{P}_i^1$ . A coalition is no-influential on  $i$  if the latter has some probability to say no when agents of the coalition  $S$  are saying no. Removing any agent of  $S$  leads agent  $i$  to say yes for sure. In some sense, a yes or a no-influential coalition is decisive for agent  $i$ .

Figure 5 illustrates these definitions. The symmetric difference of two sets linked in blue is a yes influential coalition. The aggregation function represented is the one of a conformist agent, i.e.,  $\bar{B}_i = N$ . We have for example  $\emptyset \Delta 23 = 23 \in \mathcal{Y}_i$  since  $A_i(1_{23}) > 0$  and  $A_i(1_S) = 0$  for any  $S \subset 23$ . In a conformist aggregation function, the yes influential coalitions coincide with sets.

This notion extends well to our general framework.

**Definition 9.** (General definition) Let  $i \in N$ .

A set  $S$  is a yes-influential coalition on  $i \in N$  if:

- $A_i(1_{f_{\bar{B}_i}(S)}) > 0$ ,
- $A_i(1_{f_{\bar{B}_i}(S')}) = 0$  for all  $S' \subset S$ .

A set  $S$  is a no-influential coalition on  $i \in N$  if:

- $A_i(1_{f_{\bar{B}_i^c}(S)}) < 1$ ,

- $A_i(1_{f_{\overline{B}_i^c}(S')}) = 1$  for all  $S' \subset S$ .

The collections of yes and no-influential coalitions on  $i$  are denoted by  $\mathcal{Y}_i$  and  $\mathcal{N}_i$ .<sup>3 4</sup>

The interpretation of yes and no-influential coalitions in the general model is essentially the same as in the conformist polarisation. We represent in Figure 4 below, a coalition  $S \in \mathcal{Y}_i$  (similar reasoning and picture for the no-influential coalitions). Let us write  $T := \overline{B}_i^c \Delta S$  (since  $S$  is typically close to  $\emptyset$ ,  $T$  is typically close to  $\overline{B}_i^c$ ). Then,  $A_i(1_T) > 0$  by definition of a yes-influential coalition. If we take  $S \subset S'$ , then either the part of  $T$  inside  $\overline{B}_i$  is smaller, or its part outside  $\overline{B}_i$  is bigger. The inclination of the agent to say yes disappears.

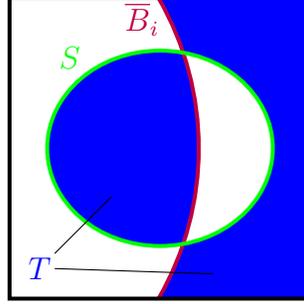


Figure 4: A yes or a no-influential coalition  $S$

**Property 5.** Assume that Assumption BC holds. Then:

$$\bigcup \mathcal{Y}_i \cup \bigcup \mathcal{N}_i \subseteq \{j \mid \phi(j \rightarrow i) \neq 0\} \quad (= \underline{B}_i \cup \overline{B}_i^c) \quad (21)$$

When  $A_i$  is Boolean, equality holds.

The reverse inclusion of Property 5 does not hold in general, as Example 7 shows.

**Example 7.** Let  $\mathcal{P}_i^1 = [1, 123]$ ,  $\mathcal{P}_i^0 = \{\emptyset, 2\}$ ,  $A_i(1_3) = \frac{1}{2}$  and  $A_i(1_{23}) = \frac{3}{4}$ . Then  $\mathcal{Y}_i = \{\{1\}, \{3\}\}$  and  $\mathcal{N}_i = \{\{1\}\}$ . However,  $\phi(2 \rightarrow i) \neq 0$ .

Since  $(\mathcal{P}_i^0, \mathcal{P}_i^1)_{i \in N}$  and  $\tilde{\mathbf{B}}$  are isomorphic, the other isomorphism established in Theorem 3 tells us that  $(\mathcal{Y}_i, \mathcal{N}_i)_{i \in N}$  is isomorphic to  $\tilde{\mathbf{B}}$  as well. Theorem 3 is illustrated by Figure 5.

**Theorem 3.** (Properties of  $\mathcal{Y}_i$  and  $\mathcal{N}_i$ )  $\mathcal{Y}_i$  (respectively  $\mathcal{N}_i$ ) and  $\mathcal{P}_i^0$  (respectively  $\mathcal{P}_i^1$ ) correspond bijectively.<sup>5</sup>

<sup>3</sup>The reader might wonder why the definition has been stated with  $\overline{B}_i$  rather than any other  $B \in \mathcal{M}_i$ . The answer is that it would be actually possible, because yes and no-influential coalitions contain no irrelevant agents on  $i$ . This is stated by Property 5. See also Figure 5.

<sup>4</sup>We must be careful with the terminology: a yes influential coalition is not a set of agents exerting a positive influence, and a no-influential coalition is not a set of agents exerting a negative influence. Assume that  $A_i$  is such that  $B \in \mathcal{M}_i$  and let the aggregation function  $A'_i$  with monotonicity collection  $\mathcal{M}'_i$  be obtained from  $A_i$  by:  $A'_i(1_S) := A_i(1_{f_B(S)})$ . Since for all  $Y, M, B \in 2^N$  we have  $\emptyset \Delta f_B(B^c \Delta Y) = Y$  and  $N \Delta f_B(B \Delta M) = M$ , the yes and no-influential coalitions of  $A_i$  and  $A'_i$  are identical. Therefore, the yes and no-influential coalitions have nothing to do with  $\mathcal{M}_i$ .

<sup>5</sup>Despite that  $\mathcal{Y}_i$  corresponds bijectively to  $\mathcal{P}_i^0$ , the yes-influential coalitions do not ‘cover’  $\mathcal{P}_i^0$ , in the sense that the collection inclusion  $\mathcal{P}_i^0 \subset \bigcup_{S \in \mathcal{Y}_i} \downarrow_{\overline{B}_i} f_{\overline{B}_i}(S)$  is wrong. This can be seen in Figure 5. The state 12 is not ‘below’ any yes-influential coalition: there exists no  $S \in \mathcal{Y}_i$  such that  $12 \subset S$ . The same remark holds for no-influential coalitions.

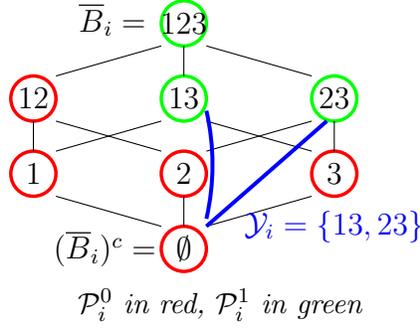


Figure 5: Theorem 3:  $\mathcal{Y}_i$  (blue) and  $\mathcal{P}_i^0$  (red) correspond bijectively.

We now want to localise the yes and no-influential coalitions from the transition graph. The first point of Theorem 4 below is adapted from Theorem 1 in Grabisch and Rusinowska (2013). When the considered transition is sure, then the second point of the theorem brings a precision to the first point. It generalises Lemma 4 in Grabisch and Rusinowska (2013).

**Theorem 4.** Assume that Assumption BC holds.

(1) (From  $\tilde{\mathbf{B}}$  to  $(\mathcal{Y}_i, \mathcal{N}_i)$ )

For all  $S, T \in 2^N$ ,  $\tilde{b}_{S,T} = 1$  if and only if:

- For each  $i \in T$ , there exists a non-empty  $Y \subseteq S\Delta B_i^c$  (where  $B_i \in \mathcal{M}_i$ ) such that  $Y$  is yes-influential on  $i$  and
- For each  $i \notin T$ , there exists a nonempty  $M \subseteq S\Delta B_i$  (where  $B_i \in \mathcal{M}_i$ ) such that  $M$  is no-influential on  $i$ .

(2) (From  $\mathbf{B}$  to  $(\mathcal{Y}_i, \mathcal{N}_i)$ )

Suppose that  $S \xrightarrow{1} T$ . Then, there cannot be a yes-influential coalition  $Y \subseteq S\Delta B_i^c$  on  $i$  (where  $B_i \in \mathcal{M}_i$ ) if  $i \notin T$ , or a no-influential coalition  $M \subseteq S\Delta B_i$  on  $i$  (where  $B_i \in \mathcal{M}_i$ ) if  $i \in T$ .

Since we already know that  $\tilde{\mathbf{B}}$ ,  $(\mathcal{P}_i^0, \mathcal{P}_i^1)_{i \in N}$  and  $(\mathcal{Y}_i, \mathcal{N}_i)_{i \in N}$  correspond bijectively, the possibility to extract  $(\mathcal{Y}_i, \mathcal{N}_i)_{i \in N}$  directly from  $\tilde{\mathbf{B}}$  (without using  $(\mathcal{P}_i, \mathcal{P}_i)_{i \in N}$ ) is not a new property, but a mere shortcut.

## 2.4 Summary of the notions introduced

At this stage, let us draw a summary of the notions introduced and their relations:

- (i) The primitive data of the model is the vector of aggregation functions  $\mathbf{A} = (A_i)_{i \in N}$ .
- (ii) From  $\mathbf{A}$ , we extract the transition graph  $\mathbf{B} = (b_{S,T})_{S,T \in 2^N}$ :

$$b_{S,T} = \prod_{i \in T} A_i(1_S) \prod_{i \notin T} (1 - A_i(1_S)).$$

- (iii)  $\mathbf{A}$  and  $\mathbf{B}$  are isomorphic. That is,  $\mathbf{B}$  could also be primitive of the model.
- (iv) Absorbing classes are obtained from  $\tilde{\mathbf{B}}$ .

(v) We define the  $\preceq_B$  order and the influence index  $\phi$ :

$$\phi(j \rightarrow i) = \sum_{S \subseteq N \setminus j} \underbrace{[A_i(1_{S \cup j}) - A_i(1_S)]}_{\text{same sign } \forall S}.$$

. If  $\phi(j \rightarrow i) = 0$ ,  $j$  is said to be irrelevant on  $i$ .

(vi) The monotonicity collections  $\mathcal{M}_i$  are extracted from the aggregation functions  $A_i$ . It determines which agent has positive or negative influence on  $i$ .

$B \in \mathcal{M}_i$  if and only if for any  $S, S' \in 2^N$ :

$$S \preceq_B S' \Rightarrow A_i(1_S) \leq A_i(1_{S'})$$

We assume that  $\mathcal{M}_i \neq \emptyset$  for all  $i$  (same sign  $\forall S$  in the sum of  $\phi$ ).

(vii) We express the interval  $\mathcal{M}_i$  in terms of influence indices  $\phi(j \rightarrow i)$ :

$$\mathcal{M}_i = [\underline{B}_i, \overline{B}_i]$$

where  $\underline{B}_i := \bigcap_{B \in \mathcal{M}_i} B = \{j \in N \mid \phi(j \rightarrow i) > 0\}$  and  $\overline{B}_i := \underline{B}_i \cup K_i$  and  $K_i$  the set of irrelevant agents on  $i$ .

(viii)  $(\mathcal{P}_i^0, \mathcal{P}_i^1)_{i \in N}$  are obtained from  $A_i$ :

$$\mathcal{P}_i^1 := \{S \mid A_i(1_S) = 1\}$$

and

$$\mathcal{P}_i^0 := \{S \mid A_i(1_S) = 0\}$$

(ix) From  $(\mathcal{P}_i^0, \mathcal{P}_i^1)_{i \in N}$ , we extract  $(\mathcal{Y}_i, \mathcal{N}_i)_{i \in N}$ :

- $S$  is yes-influential on  $i$  if and only if  $A_i(1_{f_{\overline{B}_i}(S)}) > 0$  and  $A_i(1_{f_{\overline{B}_i}(S')}) = 0$  for any  $S' \subset S$ .
- $S$  is no-influential on  $i$  if and only if  $A_i(1_{f_{\overline{B}_i^c}(S')}) < 1$  and  $A_i(1_{f_{\overline{B}_i^c}(S)}) = 1$  for any  $S' \subset S$ .

These coalitions do not contain irrelevant agents.

(x)  $\tilde{\mathbf{B}}, (\mathcal{P}_i^0, \mathcal{P}_i^1)_{i \in N}$  and  $(\mathcal{Y}_i, \mathcal{N}_i)_{i \in N}$  correspond bijectively.

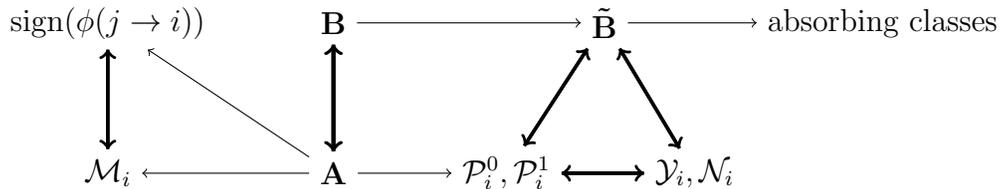


Figure 6: Summary and relations between the introduced notions

## 2.5 Special cases

### 2.5.1 Self-influence

A natural idea is that agents positively weigh their own opinion, i.e., they “agree with themselves”. Assumption SI formalises this idea, but is not assumed by default in this paper.

**Assumption SI:**  $i \in \underline{B}_i$  for all  $i$ .

Assumption SI means that all agent have a strict influence on themselves. Equivalently (Proposition 3), for all  $i$ , there exists  $S \in N \setminus i$  such that  $A_i(1_S) < A_i(1_{S \cup i})$ .<sup>6</sup>

Though Assumption SI sounds very natural, it might be relevant, in some contexts, to assume that  $\phi(i \rightarrow i) < 0$ . Here are a few examples.

- Like in SIR models, ‘saying yes’ can be seen as ‘being infected’. Stating  $\phi(i \rightarrow i) < 0$  allows to model the recovery from the disease. More generally, it can model an excitation which calls for a subsequent inhibition in a biologic cycle.
- ‘Saying yes’ can be ‘passing on a piece of information’. If there is a cost for passing the information, then the agent will convey the information for one or a few periods, until enough people received it. Then, it will stop.
- From a psychological point of view, a vote can be the acquisition of an object, and failure of Assumption SI would model state-dependant utility, where utility for the object decreases once we acquire it (see Girard (1966), Girard (1977)).
- From a managerial point of view, a vote can be a check of the quality of a product in the fabrication process. An agent probing the product at time  $t$  needs not probe it again at time  $t + 1$ . An agent anti-coordinates on himself when he gains from spreading his action over time rather than concentrating on a contained period.

We expect Assumption SI to be linked to issues of stability, which are of particular interest for us. It is indeed involved in some sufficient conditions to forbid cycles (Proposition 3 in Section 4).

### 2.5.2 Distance-based aggregation functions and Generalised Weighted Means

We first define the distance-based aggregation functions. They generalise the notion of anonymous influence (Example 5). An aggregation function is distance-based if there exists a pole  $T$  such that two states  $S, S'$  differing from  $T$  by the same number of elements are such that  $A_i(1_S) = A_i(1_{S'})$ .

**Definition 10.**  $A_i$  is said to be *distance-based* if there exists a state  $T$ , called a *pole*, such that:  $|S \Delta T| = |S' \Delta T| \Rightarrow A_i(1_S) = A_i(1_{S'})$ . If the aggregation function takes its highest value on  $T$ , then  $T$  is called the *pole* of  $A_i$ .<sup>7</sup>

Distance-based aggregation functions have a unique pole. If the agent is conformist or anti-conformist, a distance-based aggregation function is anonymous. In Example 5, the pole of  $A_i \equiv p \in \mathcal{Q}^c$  is  $N$ ; the pole of  $A_i \equiv p \in \mathcal{Q}^a$  is  $\emptyset$ . See Figures 7 and 9 for examples of

---

<sup>6</sup>One might prefer, as a weaker assumption, to assume that  $i \in \overline{B}_i$  for all  $i$  (agents do not have a strict negative influence on themselves), but our convention is easier to work with in Proposition 3.

<sup>7</sup>In this case, note that  $\mathcal{M}_i = \{T\}$ .

distance-based aggregation functions. We introduce the following notation, which will be used in Sections 4 and 5:

$$\mathcal{V}^r(T) := \{S : |S\Delta T| \leq r\}. \quad (22)$$

Note that distance-based aggregation functions have  $\mathcal{P}_i^0$  and  $\mathcal{P}_i^1$  of the form  $\mathcal{P}_i^1 = \mathcal{V}^{r_1}(T)$  and  $\mathcal{P}_i^0 = \mathcal{V}^{r_0}(T^c)$ .

We define below a generalisation of weighted means introduced in (Grabisch and Rusinowska, 2013). It is more convenient, in this definition, to map  $A_i$  into  $[-1, 1]$  rather than  $[0, 1]$ , so that passing to probabilities requires to rescale aggregation functions as follows:  $p_i(S) := \frac{1}{2}(A_i(1_S) + 1)$ .

