

Funded by the European Union’s Internal Security

Fund - Police (2014-2020)

**Internal Security Fund – Police - Programme (2014-2020)**

**823701 - YoungRes - ISFP-2017-AG-RAD**

**Strengthening European Youngsters Resilience through Serious Games -**

**YoungRes**

|  |
| --- |
| **Deliverable D4.1** |
| Software Measurement Metrics for Polarization |

|  |  |  |  |
| --- | --- | --- | --- |
| **Document information** | | | |
| **Due date of deliverable** | | 30-06-2019 | |
| **Actual submission date** | | 14-08-2019 | |
| **Organisation name of lead partner for this deliverable** | | UO | |
| **Final Version Number** | | V1 | |
|  | | | |
|  | | | |
| **Dissemination Level** | | | |
| **PU** | Public | |  |
| **RE** | Restricted to a group specified by the consortium (including the Commission Services) | |  |
| **CO** | Confidential, only for members of the consortium (including the Commission Services) | | X |

|  |  |
| --- | --- |
| **Authors list** | |
| **Author** | **Partner** |
| Mourad | UO |
|  |  |
|  |  |
|  |  |

|  |  |
| --- | --- |
| **Peer Reviewers** | |
| **Reviewer** | **Partner** |
| David Camacho | UAM |
| ? |  |
| ? |  |
|  |  |
|  |  |

|  |  |  |
| --- | --- | --- |
| **Versioning** | | |
| **Version** | **Date** | **Summary** |
| V1 | 22-05-2019 | First version |
| V2 | 27-05-2019 | Second version. Comments and logo included. |
| V3 | 30-05-2019 | Final version. |
|  |  |  |

The content of this report represents the views of the author only and is his/her sole responsibility. The European Commission does not accept any responsibility for use that may be made of the information it contains.

Contents

[EXECUTIVE SUMMARY 3](#_Toc16687124)

[POLARIZATION & SOFTWARE METRIC MEASUREMENTS 4](#_Toc16687125)

[2 Polarization as a multidisciplinary field 4](#_Toc16687126)

[2.1 What does polarization involve? 5](#_Toc16687127)

[2.2 Group polarization 6](#_Toc16687128)

[2.3 Polarization and irrational belief 7](#_Toc16687129)

[3 Polarization measurement viewpoints 8](#_Toc16687130)

[3.1 Polarization measurement in economics studies 9](#_Toc16687131)

[3.2 Social polarization indices 11](#_Toc16687132)

[3.3 Ethics Polarization 15](#_Toc16687133)

[3.4 Group polarization measurement 16](#_Toc16687134)

[3.5 Political party polarization measures 17](#_Toc16687135)

[3.6 Network polarization 19](#_Toc16687136)

[4 Reconciling Polarization measures 21](#_Toc16687137)

[4.1 Senses of polarization 21](#_Toc16687138)

[4.2 Major polarization models and the nine senses 27](#_Toc16687139)

[4.3 Other opinion dynamic based polarization models 33](#_Toc16687140)

[5 Causes of polarization 35](#_Toc16687141)

[5.1 Social comparison versus persuasive argument theories 35](#_Toc16687142)

[5.2 Factors influencing group polarization 37](#_Toc16687143)

[5.3 Impact of Computer mediated Communication 38](#_Toc16687144)

[5.4 Impact of online discussion forums 39](#_Toc16687145)

[6 Open questions, challenges for YOUNGRES community and ways forward 40](#_Toc16687146)

[6.1 Mathematical constructions of polarization measures 41](#_Toc16687147)

[6.2 Relevancy of Statistical models 41](#_Toc16687148)

[6.3 Cascade effects 42](#_Toc16687149)

[6.4 Iterated Polarization Games 42](#_Toc16687150)

[6.5 Rhetorical Asymmetry 43](#_Toc16687151)

[6.6 Role of communication 43](#_Toc16687152)

[6.7 Social media and online forum polarization 45](#_Toc16687153)

[6.8 Level of interactions 46](#_Toc16687154)

[6.9 Brandsma’s polarization model promoted by EU RAN specialists 47](#_Toc16687155)

# EXECUTIVE SUMMARY

This report is intended to provide an overview of the various aspects of the polarization phenomenon in order to comprehend the metrics that can be employed for the polarization measures. The report stresses on the multidisciplinary aspect of the polarization where sociologists, psychologists, economists, physicians, computer scientists as well as security specialists have all tackled some aspects of polarization phenomenon. Conceptual differences and theoretical foundations that govern the emergence of group polarization are elicited. Due to the discrepancy in expectations and background among distinct disciplines and scholars, various polarization measures have been put forward and reviewed. The report contrasts the axiomatically elaborated polarization measures in econometrics that tackle income distribution over population with the virtually less-elaborated measures dedicated for political party polarization, social polarization and social network polarization. A review of literature was provided and shed the light on the key research questions that would pave the way for future developments within YoungRes project have been explored. In the context of group polarization, the report emphasizes the role of the opinion dynamic and the inherent sensibility of the polarization measures to the topology of the social network and the opinion formation models. In this respect, crucial importance has been raised to the potential effect of opinion bias, prior knowledge and echo-chamber that can fuel fake news and irrational behavior.