**Definition 11.** The family of *Generalised Weighted Means* (GWM) is defined by aggregation functions such that  $A(x_1, \dots, x_n) = f^{-1}\left(\sum_{j=1}^n w_j f(x_j)\right)$  where  $f$  is a continuous automorphism on  $[-1, 1]$ ,  $(x_1, \dots, x_n) \in \{-1, 1\}^n$  and  $w_1, \dots, w_n \in [-1, 1]$ , called the *weights*, are such that  $\sum_{i=1}^n |w_j| = 1$ . Such aggregation functions are called a *GWM aggregation function*, and a society where all agents aggregate according to a GWM function is called a *GWM society*.

The GWM aggregation function of an agent  $i$  is unbiased (See Remark 1) if and only if  $f$  is an odd function.

Proposition 1 gives some precisions about the GWM aggregation functions. The last point relates the distance-based with the GWM aggregation functions.

**Proposition 1.** Assume that  $i$  is an agent who aggregates opinions according to a GWM aggregation function with weights  $(w_j^i)_{j=1, \dots, n}$ . Then:

- (1)  $\mathcal{M}_i \neq \emptyset$  for all  $i$  and Assumption BC hold.
- (2)  $\phi(j \rightarrow i) > 0$  (resp.  $< 0$ ) if and only if  $w_j^i > 0$  (resp.  $< 0$ ).
- (3) The sets in  $\mathcal{N}_i$  and  $\mathcal{Y}_i$  are singletons.
- (4)  $j$  is yes-influential on  $i$ , if and only if  $j$  is no-influential on  $i$ , if and only if  $w_j^i \neq 0$ .
- (5) Equality holds in Property 5.
- (6)  $A_i$  is a distance-based aggregation if and only if  $|w_i^j| = \frac{1}{n}$  for all  $j$ . In this case, its pole is  $T = \{j \mid w_j^i > 0\}$ .

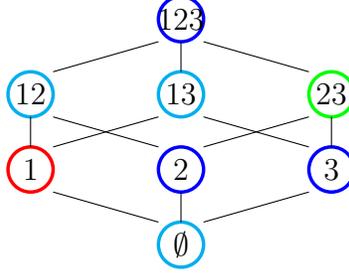
The aggregation function represented in Figure 7 is a GWM function, with  $w_2^i = w_3^i = \frac{1}{3}$  and  $w_1^i = -\frac{1}{3}$ . This aggregation function is also distance-based, and its pole is  $T = 23$  (Proposition 1.(6)).

### 2.5.3 Characterisation of absorbing states and intervals

The characterisation of absorbing class in Grabisch and Rusinowska (2013) extends well to polarisations with negative influence. For example, Theorem 5 below is the adaptation of their Theorem 3.

**Theorem 5.** Consider an influence process  $\mathbf{B}$  based on aggregation functions  $\mathbf{A}$ . Then,  $S$  is an absorbing state if and only if:

- (i) For any  $i \in N$  and  $B_i \in \mathcal{M}_i$ , there exists no no-influential coalition  $T \preceq_{B_i} S^c$  on some  $i \in S$ .



$$\begin{aligned}
 A_i(1_S) &= 1 \text{ at green states. } A_i(1_S) = \frac{2}{3} \text{ at blue states.} \\
 A_i(1_S) &= \frac{1}{3} \text{ at cyan states. } A_i(1_S) = 0 \text{ at red states.} \\
 f &= \text{Identity}
 \end{aligned}$$

Figure 7: A GWM aggregation function, which is also distance-based

- (ii) For any  $i \in N$  and  $B_i \in \mathcal{M}_i$ , there exists no yes-influential coalition  $T \preceq_{B_i} S$  on some  $i \notin S$ .

When the polarisation of the society is not specified, the previous theorem is cumbersome and of little use. Indeed, the difference with the conformist model is that we must specify a different preorder for each agent<sup>8</sup>. Similar conditions can be produced for interval absorbing classes (Theorem 4 in Grabisch and Rusinowska (2013)):

**Theorem 6.** Consider an influence process  $\mathbf{B}$  based on aggregation functions  $\mathbf{A}$  and assume that  $[S, S \cup K]$  is strongly connected. This interval is an absorbing interval if and only if:

- (i) For any  $i \in N$  and  $B_i \in \mathcal{M}_i$ , there exists no no-influential coalition  $T \preceq_{B_i} S^c$  on some  $i \in S$ .
- (ii) For any  $i \in N$  and  $B_i \in \mathcal{M}_i$ , there exists no yes-influential coalition  $T \preceq_{B_i} S$  on some  $i \notin S \cup K$ .

## 3 Groups

### 3.1 Definition

There are two main approaches for defining groups. Either we adopt a topological approach by considering cliques, conventions (Jackson and Storms, 2018), etc., that is, we exploit the properties of the graph, or we base the definition of groups on the signs of influence. We adopt the latter approach: our notion of groups is non-topological.

**Definition 12.** Let  $(N_1, \dots, N_g)$  be a partition of  $N$  into  $g$  blocks of agents. These blocks are called *groups* if  $\bigcap_{i \in N_k} \mathcal{M}_i \neq \emptyset$  for all  $k \in \{1, \dots, g\}$ . We write  $N_{(i)}$  for the group that agent  $i$  belongs to.

Agents belong to the same group if their aggregation functions can be polarised in the same direction. There exists a common  $\preceq_B$  that orientates the aggregation function of all agents

<sup>8</sup>However, in leadership models, introduced in the next subsection, a common preorder exists for all agents. In this case, Theorem 5 becomes more useful.

of the group. More precisely, they share a common set of agents who influence them in a nonnegative way. However, as illustrated in Example 8, taking two agents  $i, j$  in a group, an agent  $k$  may have a positive (or negative) influence for  $i$  but be irrelevant (no influence) for  $j$ , and vice versa for another agent  $k'$ . This makes our definition of groups flexible.

**Example 8.** (Flexibility of the definition)  $\mathcal{M}_1 = [13, 123]$ ;  $\mathcal{M}_2 = [12, 123]$ . Then,  $\mathcal{M}_1 \cap \mathcal{M}_2 = \{123\} \neq \emptyset$ : Agents 1 and 2 can belong to the same group. However, Agent 3 has a strict positive influence on Agent 1, while Agent 2 is irrelevant for him. For Agent 2, the roles are inverted: Agent 3 is irrelevant and Agent 2 exerts a strict positive influence on himself.

The following constraint, not assumed by default in this paper, requires that no coarser partition exists.

**Assumption C:** For all  $j, k \in \{1, \dots, g\}$ ,  $\bigcap_{i \in N_j \cup N_k} \mathcal{M}_i = \emptyset$ .

The partition of  $N$  into groups always exists, as shown by the following algorithm. (i) Start from a partition of  $N$  into single agents. (ii) For each  $(k, p)$ , check whether  $\bigcap_{i \in N_p \cup N_k} \mathcal{M}_i \neq \emptyset$ . If this is the case, merge groups  $p$  and  $k$ , and continue the process until no merging is possible anymore. At the extreme, it can be that the partition is  $\{\{1\}, \dots, \{n\}\}$ . The partition is not unique in general. To see this, note that we can always split a group into several groups. In this case, one partition is coarser than the other. But even with Assumption C, the partition of  $N$  into groups is not unique in general, as shown by Example 9.

**Example 9.** (Non-uniqueness of the partition) Let  $\mathcal{M}_1 := [\emptyset, 1]$ ,  $\mathcal{M}_2 := [1, 123]$  and  $\mathcal{M}_3 := [23, 123]$ . We can partition  $N$  into  $\{12, 3\}$  or into  $\{1, 23\}$ .

It is important to note that agents of the same group do not necessarily exert positive influence on each other. Indeed, in Example 9,  $\phi(2 \rightarrow 1) < 0$ . However,  $\{12, 3\}$  is a possible partition into groups.

**Definition 13.** Let  $N = N_1 \cup \dots \cup N_g$  be a partition into groups of  $N$ . We say that  $S$  is *made of blocks* if there exists  $J \subseteq [1, g]$ , called the *corresponding indices of  $S$* , such that  $S$  is a union of groups:  $S = \bigcup_{j \in J} N_j$ . We say that a class  $\mathcal{C}$  is *made of blocks* if  $\mathcal{C}$  is a singleton made of blocks, or a cycle where each state is made of blocks, or a periodic class where the states  $S_j, K_j$  of point (iii) of Theorem 1 are made of blocks.

Recall that Theorem 2 states that  $\mathcal{M}_i$  is an interval, i.e., of the form  $[\underline{B}_i, \overline{B}_i]$ . Assumption BA (“blocks assumption”) states that the set of agents influencing a given agent is made of blocks. It is not assumed by default in this paper.

**Assumption BA:** There exists a partition into groups such that for all  $i = 1, \dots, n$ ,  $\underline{B}_i$  and  $\overline{B}_i$  are made of blocks.

It can be shown that under Assumptions SI and BA agents of the same group have non-negative influence on each other.

A particular case of partition where agents exert positive influence on each other is the one of communitarian groups, i.e., communities of agents who not only have a positive perception of agents from their group, but also a negative perception from agents of the other groups. Example 10 introduces this notion.

**Example 10.** A *communitarian group* is a group  $N'$  such that  $N' \in \mathcal{M}_i$  for all  $i \in N'$ . When all groups are communitarian, we say that the society is *communitarian*. If the aggregation function of an agent  $i$  is distance-based, then the pole is  $N_{(i)}$ . The agent says yes when the state of the society is ‘not too far’ from its group. The case of two communitarian groups is referred to in the literature as ‘anti-coordination model’ (Bramoullé et al. (2004), Bramoullé (2007) and López-Pintado (2009)).

We end this section by making a remark on the relation between groups and absorbing classes. Intuitively, one would expect that absorbing classes are somehow reflecting the structure of the society into groups. Indeed, we think of groups as sets of agents behaving in the same manner (hanging out together, wearing in the same manner, adopting the same opinion, etc.). This is however not true in general in our framework. The main reason is that, as it can be seen on Figure 6, the monotonicity collections, on which the definition of groups is grounded, have no tight link with absorbing classes. In Section 5, we will see that it is however natural in some sense that agents of the same group share the same opinion.

### 3.2 Analysis of the case $g = 2$

This subsection is devoted to the analysis of the case of two groups  $N_1$  and  $N_2$ .

We do not impose Assumption C here, but we impose Assumption BA: each agent weighs positively or negatively whole blocks of agents. Therefore, we can summarise the polarisation of the society  $(\mathcal{M}_i)_{i \in N}$  by a signed matrix:

$$\mathcal{V} := \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \quad (23)$$

where  $k_{ij} = +$  if and only if  $N_j \subseteq \overline{B}_i$ , and  $k_{ij} = -$  if and only if  $N_j \subseteq \overline{B}_i^c$ , where  $i, j \in \{1, 2\}$ ,  $\bigcap_{i \in N_1} \mathcal{M}_i = [\underline{B}_1, \overline{B}_1]$  and  $\bigcap_{i \in N_2} \mathcal{M}_i = [\underline{B}_2, \overline{B}_2]$ . If, for example  $k_{12} = +$ , then the whole group  $N_2$  has a positive influence on the whole group  $N_1$ .

Four (up to symmetries) among the sixteen possible types of polarisations call for particular interest: the conformist, mixed, leadership and communitarian polarisations, which we detail below (see Table 1 for a summary of these types), together with some general properties about convergence (they will be clarified in Section 5; see also Table 2 giving the transitions from  $\emptyset, N_1, N_2$  and  $N$  under Assumption BC).

Conformist polarisation C $N \in \mathcal{M}_i$ for all $i \in N$	Mixed polarisation M <sub>1</sub> $N \in \mathcal{M}_i$ for $i \in N_1$ $\emptyset \in \mathcal{M}_i$ for $i \in N_2$
Leadership polarisation L <sub>1</sub> $N_1 \in \mathcal{M}_i$ for all $i \in N$	Communitarian polarisation R <sub>1</sub> $N_1 \in \mathcal{M}_i$ for $i \in N_1$ $N_2 \in \mathcal{M}_i$ for $i \in N_2$

Table 1: The C, M<sub>1</sub>, L<sub>1</sub> and R<sub>1</sub> polarisations.

(C) *Conformist polarisations.*  $\mathcal{V}_C := \begin{pmatrix} + & + \\ + & + \end{pmatrix}$

Conformist polarisation C $\emptyset \xrightarrow{1} \emptyset$ $N \xrightarrow{1} N$	Mixed polarisation M <sub>1</sub> $\emptyset \xrightarrow{1} N_2$ $N \xrightarrow{1} N_1$
Leadership polarisation L <sub>1</sub> $N_2 \xrightarrow{1} \emptyset$ $N_1 \xrightarrow{1} N$	Communitarian polarisation R <sub>1</sub> $N_1 \xrightarrow{1} N_1$ $N_2 \xrightarrow{1} N_2$

Table 2: Some important transitions

$N_1 \subseteq \overline{B}_1$ ,  $N_1 \subseteq \overline{B}_2$ ,  $N_2 \subseteq \overline{B}_1$  and  $N_2 \subseteq \overline{B}_2$ . Hence,  $\overline{B}_1 = \overline{B}_2 = N$ : a partition into a single group is possible. We already introduced the conformist polarisation as being one where all agents are conformist, i.e., such that  $N \in \bigcap_{i \in N} \mathcal{M}_i$ . In particular, the society can

be partitioned into a single group. Generally speaking, opinions in conformist polarisations are prone to reach an absorbing state.

(R<sub>1</sub>) *Communitarian polarisations.*  $\mathcal{V}_R := \begin{pmatrix} + & - \\ - & + \end{pmatrix}$

$N_1 \subseteq \overline{B}_1$ ,  $N_2 \subseteq \overline{B}_2$ ,  $N_2 \subseteq \overline{B}_1^c$  and  $N_1 \subseteq \overline{B}_2^c$ . A communitarian polarisation is one with two communitarian groups (cf. Example 10). It is such that  $\phi(i \rightarrow j) \geq 0 \Leftrightarrow \phi(j \rightarrow i) \geq 0$  and  $\phi(i \rightarrow j) \leq 0 \Leftrightarrow \phi(j \rightarrow i) \leq 0$ . Generally speaking, polarisations R are prone to reach an absorbing state.

(M<sub>1</sub>) *Mixed polarisations.*  $\mathcal{V}_M := \begin{pmatrix} + & + \\ - & - \end{pmatrix}$

$N_1 \subseteq \overline{B}_1$ ,  $N_2 \subseteq \overline{B}_1$ ,  $N_1 \subseteq \overline{B}_2^c$  and  $N_2 \subseteq \overline{B}_2^c$ . A mixed polarisation is one where agents in  $N_1$  are conformists ( $\overline{B}_1 = N$ ) and agents in  $N_2$  are anti-conformists ( $\underline{B}_2 = \emptyset$ ). (Grabisch et al., 2019) investigates the dynamics of mixed polarisations. Generally speaking, mixed polarisations are prone to cycling opinions.

(L<sub>1</sub>) *Leadership polarisations.*  $\mathcal{V}_L := \begin{pmatrix} + & - \\ + & - \end{pmatrix}$

$N_1 \subseteq \overline{B}_1$ ,  $N_2 \subseteq \overline{B}_1^c$ ,  $N_1 \subseteq \overline{B}_2$  and  $N_2 \subseteq \overline{B}_2^c$ . Hence,  $\overline{B}_1 = \overline{B}_2 = N_1$ : a partition into a single group is possible.

A leadership polarisation is such that  $\bigcap_{i \in N} \mathcal{M}_i \neq \emptyset$ . In particular, the society can be partitioned into a single group. Taking  $B \in \bigcap_{i \in N} \mathcal{M}_i$ , the society can be split into  $B$  and  $B^c$ : those who are unanimously followed ( $B$ ), those who are unanimously disesteemed, including by themselves ( $B^c$ ).

Writing  $\bigcap_{i \in N} \mathcal{M}_i = [\underline{B}, \overline{B}]$ , where  $\underline{B} := \bigcup_{i=1}^n \underline{B}_i$  and  $\overline{B} := \bigcap_{i=1}^n \overline{B}_i$ ,

we have  $\underline{B} \cap \bigcup_{i=1}^n \overline{B}_i^c = \emptyset$  and  $\overline{B}^c \cap \bigcup_{i=1}^n \underline{B}_i = \emptyset$ , from which we infer that agents in  $\underline{B}$  are

weighted strictly negatively by no agents, and agents outside  $\overline{B}$  are weighted strictly positively by no agents. Despite that the conformist polarisation is a particular leadership polarisation, the dynamics and absorbing classes of the leadership polarisation are not a simple transposition of the absorbing classes and dynamics of the conformist polarisation.

Generally speaking, leadership polarisations are prone to cycling opinions.

Polarisations  $M_2$ ,  $L_2$  and  $R_2$  are defined by exchanging the roles of  $N_1$  and  $N_2$ . If for  $M_2$  and  $L_2$ , this does not change the type of polarisation,  $R_2$  can be considered as a new type which we may call ‘anti-communitarianism’: agents weight positively agents outside their group, and negatively agents of their group.

Example 11 analyses the four polarisations under majority influence.

**Example 11.** (Majority influence) We assume that the society can be partitioned into two groups  $N_1$  and  $N_2$ , that all agents are unbiased and that they aggregate according to distance-based aggregation functions. For a conformist agent, this corresponds to following the majority. For the other agents, the aggregation is also based on majority, but the opinions are positively or negatively considered. Our assumptions impose  $\underline{B}_i = \overline{B}_i$  for all  $i \in N$ , i.e., the monotonicity collections of any  $i$  is of the form  $\mathcal{M}_i = \{B_i\}$  where: (i)  $B_i = N$  for conformist agents; (ii)  $B_i = \emptyset$  for anti-conformist agents; (iii)  $B_i = N_{(i)}$  for communitarian agents. Distance-based aggregation functions under majority influence (taking  $n$  odd), translate into  $\mathcal{P}_i^1 = \mathcal{V}^{[n/2]}(B_i)$  and  $\mathcal{P}_i^0 = \mathcal{V}^{[n/2]}(B_i^c)$ . In Table 3 (Appendix B) are summarised the transitions from  $\emptyset, N_1, N_2$  and  $N$  in each of the four polarisations. Without much surprise, cycles exist and there is no ‘continuity’ of absorbing classes with respect to  $\frac{n_1}{n_1+n_2}$ . For example, a leadership polarisation  $L_1$  ( $N_1 \in \mathcal{M}_i$  for all  $i$ ), under Assumption BC, has transitions  $N_2 \xrightarrow{1} \emptyset \xrightarrow{1} \emptyset$  when  $n_1 > n_2$ , and  $N_2 \xrightarrow{1} \emptyset \xrightarrow{1} N \xrightarrow{1} \emptyset$  when  $n_1 < n_2$ . That is, a small change in the cardinalities of the groups can provoke a jump from an absorbing state to a cycle.

We now investigate isomorphisms between the four important classes of polarisations, C, M, R and L. To do this, we introduce the following transformations on the weighted transition graphs:

**Definition 14.** Consider the transition  $S \xrightarrow{p} T$  and  $Z \subseteq N$ .

- The  $\Delta(Z)$ -right transformation of the transition  $S \xrightarrow{p} T$  is defined by  $S \xrightarrow{p} T\Delta Z$ .
- The  $\Delta(Z)$ -left transformation of the transition  $S \xrightarrow{p} T$  is defined by  $S\Delta Z \xrightarrow{p} T$ .
- The  $\Delta(Z)$ -left-right transformation of the transition  $S \xrightarrow{p} T$  is defined by  $S\Delta Z \xrightarrow{p} T\Delta Z$ .

A vector of aggregation functions  $\mathbf{A}'$  is said to be obtained from  $\mathbf{A}$  by a right/left/left-right transformation when a right/left/left-right transformation is operated on all transitions of  $\Gamma^{\mathbf{A}}$  to produce  $\Gamma^{\mathbf{A}'}$ .

We present here three related propositions in a row. Theorem 7 expresses the monotonicity collections obtained from the  $\Delta$  transformations. Corollary 1 applies this theorem to the four particular polarisations we are particularly interested in. Finally, Proposition 2 focuses on the absorbing classes obtained from a left-right transformation; in particular, switching from the absorbing classes of C to the ones of R, or from the ones of M to the ones of L (and vice versa), is extremely simple.

**Theorem 7.** ( $\mathcal{M}_i$  and  $\Delta$  transformations on transition graphs)

Assume that  $\mathbf{A}'$  is obtained from  $\mathbf{A}$  by :

- (i) a  $\Delta(Z)$ -left transformation.

Then,  $\mathcal{M}_i^{\mathbf{A}'} = (\mathcal{M}_i^{\mathbf{A}})^{\Delta(Z)}$  for all  $i \in N$ .

(ii) a  $\Delta(Z)$ -right transformation.

Then:  $\mathcal{M}_i^{\mathbf{A}'} = \mathcal{M}_i^{\mathbf{A}}$  for all  $i \notin Z$ , and  $\mathcal{M}_i^{\mathbf{A}'} = (\mathcal{M}_i^{\mathbf{A}})^{(c)}$  for all  $i \in Z$ .

(iii) a  $\Delta(Z)$ -left-right transformation.

Then:  $\mathcal{M}_i^{\mathbf{A}'} = (\mathcal{M}_i^{\mathbf{A}})^{\Delta(Z)}$  for all  $i \notin Z$ , and  $\mathcal{M}_i^{\mathbf{A}'} = (\mathcal{M}_i^{\mathbf{A}})^{\Delta(Z^c)}$  for all  $i \in Z$ .

**Corollary 1.** (Relations between the polarisations C, M, L and R) Assume that the society is split into the groups  $N_1$  and  $N_2$ .

(i) Let  $\mathbf{A}'$  be the aggregation function obtained from the aggregation function  $\mathbf{A}$  with the  $\Delta(N_2)$ -right transformation. Then:

- (a) The polarisation of  $\mathbf{A}$  is conformist if and only if the polarisation of  $\mathbf{A}'$  is mixed.
- (b) The polarisation of  $\mathbf{A}$  is leadership if and only if the polarisation  $\mathbf{A}'$  is communitarian.

(ii) Let  $\mathbf{A}'$  be the aggregation function obtained from the aggregation function  $\mathbf{A}$  with the  $\Delta(N_2)$ -left-right transformation. Then:

- (a) The polarisation of  $\mathbf{A}$  is conformist if and only if the polarisation of  $\mathbf{A}'$  is communitarian.
- (b) The polarisation of  $\mathbf{A}$  is leadership if and only if the polarisation of  $\mathbf{A}'$  is mixed.

(iii) Let  $\mathbf{A}'$  be the aggregation function obtained from the aggregation function  $\mathbf{A}$  with the  $\Delta(N_2)$ -left transformation. Then:

- (a) The polarisation of  $\mathbf{A}$  is conformist if and only if the polarisation of  $\mathbf{A}'$  is leadership.
- (b) The polarisation of  $\mathbf{A}$  is mixed if and only if the polarisation  $\mathbf{A}'$  is communitarian.

**Proposition 2.** (Left-right transformation)

Assume  $\mathbf{A}'$  to be the aggregation function obtained from  $\mathbf{A}$  by a  $\Delta(Z)$ -left-right transformation of its weighted transition graph. Then  $\{S_1, \dots, S_p\}$  is an absorbing class of  $\mathbf{A}$  if and only if  $\{S_1\Delta Z, \dots, S_p\Delta Z\}$  is an absorbing class of  $\mathbf{A}'$ . If the polarisation of  $\mathbf{A}$  is conformist or communitarian, then there exists no periodic class of length  $k > \binom{n}{\lfloor n/2 \rfloor}$ .

It follows from Corollary 1.(ii) and the first part of Proposition 2 that the dynamics of communitarian and the conformist polarisations are the same and that the dynamics of leadership and the mixed polarisations are the same. Therefore, the opinions in communitarian polarisations, like in the conformist polarisations, tend to reach an absorbing state, while the opinions in leadership and mixed polarisations tend to end up into cycles.

Moreover, not only the conformist and the communitarian polarisations are prone to reaching absorbing states, but also their cycles cannot be too long, as stated in the second part of Proposition 2. The simple transformation of absorbing classes obtained in Proposition 2 is specific to left-right transformations; in general, even though weighted transition graphs are isomorphic, absorbing classes are not. To see this, consider Proposition 2, which states that lengths of cycles in polarisations C and R do not exceed  $\binom{n}{\lfloor n/2 \rfloor}$ ; however we can show that there exists leadership and mixed societies with cycles which exceed  $\binom{n}{\lfloor n/2 \rfloor}$ . Therefore, C and M on the one hand, and L and R on the other hand, are polarisations whose dynamics belong to different universes.

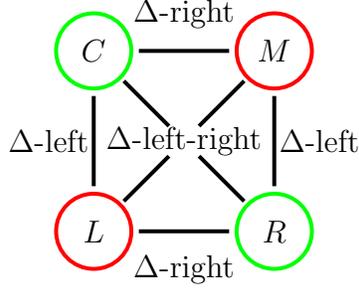


Figure 8: Corollary 1 and Proposition 2: polarisations C and R on the one hand, and M and L on the other hand, have the same dynamics.

## 4 Stability

We now investigate conditions under which opinions become stable. We refer to opinions as being *stable* when an absorbing state is reached. Admittedly, this definition of stability is strong. We could also define measures of stability. For example, we could investigate the invariants  $\mathcal{I}(\mathcal{C})$  of an absorbing class  $\mathcal{C}$ , defined by:

$$I(\mathcal{C}) := \left( \bigcap_{S \in \mathcal{C}} S \right) \cup \left( \bigcup_{S \in \mathcal{C}} S \right)^c$$

which is the set of agents who never change their opinion in  $\mathcal{C}$ . Another approach to stability could be to compute the average cardinality of states in  $\mathcal{C}$ , and its variance. Finally, we could also introduce a concept of asymptotic stability, where an absorbing state  $S$  would be asymptotically stable when for any initial state  $S_0$  close enough to  $S$ , that is, such that  $|S \Delta S_0| \leq p$  for some  $p$ , opinions starting from state  $S_0$  would end up in  $S$  with probability 1.

In social choice theory, the aggregation rules allowing for cycles (e.g., the Condorcet rule) are regarded as undesirable. In our context, the societies where opinions do not reach an absorbing state, are such that the outcome of the vote depends on the voting day. This remark motivates investigating influence process which does not converge towards an absorbing state, i.e., by Theorem 1, reaching a cycle or an interval collection. In what follows, we focus on cycles and we leave aside the absorbing classes of the third kind, whose analysis is more complex.

At this level of generality, delivering necessary and sufficient conditions to forbid cycles seems to be very complex. We only propose a few sufficient conditions.

**Proposition 3.** (Stability of opinions) If one of the following statements holds, then there is no cycle.

- (i) The society is GWM and Assumption SI holds.
- (ii) Assumption SI holds and there exists  $i \in N$  such that  $\mathcal{Y}_i = \mathcal{N}_i = \{\{1\}, \dots, \{n\}\}$ .
- (iii) All aggregation functions are Boolean and for all  $i$ :  $\mathcal{P}_i^1 = \mathcal{V}^p(N_{(i)})$ , where  $p < \min_{n_j \neq n_l} |n_j - n_l|$ .
- (iv) The polarisation is conformist or communitarian, and all agents have distance-based aggregation functions.

- (v) Each  $\mathcal{P}_i^*$  can be partitioned into  $\mathcal{P}_i^* = \mathcal{P}_i \cup \mathcal{P}'_i$  such that the vector of Boolean aggregation functions  $\mathbf{A}'$  verifies one of the previous statements, where  $A'_i$  is determined by  $\mathcal{P}_i^{\prime 1} := \mathcal{P}_i^1 \cup \mathcal{P}_i$  and  $\mathcal{P}_i^{\prime 0} := \mathcal{P}_i^0 \cup \mathcal{P}'_i$ .

**Corollary 2.** If all aggregation functions are Boolean and if one of the statements of Proposition 3 holds, then the society converges towards an absorbing state.

*Remark 2.* Assumption SI, or its weaker version ( $\phi(i \rightarrow i) \geq 0$  for all  $i$ ), is involved in all points (i) to (iv), either explicitly or implicitly. Though it is not a necessary condition for reaching an absorbing state, it is difficult to bypass it when designing sufficient conditions.

## 5 Entropy

### 5.1 Motivation

At this stage of the paper,  $(\mathcal{M}_i)_{i \in N}$  does not give a clue, neither about the absorbing classes, nor about which of the three dynamics described by Theorem 1 will be reached. Conversely, we cannot extract the monotonicity collections from the absorbing classes. For example, we know that cycles are possible in a conformist polarisation (Grabisch and Rusinowska, 2013). However, this is a rather exceptional and quite atypical situation, as a conformist polarisation leads most of the time to convergence in an absorbing state. This suggests to redirect the focus toward a statistical approach which would allow us to say, for example, that absorbing states are more natural than cycles in a conformist model. In short, we would like to define a notion of natural dynamics.

We expect that natural dynamics could help for the prediction of opinion. If indeed we can see a few transitions (typically, time series, pools of opinions), we would like to be able to say something like: ‘These opinions must have been produced by a society made of 12% of anti-conformist agents’ or: ‘The society should be composed of six communitarian groups of this and that size’, and then use this guess to predict the future opinions. This is what econometricians do when computing the best fit for their models. In the same spirit, we introduce in this section a measure of disorder or abnormality, called entropy, which we use to extract information on the polarisation of the society and to predict future opinions.

Since our question is situated at the observation level, the problem of counterfactuals arises. Observing a transition from state  $S$  to state  $T$  does not give  $b_{S,T}$ . To dodge this issue and focus our analysis on the mechanism we are interested in, we make the following assumption:

**Assumption.** In Section 5, we assume that all aggregation functions are Boolean and non constant.

Therefore, all transitions are deterministic.

### 5.2 Definition and properties

Consider a transition  $S \xrightarrow{1} T$  and assume that  $\phi(i \rightarrow j) > 0$ . Then, we expect  $j$  to adopt  $i$ ’s opinion, in the sense that, if for example  $i \in S$ , then  $j \in T$ . By the same token, if  $i \notin S$  and  $\phi(i \rightarrow j) > 0$ , we expect  $j$  to follow  $i$ ’s opinion, and therefore, that  $j$  will say no at the next period: we expect  $j \notin T$ . These implications actually fail in general, but we consider these failures as some kind of abnormality, hence our definition of entropy as the number of abnormalities over the  $n^2$  pairs  $(i, j)$ :

**Definition 15.** (Entropy of a sure transition)

Assume that  $S \xrightarrow{1} T$ . Let  $\mathbf{m}_{S \xrightarrow{1} T}$  be the Boolean function on  $N \times N$  defined by  $\mathbf{m}_{S \xrightarrow{1} T}(i, j) = 1$  (an ‘abnormality’) if and only if one of the following assertions holds: (i)  $\phi(i \rightarrow j) > 0$  and  $i \in S$  and  $j \notin T$ ; (ii)  $\phi(i \rightarrow j) > 0$  and  $i \notin S$  and  $j \in T$ ; (iii) (i)  $\phi(i \rightarrow j) < 0$  and  $i \in S$  and  $j \in T$ ; (iv)  $\phi(i \rightarrow j) < 0$  and  $i \notin S$  and  $j \notin T$ .

The *entropy* of  $S \xrightarrow{1} T$  is defined by:  $\mu(S \xrightarrow{1} T) := \sum_{i,j} \mathbf{m}_{S \xrightarrow{1} T}(i, j)$ .

The *entropy of a cycle*<sup>9</sup>  $\mathcal{C}: S_1 \xrightarrow{1} \dots \xrightarrow{1} S_k \xrightarrow{1} S_1$ , is defined by  $\mu(\mathcal{C}) := \frac{1}{k} \sum_{p=1}^k \mu(S_p \xrightarrow{1} S_{p+1})$ .

The *normalised entropy* of  $\mu$  is  $\mu_0 := \mu/n^2$ .

Since transitions are deterministic, as it is assumed in this section, we can unambiguously write  $\mathbf{m}_S$  and  $\mu(S)$  instead of  $\mathbf{m}_{S \xrightarrow{1} T}$  and  $\mu(S \xrightarrow{1} T)$  and refer to entropy as being defined on states<sup>10</sup>.

In information theory, entropy accounts for the degree of unpredictability (Shannon, 1948). In thermodynamics, the interpretation of entropy in terms of order and disorder traces back to Clausius and Helmholtz (Clausius (1879), Anderson (2005)). Being a count of abnormalities, our terminology is in accordance with this idea. As a consequence, entropy measures how consistent, and therefore how informative, is a transition about the underlying society.

Example 12 illustrates this idea. Proposition 6 shows that entropy has nothing to do with signs of influences, since entropy is left unchanged under some transformations which affect the signs of influence. When applying a  $\Delta$ -transformation on the influence graph, *both* the transitions and the monotonicity collections are transformed. As a consequence,  $\mathbf{m}$  (and therefore  $\mu$ ) remains unchanged.

**Example 12.** In conformist societies, entropy is null at the trivial absorbing states  $\emptyset$  and  $N$ . Unanimity is consistent with the idea we have about conformism. In communitarian societies with two groups  $N_1$  and  $N_2$ , the entropy is null at states  $N_1$  and  $N_2$ . In this context, the dichotomy of opinions is some kind of collective success.

<sup>9</sup>The definition of entropy on a collection of transitions as a sum over the transitions is to avoid combinatorial complexity.

<sup>10</sup>Contrary to physics where a ‘photography of the system’ suffices to see in which direction it will evaluate, observing a state  $S$  without any knowledge on  $\mathbf{A}$  does not give a clue about the evolution of opinions, hence the definition of entropy on transitions. The entropy on non-deterministic transitions (which, as announced, is not investigated in this paper) can be defined as follows:

$$\mu(S \rightarrow T) := \sum_{i,j} f(b_{S,T}) \mathbf{m}_{S,T}(i, j) = \sum_{\substack{i,j \\ \mathbf{m}(S \rightarrow T, i, j) = 1}} f(b_{S,T})$$

where  $f$  is an increasing function on  $]0, 1]$ . We suspect that  $f = \log$  would be the most appropriate: (i) this function is consistent with the fact that  $S \xrightarrow{0} T$  is not observable; (ii) since  $\log(1) = 0$ , we cannot use the non-deterministic entropy index in the deterministic case, which is consistent with the fact that entropy in information theory and statistical thermodynamics are not formulated in terms of threshold models; (iii) a low  $b_{S,T}$  is consistent with an abnormality  $\mathbf{m}_{S \rightarrow T}(i, j) = 1$  and a high  $b_{S,T}$  is consistent with  $\mathbf{m}_{S \rightarrow T}(i, j) = 0$ , but the former (a combination of two abnormal events) is more informative than the latter (a combination of two normal events). Hence, a concave function is more appropriate; (iv)  $\log(b_{S,T}) = \sum_{i \in T} \log(p_i(S)) + \sum_{i \notin T} \log(1 - p_i(S))$  obtains an expression which recalls

the Shannon entropy and the Gibbs entropy formula, where the sum of the  $-p_i \log(p_i)$  is replaced by the sum of the  $\log(p_i)$  on the domain of abnormalities.

**Property 6.** (Invariance of entropy under a  $\Delta$ -transformation)

Let  $\mathbf{A}'$  be obtained from  $\mathbf{A}$  by a  $\Delta(Z)$ -transformation, for some  $Z \subseteq N$ . Then,  $\mathbf{m}_{S'}^{\mathbf{A}'}(i, j) = \mathbf{m}_S^{\mathbf{A}}(i, j)$  for all  $S, i, j$ , where  $S' = S$  for the right transformation, and  $S' = S\Delta Z$  for a left or left-right transformation.

Since in the next subsection we intend to use a minimum entropy principle, it is useful to have an idea on the range of the values taken by  $\mu$ . First, note from the definition that the more irrelevant agents are, the lower the entropy is. For this reason, models with distance-based aggregation functions (complete networks of influences, in particular, anonymous aggregation functions) typically exhibit high entropy transitions, while models with interval collections for  $\mathcal{P}_i^0$  and  $\mathcal{P}_i^1$  typically exhibit low entropy transitions. When there are no irrelevant agents, simulations show (like in Examples ?? and 15) that the normalised entropy in complete graphs of influence can be around  $\frac{1}{3}$  or more. We may wonder if there exist transitions with no abnormalities, or, on the contrary, transitions with  $n^2$  abnormalities. Theorem 8 answers ‘yes’ to the first question and ‘no’ to the second one.

**Theorem 8.** (Extremal values of  $\mu$ )

- (1)  $\mu(S) = 0$  for all  $S \in 2^N$  if and only if each agent has a boss. In this case, we have for each  $i \in N$ :

$$\underline{B}_i = \bigcap_{\substack{S \xrightarrow{1} T \\ T \ni i}} S = \bigcap_{\substack{S \xrightarrow{1} T \\ T \not\ni i}} S^c \quad (24)$$

$$\overline{B}_i = \bigcup_{\substack{S \xrightarrow{1} T \\ T \ni i}} S = \bigcup_{\substack{S \xrightarrow{1} T \\ T \not\ni i}} S^c \quad (25)$$

- (2) There exists no transition  $S \xrightarrow{1} T$  such that  $\mu(S) = n^2$ . In other words, for any transition  $S \xrightarrow{1} T$ , at least one of the following statements holds:

- There exist  $j \in S$  and  $i \in T$  such that  $\phi(j \rightarrow i) > 0$ .
- There exist  $j \in S^c$  and  $i \in T^c$  such that  $\phi(j \rightarrow i) > 0$ .
- There exist  $j \in S$  and  $i \in T^c$  such that  $\phi(j \rightarrow i) < 0$ .
- There exist  $j \in S^c$  and  $i \in T$  such that  $\phi(j \rightarrow i) < 0$ .

### 5.3 The minimal entropy principle and its applications

Let us now show how to use the entropy index in practical problems. From the observation of a collection  $\mathcal{O}$  of one or more transitions, we want to extract some information about the society. For example, we might want to choose which polarisation is the most likely:  $D_1$  or  $D_2$ . The *minimal entropy principle* is the following:  $D_1$  is more consistent with the observations  $\mathcal{O}$  than  $D_2$  if and only if  $\mu_{D_1}(\mathcal{O}) < \mu_{D_2}(\mathcal{O})$ , where the subscript  $D_i$  in  $\mu$  indicates that the entropy has been computed under polarisation  $D_i$ .

We assume that the observer only observes absorbing classes. Assuming that absorbing classes are associated with a minimal entropy is not a mathematical fact, but this assumption boils down to assuming that the society observed is a society in the strong meaning of the term, that is, contains agents cooperating in a project. In other words, agents who consent to live together have the common goal of minimising the number of abnormalities, collaborating to clean out the chaos.

### 5.3.1 Natural dynamics

In Example 13 we compare the trivial absorbing states in the conformist model which are  $\emptyset$  and  $N$ , with the other ones. Intuitively, the trivial absorbing states are more expected to be observed than non-trivial ones. The entropy principle justifies this intuition and leads to the concept of natural dynamics, which will be further illustrated in Example 14. In this example, we will justify in particular the general statements given in Section 3.2. In both examples, we take the simplest shape of aggregation functions: the majority influence (see Example 11).

**Example 13.** (Anonymous influence in a conformist model)

Assume that the society is conformist and that all aggregation functions are distance-based (this is to ensure that there exists no irrelevant agent). Then for any absorbing state  $S \in 2^N$ :

$$\mu(S) = \underbrace{0.0}_{S \times S} + \underbrace{s.(n-s)}_{S \times S^c} + \underbrace{(n-s).s}_{S^c \times S} + \underbrace{0.0}_{S^c \times S^c} = 2s(n-s) \quad (26)$$

which is null if and only if  $S \in \{\emptyset, N\}$ . Hence, non trivial absorbing states in the conformist model lead to higher entropy than the trivial ones. Unanimity is more natural than dichotomy.

Let us now examine the case of cycles, taking one of length 2:  $S_1 \xrightarrow{1} S_2 \xrightarrow{1} S_1$ . Then we can show that the entropy of this cycle is:

$$\mu(S_1 \xrightarrow{1} S_2 \xrightarrow{1} S_1) = 2(|S_1 \times S_2^c| + |S_1^c \times S_2|) > 0$$

where  $\times$  is the cartesian product of two sets. For  $S_1 = S_2$ , we recover the computations for the absorbing state.

Generally speaking, agents of a given group have no reason to end up adopting the same opinion, but they tend to do so when the entropy of the society is low. To see why, assume that the group considered is conformist, which by Proposition 6 induces no loss of generality. These agents exert positive influence on each other, and therefore can be considered as forming a conformist sub-society. We know by Grabisch and Rusinowska (2013) that such societies can have strange opinion dynamics like cycles, and non-trivial absorbing states and classes. However, by applying the minimum entropy principle, we can see that these cases are eliminated as having high entropy values. The natural dynamics is to converge in one of the trivial absorbing states. Therefore, in general, it is natural for agents of the same group to adopt similar opinions.

Example 14 pursues this investigation by a converse approach. While Example 13 attributed a natural absorbing class to a given polarisation, Example 14 attributes a natural polarisation to a given absorbing class.

**Example 14.** (Natural dynamics in the majority influence model under the ‘four polarisations’)

Let us continue Example 11, whose absorbing classes are all summarised in Table 3 of Appendix B. Without loss of generality, consider  $n_1 > n_2$ . Under the assumptions of Example 11, we can show that  $\mu(S) = \mu(S^c)$ . Hence, we can restrict ourselves to examining the absorbing states  $\emptyset \xrightarrow{1} \emptyset$  and  $N_1 \xrightarrow{1} N_1$ , and the cycle  $\emptyset \xrightarrow{1} N_1 \xrightarrow{1} \emptyset$ . By Property 6, with a left-right-transformation,  $\mu_C(S) = \mu_{R_2}(S \Delta N_1)$  and  $\mu_{L_2}(S) = \mu_{M_2}(S \Delta N_1)$ . Therefore, comparing  $N_1 \xrightarrow{1} N_1$  in  $(M_1)$  and  $(R_1)$ , is like comparing  $\emptyset \xrightarrow{1} \emptyset$  in  $(C)$  and  $(L_1)$ . Finally, the following points (i) and (ii) suffice to make an exhaustive analysis of the majority influence case with two groups and the four polarisations:  $(C)$ ,  $(L)$ ,  $(M)$  and  $(R)$ .

- (i) Assume that the absorbing state  $\emptyset \xrightarrow{1} \emptyset$  is observed. This transition can occur in the conformist polarisation (C) and the leadership polarisation ( $L_1$ ). Which society is the most likely: (C) or ( $L_1$ )?

- $\mu_C(\emptyset) = 0$ ,
- $\mu_{L_1}(\emptyset) = n_1 n_2$ .

The observer will infer from the entropy principle that the most likely society is a conformist one. More generally, an absorbing state is more likely, either in a conformist or a communitarian society, than in a leadership or a mixed society.

- (ii) Assume that the cycle  $N_1 \xrightarrow{1} \emptyset \xrightarrow{1} N_1$  is observed, which can occur in the communitarian polarisation ( $R_2$ ), i.e., an ‘anti-communitarian society’, and the mixed polarisation ( $M_2$ ), i.e., where agents in  $N_1$  are anti-conformists. Which society is the most likely: ( $R_2$ ) or ( $M_2$ )?

- $\mu_{R_2}(\emptyset \xrightarrow{1} N_1 \xrightarrow{1} \emptyset) = n n_2 + (n_1^2 + n_2^2)$ ,
- $\mu_{M_2}(\emptyset \xrightarrow{1} N_1 \xrightarrow{1} \emptyset) = 0 + 2n_1 n_2$ .

The observer will infer from the comparison of entropy values that the most likely society is a mixed one. More generally, a cycle is more likely, either in a leadership or a mixed society, than in a conformist or a communitarian society.

### 5.3.2 Information extraction and opinions prediction

Information extraction and opinion prediction are tightly related to each other. Typically, a statistician, by a wise guess, opts for, e.g., a Poisson distribution to model a phenomenon and computes which parameter fits best the data by a maximum likelihood estimation. An econometrician, by a wise guess, chooses a set of explanatory variables and a regression provides the most suitable coefficients. Entropy will be used exactly in the same manner. By a wise guess, the observer first chooses the most suitable polarisation of the society, and then determines the most suitable characteristics of the groups by minimising the entropy. Here are two examples.

**Example 15.** (Opinion prediction in a mixed polarisation)

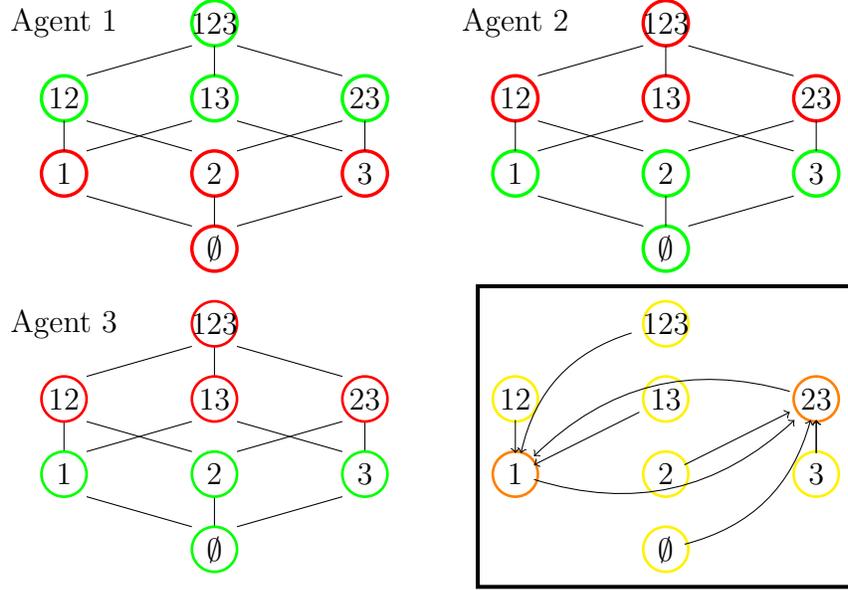
Let the society be the one of Figure 9. The observer does not know the aggregation functions of agents, he only observes the transition  $1 \xrightarrow{1} 23$ . The next transition is  $23 \xrightarrow{1} 1$ , but this is also unknown to the observer. This is what he will try to guess, using the minimal entropy principle.

The assumptions of the observer are the following:

- (i) All aggregation functions are distance-based:  $\mathcal{P}_i^1 = \mathcal{V}^{p_i}(T_i)$  for any  $i$ , where  $T_i \in \{\emptyset, N\}$ . (Mixed society).
- (ii)  $p_i = p$  for all  $i$ .

(These assumptions are actually verified by the true society). We call  $N^c$  the set of conformist agents and  $N^a$  the set of anti-conformist agents. In the true society, we have  $N^a = 23$ . The observer computes the entropy in each of the eight candidate partitions ( $N^c, N^a$ ) of the society. We can show that:

- if  $N^a \in \{\emptyset, 12, 13\}$ , then  $\mu(1) = 5$ .
- if  $N^a \in \{1, 2, 3, 123\}$ , then  $\mu(1) = 4$ .



In this example, the mixed society produces a cycle, and this cycle is made of blocks.  
 Green states:  $A_i(1_S) = 1$ . Red states:  $A_i(1_S) = 0$ .

Figure 9: A mixed society ( $N_1 = 1$  and  $N_2 = 23$ ) with distance-based aggregation functions.

- if  $N^a = 23$ , then  $\mu(1) = 3$ .

From the minimal entropy principle, the most likely partition is  $N^a = 23$  and  $N^c = 1$  (guess is correct). Then  $p$  must be computed; since  $1 \xrightarrow{1} 23$ , the only possibility is  $p = 1$  (correct). The observer has now recovered the correct aggregation functions, from which he can deduce that the next transition is  $23 \xrightarrow{1} 1$ .

**Example 16.** (Recovering the groups) Let us present two case studies where the observer aims at recovering information from the society that produced some given transitions.

The hypotheses of the observer are the following:

- The society is communitarian. The partition into groups is  $(N_1, \dots, N_g)$  with  $n_1 \leq \dots \leq n_g$ .
- All aggregation functions are distance-based.
- Observed states are made of blocks.

Since the society is assumed to be communitarian and since (ii) excludes the presence of irrelevant agents, we can show that the entropy of any transition  $S \xrightarrow{1} T$  is:

$$\mu(S) = \sum_{j=1}^g (|(N_j \cap T) \times (S \Delta N_j)| + |(N_j \setminus T) \times (S \Delta N_j)^c|). \quad (27)$$

*Situation 1.* The absorbing state  $S \xrightarrow{1} S$  is observed.

*Situation 2.* A cycle is observed. To simplify the analysis, let us take  $S \xrightarrow{1} N \xrightarrow{1} S$  for some  $S \subset N$ .

*Solving Situation 1.* From (iii), we can write  $S = \bigcup_{j \in J} N_j$ ; that is,  $J$  is the set of indices corresponding to  $S$ . From (27) with  $T = S$ :

$$\begin{aligned}
\mu(S) &= \sum_j (|(N_j \cap S) \times (S \Delta N_j)| + |(N_j \setminus S) \times (S \Delta N_j)^c|) \\
&= \sum_{j \in J} n_j \sum_{k \in J \setminus \{j\}} n_k + \sum_{j \notin J} n_j \sum_{k \notin J \cup \{j\}} n_k \\
&= \sum_{\substack{j, k \in J \\ j \neq k}} n_j n_k + \sum_{\substack{j, k \notin J \\ j \neq k}} n_j n_k
\end{aligned} \tag{28}$$

where by convention  $n_k = 0$  if  $k \in \emptyset$ . This expression is null (and therefore minimal) if and only if  $J$  and  $J^c$  contain only one group, i.e., if we have  $g = 2$ ,  $N_1 = S$  and  $N_2 = S^c$ . This solution is intuitive: the society is simply anti-coordinating. There are two groups: one says yes, the other one says no.

*Solving Situation 2.* From (iii), we can write  $S = \bigcup_{j \in J} N_j$ . From (27) with  $S \xrightarrow{1} N$  and  $N \xrightarrow{1} S$ :

$$\begin{aligned}
\mu(S) &= \sum_{j \in J} |N_j \times (S \setminus N_j)| + \sum_{j \notin J} |N_j \times (S \cup N_j)| \\
&= \sum_{j \in J} n_j (s - n_j) + \sum_{j \notin J} n_j (s + n_j)
\end{aligned} \tag{29}$$

and:

$$\begin{aligned}
\mu(N) &= \sum_{j \in J} |N_j \times (N \setminus N_j)| + \sum_{j \notin J} |N_j \times N_j| \\
&= \sum_{j \in J} n_j (n - n_j) + \sum_{j \notin J} n_j^2
\end{aligned} \tag{30}$$

So that:

$$\begin{aligned}
\mu(\mathcal{C}) &= \sum_{j \in J} n_j (s - n_j) + \sum_{j \notin J} n_j (s + n_j) + \sum_{j \in J} n_j (n - n_j) + \sum_{j \notin J} n_j^2 \\
&= \sum_{j \in J} n_j (n + s - 2n_j) + \sum_{j \notin J} n_j (s + 2n_j)
\end{aligned} \tag{31}$$

Finding the partition that minimises  $\mu(\mathcal{C})$  is equivalent to finding the partition that minimises the following quantity:

$$\sum_{j \notin J} n_j^2 - \sum_{j \in J} n_j^2. \tag{32}$$

By the Cauchy-Schwartz inequality, we have for any partition of the society:  $\sum_{j \notin J} n_j^2 \geq n - s$  and  $\sum_{j \in J} n_j^2 \leq s^2$ , where equality holds if and only if  $J^c$  is made of  $n - s$  singletons and  $J$  is

made of a single element of size  $s$ . Replacing in (31) gives  $\mu_R^*(\mathcal{C}) = s(n - s) + (n - s)(s + 2) = 2(s + 1)(n - s)$  ( $= 2\mu(S) = 2\mu(N)$ ). The solution of this minimisation problem seems to be of bad quality: obtaining a partition made of one group equal to  $S$  and a fragmentation of  $S^c$  into singletons sounds weird. This is not surprising. As shown in Example 14, a cycle is not a natural dynamics for communitarian polarisations. Taking a communitarian polarisation as an assumption to explain a cycle was not a judicious choice.

Let us rather assume that the polarisation is a mixed one, partitioned into conformist and anti-conformist agents ( $N^c, N^a$ ), aggregating with distance-based aggregation functions. We must determine which one of the following two cases minimises  $\mu$ ; (i)  $N^a = S$ ; (ii)  $N^a = S^c$ . We report below the entropy values in both cases.

(i) If  $N^a = S$ , then  $\mu(S) = s^2 + (n - s)^2$  and  $\mu(N) = n^2$ .

(ii) If  $N^a = S^c$ , then  $\mu(S) = 2s(n - s)$  and  $\mu(N) = 0$ .

For any  $S$ , taking  $N^c = S$  and  $N^a = S^c$  minimises the entropy of the cycle  $S \xrightarrow{1} N \xrightarrow{1} S$ , equal to  $\mu_M^*(\mathcal{C}) = 2s(n - s)$ . We can also see that  $\mu_M^*(\mathcal{C}) < \mu_R^*(\mathcal{C})$  for any  $S \in 2^N \setminus \{N\}$ , which shows that choosing a mixed polarisation to explain this cycle is a more judicious choice.

## 6 Concluding remarks

In this paper we developed a tractable model of opinion diffusion under mutual influence. Any kind of diffusion can be hosted by our formalism, as long as: (i) alternatives are binary; (ii) decision making is reversible; (iii) updating is synchronous and (iv) the process is Markovian; (v) aggregation functions are fixed.

Section 2 showed that negative influence can be treated similarly as the positive one, by choosing an appropriate pre-order for each agent. The difficulty inherent to models with negative influence is not the negative influence per se but the possible heterogeneity in such pre-orders. However, the characterisation of absorbing classes is the same as in the conformist model. Little assumptions on aggregation functions were imposed, the only indispensable one being that  $\mathcal{M}_i \neq \emptyset$  for all  $i$ , which means the influences received by the agents do not qualitatively depend on the state of the world. Perhaps surprisingly, Assumption SI, stating that agents weigh positively their own opinion, does not play an important role in the formulation of this model. The absence of utility functions, replaced by aggregation functions playing the role of best responses, makes the model very handy.

In Section 3, we proposed a notion of groups which is entirely non-topological: agents belong to the same group if they qualitatively receive the same influences. We uncovered simple transformations between four important polarisations: conformism, anti-conformism, leadership and communitarianism (anti-coordination). However, agents of the same groups do not necessarily end up adopting the same opinion. This inconvenient fact motivates the introduction of a measure of abnormalities, called entropy, developed in Section 5. This index measures how consistent dynamics with aggregation functions of agents is. For example, a natural idea is that agents of the same group end up adopting the same opinion, or that there are no cycling opinions in conformist models. Though this does not hold in general, it does however in societies characterised by a low entropy. We introduced the minimal entropy principle, which can be applied to information extraction and opinion predictions. The strategy is similar to the minimal entropy principle. A given shape of the society is assumed ex-ante and the most likely parameters of the society are recovered from the observations.

A significant improvement of this model would be to study the properties of its asynchronous version. Most of the tools of this article can be exported easily: aggregation functions, yes/no influential coalitions, influence graphs, monotonicity collections, polarisations and groups. One important question is: would it be possible to obtain the dynamics of the asynchronous version of the model from the dynamics of the synchronous one? This question is essential, since the synchronous model is more tractable than the asynchronous one, while the asynchronous model is more realistic. Some dynamics, however, can be proved to be similar in both models, which suggests that the gap between the two approaches, actually, is not so wide.

## Bibliography

- R. P. Abelson. Mathematical models of the distribution of attitudes under controversy. In N. Fredericksen and H. Gullicksen, editors, *Contributions to mathematical psychology*, pages 142–160. 1964.
- G. Anderson. *Thermodynamics of Natural Systems*. Cambridge University, UK, 2005.
- K. J. Arrow. *Social Choice and Individual Values*. New York: Wiley, 1963.
- C. Asavathiratham. *Influence model: a tractable representation of networked Markov chains*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2000.
- A. V. Banerjee. A simple model of herd behavior. *Quarterly Journal of Economics*, 107(3): 797–817, 1992.
- S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100:992–1026, 1992.
- Y. Bramoullé. Anti-coordination and social interactions. *Games and Economic Behavior*, 58: 30–49, 2007.
- Y. Bramoullé, D. López-Pintado, S. Goyal, and F. Vega-Redondo. Network formation and anti-coordination games. *International Journal of Game Theory*, 33(1):1–19, 2004.
- B. Buechel, T. Hellmann, and S. Klößner. Opinion dynamics and wisdom under conformity. *Journal of Economic Dynamics and Control*, 52:240–257, 2015.
- C. Castellano, M. Muñoz, and R. Pastor-Satorras. Nonlinear q-voter model. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 80:041129, 2009.
- R. Clausius. *The Mechanical Theory of Heat*. London: MacMillan and co., 1879.
- P. Clifford and A. Sudbury. A model of spatial conflict. *Biometrika*, 60(3):581, 1973.
- M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345): 118–121, 1974.
- M. Förster, M. Grabisch, and A. Rusinowska. Anonymous social influence. *Games and Economic Behavior*, 82:621–635, 2013.
- J. French. A formal theory of social power. *Psychological Review*, 63(3):181–194, 1956.

- N. Friedkin and E. Johnsen. Social influences and opinion. *Journal of Mathematical Sociology*, 15:193–205, 1990.
- S. Galam. Majority rule, hierarchical structures, and democratic totalitarianism: A statistical approach. *Journal of Mathematical Psychology*, 30(4):426–434, 1986.
- S. Galam. Contrarian deterministic effects on opinion dynamics: the hung elections scenario. *Physica A*, 333:453–460, 2004.
- S. Galam. Sociophysics: A review of galam models. *International Journal of Modern Physics C*, 19, 2008.
- S. Galam and F. Jacobs. The role of inflexible minorities in the breaking of democratic opinion dynamics. *Physica A: Statistical Mechanics and its Applications*, 381:366–376, 2007.
- R. Girard. *Deceit, Desire and the Novel: Self and Other in Literary Structure*. Baltimore: Johns Hopkins University, 1966.
- R. Girard. *Violence and the Sacred*. Baltimore: Johns Hopkins University, 1977.
- M. Grabisch. Set functions, games and capacities in decision making. *Springer International Publishing AG*, 2016.
- M. Grabisch and F. Li. Anti-conformism in the threshold model of collective behavior. *ExSIDE Working Paper Series*, 10, 2019.
- M. Grabisch and A. Rusinowska. Iterating influence between players in a social network. *Documents de travail du Centre d’Economie de la Sorbonne 2010.89*.
- M. Grabisch and A. Rusinowska. A model of influence based on aggregation functions. *Mathematical Social Sciences*, 66:216–330, 2013.
- M. Grabisch and A. Rusinowska. Determining models of influence. *Operations Research and Decisions*, 26:69–85, 2016.
- M. Grabisch, A. Poindron, and A. Rusinowska. A model of anonymous influence with anti-conformist agents. *Journal of Economic Dynamics and Control*, 2019. Forthcoming.
- M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83: 1420–1443, 1978.
- X. Hu and S. Shapley. On authority distributions in organizations: Equilibrium. *Games and Economic Behavior*, 45:132–152, 2003a.
- X. Hu and S. Shapley. On authority distributions in organizations: Controls. *Games and Economic Behavior*, 45:153–170, 2003b.
- E. Ising. *Contribution to the Theory of Ferromagnetism*. PhD thesis, 1925.
- M. O. Jackson and E. C. Storms. Behavioral communities and the atomic structure of networks. 2018.

- J. S. Jull and M. A. Porter. Hipsters on networks: How a minority group of individuals can lead to an antiestablishment majority. *Physical Review*, E 99, 022313, 2019.
- T. M. Liggett. *Interacting Particle Systems*. Springer, 1985.
- T. M. Liggett. *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*. Springer, Heidelberg, 1999.
- D. López-Pintado. Network formation, cost sharing and anti-coordination. *International Game Theory Review*, 11(1):53–76, 2009.
- S. Morris. Contagion. *Review of Economic Studies*, 67(1):57–78, 2000.
- P. Nyczka and K. Sznajd-Weron. Anticonformity or independence? Insights from statistical physics. *Journal of Statistical Physics*, 151:174–202, 2013.
- P. Nyczka, K. Sznajd-Weron, and J. Cislo. Phase transitions in the q-voter model with two types of stochastic driving. *Physical Review*, 86, 2012.
- C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.
- K. Suzumura. *Rational choice, collective decisions, and social welfare*. Cambridge Cambridge University Press, 1983.
- K. Sznajd-Weron and J. Sznajd. Opinion evolution in closed community. *International Journal of Modern Physics*, 11(6), 2001.
- M. Taylor. Towards a mathematical theory of influence and attitude change. *Human Relations*, 21:121–139, 1968.
- J. Touboul. The hipster effect: When anticonformists all look the same. *arxiv*, 2014.
- S. M. Ulam. A collection of mathematical problems. 1960.
- J. von Neumann. *Theory of Self-Reproducing Automata*. University of Illinois Press, Champaign, IL, USA, 1966.
- D. J. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(9):5766–5771, 2002.

## Appendix A. Proofs

*Proof.* (Property 1) Inverting the identity (1). For any  $S$ ,  $p_i(S)$  can be recovered from  $\mathbf{B}$  as follows. For any  $S \in 2^N$ ,  $\mathbf{A}(1_S)$  is of the form  $\mathbf{A}(1_S) = (\mathbf{1}_T, \mathbf{x}_K, \mathbf{0}_{(T \cup K)^c})$  with  $\mathbf{x}_K \in (0, 1)^K$ . Hence, there exists  $T, K$  such that  $S \xrightarrow{1} [T, T \cup K]$ . Noticing that  $A_i(1_S) = p_i(S)$  is the probability that the next state contains  $i$ :  $p_i(S) = \sum_{\substack{T' \ni i \\ T' \in [T, T \cup K]}} b_{S, T'}$ .

(This property can be found in Grabisch and Rusinowska) □

*Proof.* (Property 2) Assume that  $f_B(S) \subseteq f_B(S')$ . Since the lattice isomorphism  $f_B$  is an involution, this is equivalent to  $S \preceq_B S'$ . Since  $B \in \mathcal{M}_i$ ,  $A_i(1_S) \leq A_i(1_{S'})$ . That is,  $A'_i(1_{f_B(S)}) \leq A'_i(1_{f_B(S')})$ . Since  $f_B(2^N) = 2^N$ , we have proved that  $S \subseteq S' \Rightarrow A'_i(1_S) \leq A'_i(1_{S'})$ , which implies that  $N \in \mathcal{M}'_i$ .  $\square$

*Proof.* (Property 3)

( $\Rightarrow$ ) Since  $\mathcal{M}_i \neq \emptyset$ , there exists  $B \in \mathcal{M}_i$ . Let  $j \in N$ . Either  $j \notin B$ , and then  $A_i(1_{S \cup j}) \leq A_i(1_S)$  for all  $S \in 2^N$ , or  $j \in B$  and then,  $A_i(1_S) \leq A_i(1_{S \cup j})$  for all  $S \in 2^N$ .

( $\Leftarrow$ ) Assume that there exists  $j \in N$  such that (a) there exists  $S$  such that  $A_i(1_{S \cup j}) > A_i(1_S)$ , and that: (b) there exists  $S'$  such that:  $A_i(1_{S' \cup j}) < A_i(1_{S'})$ . If there exists  $B \in \mathcal{M}$ , then either  $j \in B$ , which contradicts (b), or  $j \notin B$ , which contradicts (a).  $\square$

*Proof.* (Theorem 2)

For any sets  $A, B \in 2^N$ , we write  $[[A, B]] := [A \cap B, A \cup B]$ .

Step 1. Assume that  $B_1, B_2 \in \mathcal{M}_i$ . Then,  $[[B_1, B_2]] \subseteq \mathcal{M}_i$ .

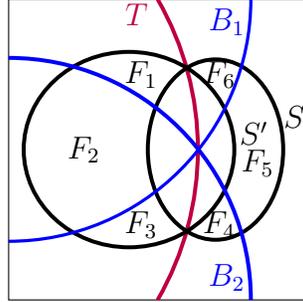


Figure 10:  $T \in [[B_1, B_2]]$

Let  $T \in [[B_1, B_2]]$ . Assume that  $S \preceq_T S'$  and let:

- $S_1 := S \cup F_1$  with  $F_1 = (S \Delta S') \cap (T \setminus B_2)$
- $S_2 := S_1 \cup F_2$  with  $F_2 = (S \Delta S') \cap (B_1 \cap B_2)$
- $S_3 := S_2 \cup F_3$  with  $F_3 = (S \Delta S') \cap (T \setminus B_1)$
- $S_4 := S_3 \setminus F_4$  with  $F_4 = (S \Delta S') \cap (B_2 \setminus T)$
- $S_5 := S_4 \setminus F_5$  with  $F_5 = (S \Delta S') \cap (B_1 \cup B_2)^c$
- $S' = S_5 \setminus F_6$  with  $F_6 = (S \Delta S') \cap (B_1 \setminus T)$

Noticing respectively that:  $F_1 \in B_1$ ;  $F_2 \in B_1$ ;  $F_3 \in B_2$ ;  $F_4 \in B_1^c$ ;  $F_5 \in B_1^c$  and  $F_6 \in B_2^c$ , we deduce the following inequalities:  $A_i(1_S) \leq A_i(1_{S_1}) \leq A_i(1_{S_2}) \leq A_i(1_{S_3}) \leq A_i(1_{S_4}) \leq A_i(1_{S_5}) \leq A_i(1_{S'})$ .

Step 2.  $\mathcal{M}_i = [\underline{B}_i, \overline{B}_i]$ .

The previous step shows that  $\mathcal{M}_i$  is an interval. From the definition of  $\underline{B}_i$  and  $\overline{B}_i$ , we actually have  $\mathcal{M}_i = [\underline{B}_i, \overline{B}_i]$ .

Step 3.  $\phi(j \rightarrow i) = 0 \Leftrightarrow [B \setminus j \in \mathcal{M}_i \Leftrightarrow B \cup j \in \mathcal{M}_i]$ .

If  $\phi(j \rightarrow i) = 0$ , then by Property 3: for all  $S \in 2^N$ :  $A_i(1_{S \setminus j}) = A_i(1_{S \cup j})$ . In particular:  $B \setminus j \in \mathcal{M}_i \Leftrightarrow B \cup j \in \mathcal{M}_i$ . Converse: Assume that  $\phi(j \rightarrow i) \neq 0$ . We want to show that

the equivalence  $[B \setminus j \in \mathcal{M}_i \Leftrightarrow B \cup j \in \mathcal{M}_i]$  does not hold. By Property 3 and without loss of generality,  $A_i(1_{S \setminus j}) \leq A_i(1_{S \cup j})$  for all  $S$ , and exists  $S^*$  such that  $A_i(1_{S^* \setminus j}) < A_i(1_{S^* \cup j})$ . Let  $B \in \mathcal{M}_i$ ; let us show that necessarily  $j \in B$  (and therefore, that there is no  $B \in \mathcal{M}_i$  such that  $B \setminus j \in \mathcal{M}_i$ ). Assume that  $j \in B^c$ . Then  $A_i(1_{S \cup j}) \leq A_i(1_{S \setminus j})$  for all  $S$ . In particular for  $S = S^*$ , then  $A_i(1_{S^* \cup j}) \leq A_i(1_{S^* \setminus j})$ , a contradiction.

Step 4. Assume that  $B_1, B_2 \in \mathcal{M}_i$ . For all  $j \in B_1 \Delta B_2$ :  $\phi(j \rightarrow i) = 0$ . Conversely, if  $\phi(j \rightarrow i) = 0$ , there exists  $B_1, B_2 \in \mathcal{M}_i$ :  $j \in B_1 \Delta B_2$ .

( $\Leftarrow$ ) Let  $B_1, B_2 \in \mathcal{M}_i$ . By the Step 1:  $[[B_1, B_2]] \subseteq \mathcal{M}_i$ . Let  $j \in (B_1 \cup B_2) \setminus (B_1 \cap B_2)$ : then for any  $B \in [[B_1, B_2]]$ :  $[B \setminus j \in \mathcal{M}_i \Leftrightarrow B \cup j \in \mathcal{M}_i]$ . Result follows from the converse direction of the previous step. ( $\Rightarrow$ ) If  $\phi(j \rightarrow i) = 0$ , then for any  $B \in \mathcal{M}_i$ , by Step 3: either  $B \ni j$ , and therefore:  $B \setminus j \in \mathcal{M}_i$ , or  $j \notin B$ , and therefore  $B \cup j \in \mathcal{M}_i$ . In both cases we have found  $B_1, B_2 \in \mathcal{M}_i$  such that  $j \in B_1 \Delta B_2$ .

Step 5.  $\phi(j \rightarrow i) = 0$  if and only if  $j \in \overline{B}_i \setminus \underline{B}_i$ .

From  $\overline{B}_i \setminus \underline{B}_i = \overline{B}_i \Delta \underline{B}_i = \{j \mid \exists B_1, B_2 \in \mathcal{M}_i : j \in B_1 \Delta B_2\}$  and the previous step.

Step 6. If there exists  $B \in \mathcal{M}_i$  such that  $j \in B$ , then  $\phi(j \rightarrow i) \geq 0$ .

Assume that  $\phi(j \rightarrow i) < 0$  and  $j \in B \in \mathcal{M}_i$ . By  $\phi(j \rightarrow i) < 0$ , there exists  $S^*$  such that  $A_i(1_{S^* \cup j}) < A_i(1_{S^*})$ . By  $j \in B \in \mathcal{M}_i$ , it holds that  $S \prec_B S \cup j$  for and all  $S$ . In particular  $S^* \prec_B S^* \cup j$ , which implies  $A_i(1_{S^*}) \leq A_i(1_{S^* \cup j})$ , a contradiction. By the previous step, the inequality is strict if and only if  $j \in \underline{B}_i$ .

□

*Proof.* (Property 5) For all  $S \in \mathcal{N}_i$  and  $j \in S$ :  $A_i(1_{\overline{B}_i \Delta S}) < A_i(1_{\overline{B}_i \Delta (S \setminus j)}) = 1$ ; therefore,  $\phi(j \rightarrow i) \neq 0$ . For all  $S \in \mathcal{Y}_i$  and  $j \in S$ :  $A_i(1_{\overline{B}_i^c \Delta S}) > A_i(1_{\overline{B}_i^c \Delta (S \setminus j)}) = 0$ ; therefore,  $\phi(j \rightarrow i) \neq 0$ . That is:

$$\bigcup \mathcal{Y}_i \cup \bigcup \mathcal{N}_i \subseteq \{j \mid \phi(j \rightarrow i) \neq 0\} = \underline{B}_i \cup \overline{B}_i^c, \quad (33)$$

where the last equality comes from Theorem 2.

Let us establish the reverse inclusion for Boolean aggregation functions. Assume that  $i$  is conformist, which, from Property 2, is without loss of generality. Let  $j$  such that  $\phi(j \rightarrow i) \neq 0$ , i.e., since the agent is conformist, that  $\phi(j \rightarrow i) > 0$ . There must exist  $S$  such that  $A_i(1_{S \cup j}) = 1$  and  $A_i(1_S) = 0$ . Therefore there exists a yes-influential coalition in  $\downarrow (S \cup j)$  that contains  $j$ , i.e., there exists  $S' \subseteq S \setminus j$  such that  $(S \cup j) \setminus S' \in \mathcal{Y}_i$ . By the same token, there exists  $S'$  such that  $(N \setminus S) \setminus S'$  is a no-influential coalition, with  $j \notin S$  and  $j \notin S'$ , that is, a no-influential coalition that contains  $j$ . Finally, equality holds in the previous inclusion.

(Moreover, we notice that  $\bigcup_{S \in \mathcal{Y}_i} S = \bigcup_{S \in \mathcal{N}_i} S = \bigcup_{S \in \mathcal{N}_i \cup \mathcal{Y}_i} S$ ).

□

*Proof.* (Theorem 3) By definition,  $\mathcal{P}_i^0$  (resp.  $\mathcal{P}_i^1$ ) uniquely determines  $\mathcal{Y}_i$  (resp.  $\mathcal{N}_i$ ). We must show, conversely, that  $\mathcal{Y}_i$  (resp.  $\mathcal{N}_i$ ) uniquely determines  $\mathcal{P}_i^0$  (resp.  $\mathcal{P}_i^1$ ). Let us show that for any  $B \in \mathcal{M}_i$ :  $\mathcal{P}_i^* \cup \mathcal{P}_i^1 = \bigcup_{Y \in \mathcal{Y}_i} \uparrow_B Y$  and  $\mathcal{P}_i^* \cup \mathcal{P}_i^0 = \bigcup_{M \in \mathcal{N}_i} \downarrow_B (N \setminus M)$ . Let  $B \in \mathcal{M}_i$ . We have

$\mathcal{P}_i^* \cup \mathcal{P}_i^1 \supseteq \bigcup_{Y \in \mathcal{Y}_i} \uparrow_B Y$ , otherwise the monotonicity conditions would be violated. Conversely,

$\mathcal{P}_i^* \cup \mathcal{P}_i^1 \subseteq \bigcup_{Y \in \mathcal{Y}_i} \uparrow_B Y$ ; indeed, if there existed  $S \notin \bigcup_{Y \in \mathcal{Y}_i} \uparrow_B Y$  such that  $A_i(1_S) > 0$ , then there

would exist  $Y \in \downarrow_B S$  such that  $Y \in \mathcal{Y}_i$ , i.e.,  $S \in \uparrow_B Y$ , which is absurd. (Similar proof for no-influential coalitions).  $\square$

*Proof.* (Theorem 4)

- (1)  $S \rightarrow T$  if and only if  $A_i(1_S) > 0$  for all  $i \in T$  and  $A_i(1_S) < 1$  for all  $i \notin T$ . For  $B_i \in \mathcal{M}_i$ , from the monotonicity of  $A_i$  with respect to the  $\preceq_{B_i}$  order and (from Assumption BC) the fact that  $A_i(1_{B_i^c}) = 0$ , we deduce that  $A_i(1_S) > 0$  if and only if there exists a yes-influential coalition  $Y$  on  $i$  with  $Y \subseteq S \Delta B_i^c$ . On the other hand, since  $A_i$  is monotonous with respect to the  $\preceq_{B_i}$  order and (from Assumption BC)  $A_i(1_{B_i}) = 1$ ; therefore,  $A_i(1_S) < 1$  if and only if there exists a no-influential coalition  $M$  on  $i$  with  $M \subseteq S \Delta B_i$ .

(It is not hard to show that Theorem 4.(1) actually recovers *all* the  $\mathcal{Y}_i$  and the  $\mathcal{N}_i$ ).

- (2)  $S \xrightarrow{1} T$  is equivalent to  $A_i(1_S) = 1$  if  $i \in T$  and  $A_i(1_S) = 0$  if  $i \notin T$ . If  $Y \subseteq S \Delta B_i^c$  (by (10), this is equivalent to  $Y \Delta B_i^c \preceq_{B_i} S$ ) is yes-influential on  $i \notin T$ ,  $0 < A_i(1_{Y \Delta B_i^c}) \leq A_i(1_S) = 0$ , a contradiction. By the same token, if  $M \subseteq S \Delta B_i$  (by (10), this is equivalent to  $M \Delta B_i \preceq_{B_i^c} S$ , i.e.,  $M \Delta B_i \succ_{B_i} S$ ) is no-influential on  $i \in T$ ,  $1 = A_i(1_S) \leq A_i(1_{M \Delta B_i}) < 1$ , a contradiction.  $\square$

*Proof.* (Proposition 1)

(1) to (4) are adapted from Proposition 2 of Grabisch and Rusinowska (2013). (5) is a direct consequence of (4) and (3). For (6), note that if there exists  $j, k$  with  $|w_j^i| \neq |w_k^i|$ , then  $A_i(1_{T \Delta j}) \neq A_i(1_{T \Delta k})$ .  $\square$

*Proof.* (Theorem 7)

We will use the following straightforward property:

**Property 7.** For any  $S_1, S_2, B, Z \in 2^N$ :

$$S_1 \preceq_B S_2 \Leftrightarrow S_1 \Delta Z \preceq_{B \Delta Z} S_2 \Delta Z \quad (\Leftrightarrow f_{Z^c}(S_1) \preceq_{f_{Z^c}(B)} f_{Z^c}(S_2)). \quad (34)$$

- (i) Let  $S_1, S_2$  be any two states in the weighted transition graph characterized by  $\mathbf{A}$ . Let  $B \in \mathcal{M}_i^{\mathbf{A}}$ . Then, in the weighted transition graph characterized by  $\mathbf{A}'$  obtained from  $\mathbf{A}$  by a left transformation:  $S_1 \Delta Z \preceq_B S_2 \Delta Z \Rightarrow A'_i(1_{S_1}) \leq A'_i(1_{S_2})$ . By Property 7,  $S_1 \preceq_{B \Delta Z} S_2 \Rightarrow A'_i(1_{S_1}) \leq A'_i(1_{S_2})$ . Therefore,  $B \Delta Z \in (\mathcal{M}_i)^{\mathbf{A}'}$ . We have established  $\mathcal{M}_i^{\mathbf{A}'} \supseteq (\mathcal{M}_i^{\mathbf{A}})^{\Delta(Z)}$ ; the left-transformation being a symmetry, the same reasoning establishes the reverse inclusion.
- (ii) For any  $S \in 2^N$ , there exists  $T, K$  such that  $S \xrightarrow{1}_{\mathbf{A}} [T, T \cup K]$ . By Property 1:

$$p_i^{\mathbf{A}}(S) = \sum_{\substack{T' \ni i \\ T' \in [T, T \cup K]}} b_{S, T'}^{\mathbf{A}} \quad (35)$$

Let  $\mathbf{A}'$  obtained by a  $\Delta(Z)$ -right transformation. By definition, for all  $S, T \in 2^N$ :  $b_{S, T}^{\mathbf{A}'} = b_{S, T \Delta Z}^{\mathbf{A}}$ . That is:

$$S \xrightarrow{1}_{\mathbf{A}'} [[T \Delta Z, (T \cup K) \Delta Z]] \quad (36)$$

with:

$$p_i^{\mathbf{A}'}(S) = \sum_{\substack{T' \ni i \\ T' \in [[T\Delta Z, (T \cup K)\Delta Z]]}} b_{S, T'}^{\mathbf{A}'}. \quad (37)$$

We write  $\mathcal{R}^{\mathbf{A}}(S, i)$  the collection of sets  $T' \in [T, T \cup K]$  containing  $i$ , for a given transition  $S \xrightarrow[\mathbf{A}]{1} [T, T \cup K]$ . We rewrite:

$$p_i^{\mathbf{A}}(S) = \sum_{T' \in \mathcal{R}^{\mathbf{A}}(S, i)} b_{S, T'}^{\mathbf{A}} \quad (38)$$

Noticing that:

$$\mathcal{R}^{\mathbf{A}'}(S, i) = \begin{cases} (\mathcal{R}^{\mathbf{A}}(S, i))^{\Delta(Z)} & \text{if } i \notin Z \\ ([T, T \cup K] \setminus \mathcal{R}^{\mathbf{A}}(S, i))^{\Delta(Z)} & \text{if } i \in Z \end{cases}, \quad (39)$$

we rewrite:

$$p_i^{\mathbf{A}'}(S) = \begin{cases} \sum_{T' \in (\mathcal{R}^{\mathbf{A}}(S, i))^{\Delta(Z)}} b_{S, T'}^{\mathbf{A}'} & \text{if } i \notin Z \\ \sum_{T' \in ([T, T \cup K] \setminus \mathcal{R}^{\mathbf{A}}(S, i))^{\Delta(Z)}} b_{S, T'}^{\mathbf{A}'} & \text{if } i \in Z \end{cases} \quad (40)$$

If  $i \notin Z$ , then  $p_i^{\mathbf{A}'} = p_i^{\mathbf{A}}$ . If  $i \in Z$ , then  $p_i^{\mathbf{A}'} = 1 - p_i^{\mathbf{A}}$ .

We get that finally for all  $i$  and  $B \in \mathcal{M}_i^{\mathbf{A}'}$ :

$$S \preceq_B S' \Rightarrow \begin{cases} p_i^{\mathbf{A}'}(S) \leq p_i^{\mathbf{A}'}(S') & \text{if } i \notin Z \\ p_i^{\mathbf{A}'}(S) \geq p_i^{\mathbf{A}'}(S') & \text{if } i \in Z \end{cases} \quad (41)$$

Therefore,  $B \in \mathcal{M}_i^{\mathbf{A}'}$  for  $i \in Z$  and  $B^c \in \mathcal{M}_i^{\mathbf{A}'}$  for  $i \notin Z$ . That is,  $\mathcal{M}_i^{\mathbf{A}'} \supseteq \mathcal{M}_i^{\mathbf{A}}$  for all  $i \notin Z$ , and  $\mathcal{M}_i^{\mathbf{A}'} \supseteq (\mathcal{M}_i^{\mathbf{A}})^{(c)}$  for all  $i \in Z$ . As in the previous point, the reverse inclusions hold too.

- (iii) The monotonicity collection of a weighted transition graph obtained from a right-left transformation stems from the right and the left ones. □

*Proof.* (Corollary 1)

It is a direct application of Theorem 7. For example, let us prove that the transformation of a polarisation  $M_1$  by a  $\Delta(N_2)$ -right transformation is a polarisation  $R_1$ .

A polarisation  $M_1$  is such that  $N \in \mathcal{M}_i$  for  $i \in N_1$  and  $\emptyset \in \mathcal{M}_i$  for  $i \in N_2$ . Apply a  $\Delta(N_2)$  right transformation on the weighted transition graph determined by  $\mathbf{A}$  (Property 1): from Theorem 7, for  $i \in N_1$ :  $N\Delta N_2 = N_1 \in \mathcal{M}_i^{\mathbf{A}'}$  and for  $i \in N_2$ :  $\emptyset\Delta N_2 = N_2 \in \mathcal{M}_i^{\mathbf{A}'}$ , i.e., the polarisation obtained is a communitarian one. □

*Proof.* (Proposition 2)

It holds that  $S \xrightarrow[\mathbf{A}]{1} T \Leftrightarrow S \xrightarrow[\mathbf{A}]{1} T$ . Let  $\mathcal{C}^{\mathbf{A}}$  be an absorbing class in  $\Gamma^{\mathbf{A}}$ . Therefore: (i)  $\mathcal{C}^{\mathbf{A}}$  is strongly connected if and only if  $(\mathcal{C}^{\mathbf{A}})^{\Delta}$  is strongly connected; (ii) there is no arc from  $S \in \mathcal{C}^{\mathbf{A}}$

to some  $T \notin \mathcal{C}^{\mathbf{A}}$  if and only if there is no arc from  $S \in (\mathcal{C}^{\mathbf{A}})^{\Delta}$  to some  $T \notin (\mathcal{C}^{\mathbf{A}})^{\Delta}$ . Therefore,  $(\mathcal{C}^{\mathbf{A}})^{\Delta}$  is an absorbing class in  $\Gamma^{\mathbf{A}'}$  ( $\Delta$  being a symmetry, the reverse inclusion holds too).

*Conformist and communitarian polarisations.* Assume that  $\mathbf{A}$  is the aggregation vector of communitarian polarisation where the two groups are  $N_1$  and  $N_2$ . Let the periodic class:  $[T_1, T_1 \cup H_1] \xrightarrow{1} [T_2, T_2 \cup H_2] \xrightarrow{1} \dots \xrightarrow{1} [T_k, T_k \cup H_k] \xrightarrow{1} [T_1, T_1 \cup H_1]$ , where some  $H_p$  are possibly empty. Let  $\mathbf{A}'$  obtained from  $\mathbf{A}$  with a  $\Delta(N_2)$ -left-right transformation. Then the polarisation of  $\mathbf{A}'$  is conformist and, by the first point and Corollary 1, absorbing classes of  $\Gamma$  and  $\Gamma'$  are linked by the relation: ' $\mathcal{C}$  is an absorbing class in  $\Gamma$  if and only if  $\mathcal{C}^{\Delta}$  is an absorbing class of  $\Gamma'$ '. Consider therefore the following periodic class in a conformist model:  $[S_1, S_1 \cup K_1] \xrightarrow{1} [S_2, S_2 \cup K_2] \xrightarrow{1} \dots \xrightarrow{1} [S_k, S_k \cup K_k] \xrightarrow{1} [S_1, S_1 \cup K_1]$ . Sperner's Theorem states that the upper bound of the length of an antichain with the set order is  $\binom{n}{\lfloor n/2 \rfloor}$ . Therefore if we show that  $S_1, S_2, \dots, S_k$  must be incomparable with the order  $\subset$ , then we will have proved that  $\binom{n}{\lfloor n/2 \rfloor}$  is an upper-bound on the length of the periodic classes. Assume that  $S_p \subset S_q$  for some  $p, q$  (the equality is impossible since sets of a periodic class must be incomparable), then  $S_p \xrightarrow{1} [S_{p+1}, S_{p+1} \cup K_{p+1}]$  and  $S_q \xrightarrow{1} [S_{q+1}, S_{q+1} \cup K_{q+1}]$  with  $S_{p+1} \subseteq S_{q+1}$  and  $K_{q+1}$  possibly empty (where actually  $S_{p+1} \subset S_{q+1}$  because the class is periodic). Iterating  $q - p$  times:  $S_q \subset S_{2q-p}$ . Let  $m$  be such that  $(m + 1)q - mp \equiv 0[k]$ . Iterating  $m$  times gets:

$$S_p \subset S_q \subset S_{2q-p} \subset \dots \subset S_{(m+1)q-mp} = S_p \quad (42)$$

which is absurd. □

*Proof.* (Proposition 3)

In all these points, we assume that  $S_1 \xrightarrow{1} S_2 \xrightarrow{1} \dots \xrightarrow{1} S_k \xrightarrow{1} S_1$  with  $k \geq 2$ .

- (i) Let  $W_0^j := \{i \mid w_i^j = 0\}$ , so that  $\phi(i \rightarrow j) = 0 \Leftrightarrow i \in W_0^j$ . Let the transition  $S_p \xrightarrow{1} S_{p+1}$ . In a GWM model (from Proposition 1.(1)+(2)), we have for all  $p$  and all  $j \in S_{p+1}$ :  $S_p \setminus W_0^j = \bigcup_{w_k^j > 0} k$  and for all  $j \notin S_{p+1}$ :  $S_p \setminus W_0^j = \bigcup_{w_k^j < 0} k$ . By assumption SI,  $w_j^j > 0$  for

all  $j \in N$ . Therefore:  $S_{p+1} \subseteq S_p \setminus W_0^j$  for all  $j \in S_{p+1}$ , which implies:  $S_{p+1} \subseteq S_p$ , where the inclusion is actually strict, since the transition is embedded into a cycle. But this holds for all  $p$ , which implies  $S_1 \subset S_1$ , a contradiction.

(This is an adaptation of Proposition 1.(iii) in Grabisch and Rusinowska (2013)).

- (ii)  $\mathcal{Y}_i = \mathcal{N}_i = \{1, \dots, n\}$  implies that  $\mathcal{M}_i$  contains a single set  $B_i$ , and that  $A_i(1_S) \in \{0, 1\}$  if and only if  $S \in \{B_i, B_i^c\}$ . There is only one candidate cycle to examine:  $B_i \xrightarrow{1} B_i^c \xrightarrow{1} B_i$ . By Assumption SI,  $i \notin B_i^c$ . Since  $B_i \xrightarrow{1} B_i^c$ , we have:  $p_i(B_i) = 0$ . Since  $\mathcal{M}_i = \{B_i\}$ , this implies  $p_i(S) = 0$  for all  $S \in 2^N$ ; but  $p_i(B_i^c) = 0$  contradicts  $B_i^c \xrightarrow{1} B_i$  and  $i \in B_i$ .

(This is an adaptation of Proposition 1.(ii) in Grabisch and Rusinowska (2013)).

- (iii) Since agents of the same group have the same aggregation function, the following cycle is made of blocks:  $S_1 \xrightarrow{1} S_2 \xrightarrow{1} \dots \xrightarrow{1} S_p \xrightarrow{1} S_1$ , and we call  $J_1, \dots, J_p$  the corresponding indices of the  $(S_q)_{q=1..p}$ . This succession of  $p$  transitions imposes the following conditions:

$$\begin{cases} S_q \in \mathcal{V}^p(N_{(i)}) & \text{for all } i \in S_{q+1} \\ S_q \notin \mathcal{V}^p(N_{(i)}) & \text{for all } i \notin S_{q+1} \end{cases} \quad (43)$$

Since  $i \in S_q$  if and only if  $(i) \in J_q$ , the previous conditions translate into:

$$\left\{ \begin{array}{l} \sum_{j \in J_q} n_j - n_{(i)} \leq p \quad \text{for all } i \text{ such that } (i) \in J_{q+1} \\ \sum_{j \in J_q} n_j + n_{(i)} > p \quad \text{for all } i \text{ such that } (i) \notin J_{q+1} \end{array} \right. \quad (44)$$

(Recall that  $(n_i)_{i \in [1, g]}$  are the cardinalities of the groups).

Hence  $S_1 \xrightarrow{1} S_2 \xrightarrow{1} \dots \xrightarrow{1} S_k \xrightarrow{1} S_1$  is a cycle if and only if for all  $i \in N$ :

$$p \in \left[ \max_{q \in [1, k]} \left[ \sum_{j \in J_q} n_j - \min_{j \in J_{q+1}} n_j \right] ; \min_{q \in [1, k]} \left[ \sum_{j \in J_q} n_j + \min_{j \notin J_{q+1}} n_j \right] - 1 \right].$$

(where the interval  $[a, b]$  is empty if  $b < a$ ).

Since for any succession of transitions distinct from  $\emptyset \xrightarrow{1} \emptyset$  (impossible in a cycle), it holds

that  $\max_{q \in [1, k]} \left[ \sum_{j \in J_q} n_j - \min_{j \in J_{q+1}} n_j \right] > 0$ , we have therefore:

$$\min_{n_j \neq n_l} |n_j - n_l| \leq \max_{q \in [1, k]} \left[ \sum_{j \in J_q} n_j - \min_{j \in J_{q+1}} n_j \right].$$

If  $p$  is strictly lower than  $\min_{n_j \neq n_l} |n_j - n_l|$ , then all cycles are impossible.

- (iv) Let us show it for conformist polarisations; by Proposition 2, this holds in communitarian polarisations too.

In a society where agents have distance-based aggregation functions,  $|S| = |S'|$  and  $S \xrightarrow{1} T$  implies  $S' \xrightarrow{1} T$ . Therefore states must be of different cardinality within a cycle; hence there exists  $S_i, S_j$  such that  $|S_i| < |S_j|$ . But this implies  $|S_{i+1}| \leq |S_{j+1}|$ , because all agents are conformist. Iterating  $k$  times and using a similar argument as in the proof of Proposition 2, we get  $|S_i| < |S_i|$ , which is absurd.

(This result was established in Förster et al. (2013), Proposition 3).

- (v) If there is no cycle in a society where the set of sure transitions is larger, then there is no cycle in the primitive society. □

*Proof.* (Corollary 2) When aggregation functions are Boolean, absorbing classes are either cycles or absorbing states. Therefore, if there are no cycles, there are only absorbing states. □

*Proof.* (Property 6) This is an application of Corollary 1. Recall that for  $\mathcal{M}_i^{\mathbf{A}} = [\underline{B}_i, \overline{B}_i]$ , we have  $(\mathcal{M}_i^{\mathbf{A}})^{(c)} = [\overline{B}_i^c, \underline{B}_i^c]$ . Moreover,  $(\mathcal{M}_i^{\mathbf{A}})^{\Delta(Z)} = [\underline{B}'_i, \overline{B}'_i]$  where:

- $\underline{B}'_i := (\underline{B}_i \setminus Z) \cup (\overline{B}_i^c \cap Z)$  and:
- $(\overline{B}'_i)^c := (\overline{B}_i^c \setminus Z) \cup (\underline{B}_i \cap Z)$ .

It holds that  $\phi(i \rightarrow j) = 0 \Leftrightarrow \phi'(i \rightarrow j) = 0$  for any of the three transformations. Let us study the case (i)  $i \in S$  and  $j \notin T$ . The cases (ii)  $i \in S$  and  $j \in T$ , (iii)  $i \notin S$  and  $j \notin T$ ; (iv)  $i \notin S$  and  $j \in T$  are analogous.

- (i) *Left-transformation* ( $S \xrightarrow{1} T$  transformed into  $S\Delta Z \xrightarrow{1} T$ ).

$$\begin{cases} i \in S \setminus Z \\ (j \notin T) \end{cases} \Rightarrow \begin{cases} \text{sign}(\phi'(i \rightarrow j)) = \text{sign}(\phi(i \rightarrow j)) \\ i \in S\Delta Z \\ (j \notin T) \end{cases} .$$

$$\begin{cases} i \in S \cap Z \\ (j \notin T) \end{cases} \Rightarrow \begin{cases} \text{sign}(\phi'(i \rightarrow j)) = -\text{sign}(\phi(i \rightarrow j)) \\ i \notin S\Delta Z \\ (j \notin T) \end{cases}$$

In both cases,  $\mathbf{m}'_{S\Delta Z}(i, j) = \mathbf{m}_S(i, j)$ .

- (ii) *Right-transformation*. ( $S \xrightarrow{1} T$  transformed into  $S \xrightarrow{1} T\Delta Z$ ).

$$\begin{cases} i \in S \\ j \in Z \setminus T \end{cases} \Rightarrow \begin{cases} \text{sign}(\phi'(i \rightarrow j)) = -\text{sign}(\phi(i \rightarrow j)) \\ i \in S \\ j \in T\Delta Z \end{cases}$$

$$\begin{cases} i \in S \\ j \in (T \cup Z)^c \end{cases} \Rightarrow \begin{cases} \text{sign}(\phi'(i \rightarrow j)) = \text{sign}(\phi(i \rightarrow j)) \\ i \in S \\ j \notin T\Delta Z \end{cases}$$

In both cases,  $\mathbf{m}'_S(i, j) = \mathbf{m}_S(i, j)$ .

- (iii) *Left-right-transformation*. Composition of a left and a right-transformation.

□

*Proof.* (Theorem 8) Some lemmas need first to be exposed.

**Lemma 1.** Assume that  $\mu(S) = 0$  for all transition  $S \xrightarrow{1} T$ .

- (a)  $\phi(j \rightarrow i) > 0$  if and only if for each transition  $S \xrightarrow{1} T$ ,  $[j \in S \Leftrightarrow i \in T]$ .
- (b)  $\phi(j \rightarrow i) < 0$  if and only if for each transition  $S \xrightarrow{1} T$ ,  $[j \notin S \Leftrightarrow i \in T]$ .

*Proof.* (Lemma 1) Let us prove (a), the point (b) being analogous. ( $\Rightarrow$ ) If  $S \xrightarrow{1} T$  with  $j \in S$  and  $i \notin T$ , then  $\mathbf{m}_S(j, i) = 1$ . If  $j \notin S$  and  $i \in T$ , then  $\mathbf{m}_S(j, i) = 1$  too. ( $\Leftarrow$ ) By contraposition; two cases:  $\phi(j \rightarrow i) = 0$  and  $\phi(j \rightarrow i) < 0$ . Assume that  $\phi(j \rightarrow i) = 0$  and let  $S \xrightarrow{1} T$  be such that  $j \in S$  and  $i \in T$ . Then we have also  $S \setminus j \xrightarrow{1} T$ . That is, the equivalence  $[j \in S \Leftrightarrow i \in T]$  does not hold. Assume that  $\phi(j \rightarrow i) < 0$ . Since  $\mathbf{m}_S(j, i) = 0$  for all  $S \xrightarrow{1} T$ ,  $i$  and  $j$ , then necessarily either  $j \in S$  and  $i \notin T$ , or  $j \notin S$  and  $i \in T$ ; the equivalence  $[j \in S \Leftrightarrow i \in T]$  does not hold either. □

**Lemma 2.** If  $\mathcal{P}_i^1$  is a interval, then  $\mathcal{Y}_i = \{\underline{B}_i \cup \overline{B}_i^c\}$  and  $\mathcal{N}_i = \{\{j\} \mid j \in \underline{B}_i \cup \overline{B}_i^c\}$ .

If  $\mathcal{P}_i^0$  is a interval, then  $\mathcal{Y}_i = \{\{j\} \mid j \in \underline{B}_i \cup \overline{B}_i^c\}$  and  $\mathcal{N}_i = \{\underline{B}_i \cup \overline{B}_i^c\}$ .

*Proof.* (Lemma 2) Let us prove it for  $\mathcal{P}^1$  being an interval; the results are reversed for  $\mathcal{P}^0$ .

- $S \in \mathcal{Y}_i$  if and only if  $A_i(1_{\overline{B}_i^c \Delta S}) > 0$  (equal to 1 since  $A_i$  is Boolean) and  $A_i(1_{\overline{B}_i^c \Delta S'}) = 0$  for all  $S' \subset S$ . Since  $\mathcal{P}_i^1$  is an interval,  $\mathcal{P}_i^1 = \mathcal{M}_i$ . Therefore,  $A_i(1_T) = 1 \Leftrightarrow \underline{B}_i \subseteq T \subseteq \overline{B}_i$ . This implies  $S = \overline{B}_i^c \Delta \underline{B}_i = \overline{B}_i^c \cup \underline{B}_i$ .
- Let  $S \in \mathcal{N}_i$ . We have  $A_i(1_{\overline{B}_i \Delta S}) = 0$  and  $A_i(1_{\overline{B}_i \Delta S'}) = 1$  for all  $S' \subset S$ . That is,  $\mathcal{Q} := [[\overline{B}_i \Delta S, \overline{B}_i]] \setminus (\overline{B}_i \Delta S) \subseteq \mathcal{P}_i^1$ . If  $|S| \geq 2$ , then there exists  $\overline{B}_i \Delta S_1, \overline{B}_i \Delta S_2$  distinct sets of  $\mathcal{Q}$  linked to  $\overline{B}_i \Delta S$  in the Hasse diagram ( $S_1, S_2 \subset S$ ). But then, the fact that  $\overline{B}_i \Delta S = (\overline{B}_i \Delta S_1) \wedge_{\overline{B}_i} (\overline{B}_i \Delta S_2)$  contradicts that  $\mathcal{P}_i^1$  is a interval. Hence we must have

$|S| = 1$ . The collection of sets of  $\mathcal{P}_i^1$  linked to  $\overline{B}_i$  in the Hasse diagram is made of the sets  $\overline{B}_i \Delta j$ , where the singleton  $j$  belongs to  $\overline{B}_i \setminus \underline{B}_i$ . We finally can write  $\mathcal{N}_i = \{\{i\} \mid i \in (\overline{B}_i \setminus \underline{B}_i)^c\} = \{\{i\} \mid i \in \overline{B}_i^c \cup \underline{B}_i\}$ .

□

**Lemma 3.** The following statements are equivalent:

- (i)  $\mathcal{P}_i^0 = (\mathcal{P}_i^1)^{(c)}$ .
- (ii)  $\mathcal{P}_i^0$  and  $\mathcal{P}_i^1$  are intervals.
- (iii)  $\mathcal{P}_i^1$  is an interval of cardinality  $2^{n-1}$ .
- (iv)  $\mathcal{P}_i^0$  is an interval of cardinality  $2^{n-1}$ .
- (v)  $\mathcal{M}_i^{(c)} = \mathcal{M}_i^c$ .
- (vi) For all  $i \in N$ ,  $|\{j \mid \phi(j \rightarrow i) \neq 0\}| = 1$ .
- (vii) For each  $i$ , there exists  $j$  such that  $j$  is a boss for  $i$ .

In particular, under one of these points,  $\mathcal{P}_i^1 = \mathcal{M}_i = (\mathcal{P}_i^0)^{(c)} = ((\mathcal{M}_i)^c)^{(c)}$ .

*Proof.* (Lemma 3) (i)  $\Rightarrow$  (ii), (iii) and (iv) are obvious. Let us prove (i)  $\Rightarrow$  (v). Since  $\mathcal{P}_i^1$  is an interval and  $A_i$  is Boolean, we have  $\mathcal{M}_i = \mathcal{P}_i^1$ . Therefore,  $(\mathcal{P}_i^0)^c = (\mathcal{P}_i^0)^{(c)}$  implies  $(\mathcal{M}_i)^c = (\mathcal{M}_i)^{(c)}$ . Let us prove (v)  $\Rightarrow$  (i). We have  $[\underline{B}_i, \overline{B}_i]^c = [(\overline{B}_i)^c, (\underline{B}_i)^c]$ , then  $\mathcal{P}_i^1 = \mathcal{M}_i$  and  $(\mathcal{M}_i)^c$  are of equal cardinalities. Therefore,  $(\mathcal{M}_i)^c = \mathcal{P}_i^0$ . Let us prove (ii)  $\Rightarrow$  (vi). From Lemma 2;  $\mathcal{Y}_i$  and  $\mathcal{N}_i$  are equal, made of a single state, which state is made of a single agent. That is,  $|j \mid \phi(j \rightarrow i) \neq 0| = 1$  for all  $i$ . Let us prove (vi)  $\Rightarrow$  (v) For a given  $i$ , let us assume without loss of generality that the agent having a non-null influence on it exerts a positive one. Therefore,  $\mathcal{M}_i = [j, N]$ . We have  $(\mathcal{M}_i)^c = [\emptyset, N \setminus j] = (\mathcal{M}_i)^{(c)}$ . Equivalence between (vi) and (vii) is straightforward. □

We can now prove Theorem 8.

- (1) • First part.

( $\Leftarrow$ ) Let  $S \xrightarrow{1} T$  and  $i, j$  arbitrary. Assume that  $\phi(j \rightarrow i) < 0$  and  $\mathbf{m}_S(j, i) = 1$  (the case  $\phi(j \rightarrow i) > 0$  is analogous). By Lemma 1, either  $j \in S$  and  $i \in T$ , or  $j \notin S$  and  $i \notin T$ . In the first case, we have  $S \in \mathcal{P}_i^1 = \mathcal{M}_i = [\underline{B}_i, \overline{B}_i]$ , and therefore  $\phi(j \rightarrow i) \geq 0$  (absurd). The second case ( $i \notin T$ ) is obtained from Lemma 3: we have  $S \in \mathcal{P}_i^0 = (\mathcal{M}_i)^c = (\mathcal{M}_i)^{(c)} = [\overline{B}_i^c, \underline{B}_i^c]$ ; since  $j \notin S$ , this implies  $\phi(j \rightarrow i) \geq 0$  too (absurd).

( $\Rightarrow$ ) Assume that  $\mu(S) = 0$  for all  $S \xrightarrow{1} T$ . Then for a given transition  $S \xrightarrow{1} T$ :

- (a)  $\phi(i \rightarrow j) > 0 \Rightarrow i \in S$  and  $j \in T$ , or  $i \notin S$  and  $i \notin T$ .
- (b)  $\phi(i \rightarrow j) < 0 \Rightarrow i \in S$  and  $j \notin T$ , or  $i \notin S$  and  $i \in T$ .

Assume that there exists  $k, S_1$  such that  $S_1 \in \mathcal{P}_k^1 \setminus \mathcal{M}_k$ . Then, there exists  $S_2 \in \mathcal{P}_k^1 \setminus \mathcal{M}_k$  such that  $[[S_1, S_2]] \not\subseteq \mathcal{P}_k^1$ ; in particular,  $S_1 \neq S_2$  and, since  $\mathcal{M}_k \neq \emptyset$ ,  $S_3 := S_1 \wedge_{\overline{B}_k} S_2 \in \mathcal{P}_k^0$ . Consider two chains of sets linked in the Hasse diagram, i.e., such that  $|S_r \Delta S_{r+1}| = 1$  and  $|S'_r \Delta S'_{r+1}| = 1$  for all  $r = 3 \dots p$ :

$$S_3 \preceq_{S_1} S_4 \preceq_{S_1} \dots \preceq_{S_1} S_p \preceq_{S_1} S_1$$

and:

$$S_3 \preceq_{S_2} S'_4 \preceq_{S_2} \dots \preceq_{S_2} S'_p \preceq_{S_2} S_2$$

Let  $q$  be such that  $A_k(1_{S_q}) = 0$  and  $A_k(1_{S_{q+1}}) = 1$ , and  $q'$  such that  $A_k(1_{S'_{q'}}) = 0$  and  $A_k(1_{S'_{q'+1}}) = 1$ . Let  $j := S_q \Delta S_{q+1}$  and  $l := S'_{q'} \Delta S'_{q'+1}$  and notice that  $j \neq l$ . Without loss of generality, assume that  $j \in S_{q+1}$  and  $l \in S'_{q'+1}$ . Therefore:  $\phi(j \rightarrow k) > 0$  and  $\phi(l \rightarrow k) > 0$ . Because of (a): (i)  $j \in S \Rightarrow A_k(1_S) = 1$  and  $j \notin S \Rightarrow A_k(1_S) = 0$ : (ii)  $l \in S \Rightarrow A_k(1_S) = 1$  and  $l \notin S \Rightarrow A_k(1_S) = 0$ . Therefore,  $j$  and  $l$  are bosses for  $k$ ; but there cannot be two distinct bosses for an agent. Therefore,  $\mathcal{P}_i^1$  is an interval for all  $i \in N$ . By the same token,  $\mathcal{P}_i^0$  is an interval for all  $i$ . By Lemma 3, this implies  $\mathcal{P}_i^0 = (\mathcal{P}_i^1)^{(c)}$ .

- Second part. We have  $\mathcal{M}_i = \mathcal{P}_i^1$  for all  $i \in N$ .  
From Lemma 1 and Theorem 2:

$$\underline{B}_i = \left( \bigcap_{\substack{S \xrightarrow{1} T \\ T \ni i}} S \right) \cup \left( \bigcap_{\substack{S \xrightarrow{1} T \\ T \not\ni i}} S^c \right) \quad (45)$$

$$\overline{B}_i = N \setminus \left( \left( \bigcap_{\substack{S \xrightarrow{1} T \\ T \ni i}} S \right) \cup \left( \bigcap_{\substack{S \xrightarrow{1} T \\ T \not\ni i}} S^c \right) \right) \quad (46)$$

From the first part, since the  $\mathcal{P}^0$  and  $\mathcal{P}^1$  are intervals, they are both of cardinalities  $2^{n-1}$ . In particular, none of them contains two complement sets. Therefore,  $S \xrightarrow{1} T \Leftrightarrow S^c \xrightarrow{1} T^c$ . This implies:  $\bigcap_{\substack{S \xrightarrow{1} T \\ T \ni i}} S = \bigcap_{\substack{S \xrightarrow{1} T \\ T \not\ni i}} S^c$  and:  $\bigcap_{\substack{S \xrightarrow{1} T \\ T \not\ni i}} S = \bigcap_{\substack{S \xrightarrow{1} T \\ T \ni i}} S^c$ ,

$$\bigcap_{\substack{S \xrightarrow{1} T \\ T \ni i}} S \quad \bigcap_{\substack{S \xrightarrow{1} T \\ T \not\ni i}} S \quad \bigcap_{\substack{S \xrightarrow{1} T \\ T \ni i}} S^c \quad \bigcap_{\substack{S \xrightarrow{1} T \\ T \not\ni i}} S^c$$

which simplifies (45) and (46) and gives the forms announced.

- (2) Let us show that assuming (A) to (D) together leads to a contradiction.

For all  $j \in S$  and  $i \in T$ :  $\phi(j \rightarrow i) \leq 0$ . (A)

For all  $j \in S^c$  and  $i \in T^c$ :  $\phi(j \rightarrow i) \leq 0$ . (B)

For all  $j \in S^c$  and  $i \in T$ :  $\phi(j \rightarrow i) \geq 0$ . (C)

For all  $j \in S$  and  $i \in T^c$ :  $\phi(j \rightarrow i) \geq 0$ . (D)

These assumptions on  $\phi$  can be expressed by sets inclusions w.r.t  $\underline{B}_i$  and  $\overline{B}_i$ :

(A) and (C) together give  $S \in (\mathcal{M}_i)^{(c)}$  for all  $i \in T$ .

(B) and (D) together give  $S \in \mathcal{M}_i$  for all  $i \in T^c$ .

*Step 1.*  $\mathcal{P}_i^1 \supseteq (\uparrow S) \cup (\downarrow S)$  for all  $i \in T$  and  $\mathcal{P}_i^0 \supseteq (\uparrow S) \cup (\downarrow S)$  for all  $i \notin T$ .

Let  $S^*, S^{**}$  be such that  $S^* \subseteq S \subseteq S^{**}$ .

(A) implies  $S^* \xrightarrow{1} T'$  for some  $T' \supseteq T$  (because for  $i \in T$ :  $A_i(1_{S^*}) \geq A_i(1_S) = 1$ ).

(B) implies  $S^{**} \xrightarrow{1} T''$  with  $T'' \subseteq T$  (because for  $i \in T^c$ :  $A_i(1_{S^{**}}) \leq A_i(1_S) = 0$ ).

(C) implies  $S^{**} \xrightarrow{1} T'''$  for some  $T''' \supseteq T$  (because for  $i \in T$ :  $A_i(1_{S^{**}}) \geq A_i(1_S) = 1$ ).

(D) implies  $S^* \xrightarrow{1} T''''$  for some  $T'''' \subseteq T$  (because for  $i \in T^c$ :  $A_i(1_{S^*}) \leq A_i(1_S) = 0$ ).

(A) yields  $S^* \in \mathcal{P}_i^1$  for all  $i \in T$ .

(B) yields  $S^{**} \in \mathcal{P}_i^0$  for all  $i \notin T$ .

(C) yields  $S^{**} \in \mathcal{P}_i^1$  for all  $i \in T$ .

(D) yields  $S^* \in \mathcal{P}_i^0$  for all  $i \notin T$ .

This holds for any  $S^*, S^{**}$  such that  $S^* \subseteq S \subseteq S^{**}$ . Therefore, for any set  $U$  of any chain from  $\emptyset$  to  $N$  passing through  $S$  and for any  $i \in T$ :  $A_i(1_U) = 1$ ; in other words,  $\mathcal{P}_i^1 \supseteq (\uparrow S) \cup (\downarrow S)$ . By the same token, for any  $i \notin T$ :  $A_i(1_U) = 0$ ; in other words,  $\mathcal{P}_i^0 \supseteq (\uparrow S) \cup (\downarrow S)$ .

*Remark 3.* All these collections inclusions are actually strict. If indeed equality held, let us say for some  $i \in T$ , then we would get  $\mathcal{M}_i = \{S\}$ , that is,  $\underline{B}_i = S$ . But this would contradict  $\phi(j \rightarrow i) \leq 0$  for all  $j \in S$ .

*Step 2.*  $S \in \mathcal{M}_i$  for all  $i \in T$  and  $S \in (\mathcal{M}_i)^{(c)}$  for all  $i \notin T$ .

Let  $i \in T$  (the case  $i \notin T$  is analogous) and assume that  $S \notin \mathcal{M}_i$ .

Since: (i) aggregation functions are Boolean; (ii)  $[[S, U]] \subseteq \mathcal{P}_i^1$  for all  $U \in (\uparrow S) \cup (\downarrow S)$  and (iii)  $S \notin \mathcal{M}_i$ , there exists  $X_1$  such that  $X_1 \in \mathcal{P}_i^1 \setminus ((\uparrow S) \cup (\downarrow S))$  with  $[[S, X_1]] \not\subseteq \mathcal{P}_i^1$ . Since  $[[S, X_1]] \not\subseteq \mathcal{P}_i^1$ , there exists  $S' \in [[S, X_1]] \cap \mathcal{P}_i^0$ . The set  $S'' := S' \setminus \underline{B}_i$  belongs to  $[[S, X_1]]$  too and, since  $S' \in \mathcal{P}_i^0$ , it holds also that  $S'' \in \mathcal{P}_i^0$ . Let  $Y := (S_1 \cup X_1) \cap (\underline{B}_i)^c$ . Since by (A) we have  $S \subseteq (\underline{B}_i)^c$ , on the one hand we have  $Y \in (\uparrow S)$ , which by Step 1 implies that  $Y \in \mathcal{P}_i^1$ . On the other hand we have  $Y \preceq_B S''$  for some  $B \in \mathcal{M}_i$ . Since  $S'' \in \mathcal{P}_i^0$ , this implies  $Y \in \mathcal{P}_i^0$ , a contradiction.

*Conclusion.*

By Step 2,  $\phi(j \rightarrow i) \geq 0$  for all  $j \in S$  and  $i \in T$  and  $\phi(j \rightarrow i) \leq 0$  for all  $j \in S$  and  $i \notin T$ . Because of (A) and (D), this will imply that  $\phi(j \rightarrow i) = 0$  for all  $j \in S$  and  $i \in N$ . Since:

- $S \in \mathcal{M}_i$  for all  $i \in T$  is equivalent to  $S^c \in \mathcal{M}_i^{(c)}$  for all  $i \in T$ .
- $S \in \mathcal{M}_i^{(c)}$  for all  $i \notin T$  is equivalent to  $S^c \in \mathcal{M}_i$  for all  $i \notin T$ .

we will have also that  $\phi(j \rightarrow i) = 0$  for all  $j \notin S$  and  $i \in N$ . In other words, all agents would be irrelevant on all agents, which would contradict that the aggregation functions are non constant.

□

## Appendix B. The ‘four polarisations’. Transitions from $\emptyset, N_1, N_2$ and $N$ under majority influence

$V_1 =$ $V_2 =$	(+, +)	(+, -)	(-, +)	(-, -)
(+, +)	$N_2 \xrightarrow{1} \emptyset \xrightarrow{1} \emptyset$ $N_1 \xrightarrow{1} N \xrightarrow{1} N$ (conformist polarisation C)	$N_2 \xrightarrow{1} \emptyset \xrightarrow{1} \emptyset$ $N_1 \xrightarrow{1} N \xrightarrow{1} N$	$N \xrightarrow{1} N_2 \xrightarrow{1} N_1 \xrightarrow{1} \emptyset \xrightarrow{1} N_1$	$N \xrightarrow{1} N_2 \xrightarrow{1} N_1 \xrightarrow{1} \emptyset \xrightarrow{1} N_1$ (mixed polarisation M <sub>2</sub> )
(+, -)	$N_2 \xrightarrow{1} \emptyset \xrightarrow{1} \emptyset$ $N_1 \xrightarrow{1} N \xrightarrow{1} N$	$N_2 \xrightarrow{1} \emptyset \xrightarrow{1} \emptyset$ $N_1 \xrightarrow{1} N \xrightarrow{1} N$ (leadership polarisation L <sub>1</sub> )	$N \xrightarrow{1} N_2 \xrightarrow{1} N_1 \xrightarrow{1} \emptyset \xrightarrow{1} N_1$ (communitarian polarisation R <sub>2</sub> )	$N \xrightarrow{1} N_2 \xrightarrow{1} N_1 \xrightarrow{1} \emptyset \xrightarrow{1} N_1$
(-, +)	$\emptyset \xrightarrow{1} N_2 \xrightarrow{1} N_2$ $N \xrightarrow{1} N_1 \xrightarrow{1} N_1$	$\emptyset \xrightarrow{1} N_2 \xrightarrow{1} N_2$ $N \xrightarrow{1} N_1 \xrightarrow{1} N_1$ (communitarian polarisation R <sub>1</sub> )	$N_1 \xrightarrow{1} \emptyset \xrightarrow{1} N \xrightarrow{1} \emptyset$ $N_2 \xrightarrow{1} N \xrightarrow{1} \emptyset \xrightarrow{1} N$ (leadership polarisation L <sub>2</sub> )	$N_1 \xrightarrow{1} \emptyset \xrightarrow{1} N \xrightarrow{1} \emptyset$ $N_2 \xrightarrow{1} N \xrightarrow{1} \emptyset \xrightarrow{1} N$
(-, -)	$\emptyset \xrightarrow{1} N_2 \xrightarrow{1} N_2$ $N \xrightarrow{1} N_1 \xrightarrow{1} N_1$ (mixed polarisation M <sub>1</sub> )	$\emptyset \xrightarrow{1} N_2 \xrightarrow{1} N_2$ $N \xrightarrow{1} N_1 \xrightarrow{1} N_1$	$N_1 \xrightarrow{1} \emptyset \xrightarrow{1} N \xrightarrow{1} \emptyset$ $N_2 \xrightarrow{1} N \xrightarrow{1} \emptyset \xrightarrow{1} N$	$N_1 \xrightarrow{1} \emptyset \xrightarrow{1} N \xrightarrow{1} \emptyset$ $N_2 \xrightarrow{1} N \xrightarrow{1} \emptyset \xrightarrow{1} N$ (anti-conformist polarisation F)

Table 3: Transitions from  $\emptyset, N_1, N_2$  and  $N$  under majority influence, where  $n_1 > n_2$