The report also shed the light on the reverse phenomenon referred to *depolarization*, rising the question of asymmetry of polarization measures in order to estimate the depolarization process.

From security perspective, the report outlines the populated Brandsma’s polarization model that has been highly recommended by the European RAN specialists, highlighting the need for further mathematical elaboration and empirical research in order to lie down more rigorous theoretical foundations and elaborate appropriate polarization – depolarization metrics.

# POLARIZATION & SOFTWARE METRIC MEASUREMENTS

# Polarization as a multidisciplinary field

Due to its pervasiveness, inherent multiple causes and consequences, polarization has attracted interests from various disciplines. In political science, the fundamental divide between Democrats and Republicans in US or the spectacular rise of nationalism in European sphere in a way that put Europeans values at risk have boosted the interest to polarization as a complex social phenomenon that requires special focus. In social science, polarization appears as an inherent characteristic that governs the dynamic of communities, providing useful insights to comprehend cultural diversity, community integration and predicts the development of social norms and patterns. In security science, polarization is often viewed as a catalyst that may trigger radicalization and conflict where hate crimes, growing intolerance, xenophobia, street violence – all are manifestations of polarizations that pose serious threats to Europe fundamental values. In economics, polarization is highly influenced by the analysis of income distribution, which is mainly motivated by the phenomenon of “disappearing middle class”, often associated to a rising potential of social tensions /disturbances caused by the gap between poor lower class and a rich upper class. In physics and computational science, polarization appears as a special case of the dynamic of the social network associated to a set of agents (individuals).

Consequently, polarization becomes a topic of intense interest among social scientists, with analysis of voting patterns in political sciences, formation and dynamic of popular attitude in sociological studies, media influence in communication studies, income distribution evolution in economic studies, ethnic conflict resolution and community integration in security sciences, attitude change in social psychology, group and multi-agent dynamics in on media influence and attitude change all in search of a better understanding of central mechanisms (Fiorina and Abrams 2008).

## What does polarization involve?

* Typically, when one thinks of a polarized distribution, a picture of two large and homogeneous groups opposite to each other comes to mind, a situation that can induce tension and incompatibilities within the population.
* Polarization refers either to a distribution of opinions with multiple local maxima or to the process by which such strong divergences of opinions that divide a population come about (DiMaggio et al., 1996; Bramson et al., 2016).
* Polarization can be seen as a thought construct, based on assumptions of ‘us’ and ‘them’ identities. In a process of polarization, the dominant and active narrative is about the perceived (and often exaggerated) differences and simplistic narratives about the others. There is a neglect of what the ‘us’ and ‘them’ might have in common. Polarization therefore shows itself in negative thoughts and attitudes towards other groups, which could result in growing hostility and segregation. Ultimately this could to lead situations in which intolerance slips into hate speech and even hate crime. In such an environment, some (parts) of groups or individuals can radicalize towards violent extremism and terrorism. Recruiting for an extremist ideology is much more successful if there are susceptible groups who feel their group and identity is being insulted and even threatened.
* In social-science disciplines polarization is often considered as a process. In politics, it is a process by which the public opinion divides and goes to the extremes. In communications and psychology the process involves a social or political group dividing into two opposing sub-groups with fewer and fewer members of the group remaining neutral or holding an intermediate position.
* In the case of income polarization, the concept of polarization assumes the existence of poles, normally two. In the context of income polarization these poles are income levels.
* In their highly interesting contributions, Esteban and Ray (1994) and Wolfson (1994) attempted to provide rigorous definitions of polarization (see also Foster and Wolfson, 2009). While the Wolfson index is concerned with dispersion of the distribution of income from the median towards the extreme points, Esteban and Ray (1994) developed an axiomatic characterization of a class of polarization indices based on distances between incomes

## Group polarization

Social groups as conceived in psychology reflect the characteristics such as small size, purpose, interactions, interdependency, relations with one another, so that a comprehensive definition of group would be a collection of two or more individuals in face to face interaction, each aware of his or her membership in the group, others who belong to the group, and their positive interdependence as they strive to achieve mutual goals (Johnson & Johnson, 2009). Forsyth (2006) suggests four theoretical perspectives that explain the importance of group membership namely, the socio-biological, psychodynamic, social comparison, and the social exchange theories. From the social comparison perspective, groups can reward the members with reassurance, social support, and often afford interaction with competent and attractive others and the motivation.

Often group members are more likely to hold different opinions about an issue initially; nevertheless, after interactions and exchanges of information during the group meeting, there is a tendency for members to come closer together in their opinions. This group tendency to make consensus decisions is also sometimes referred to as group polarization, which describes the process the group discussion tends to intensify convergence of group opinions, or inversely, divergence of group members.

In terms of quantitative estimation, group polarization is often determined by calculating the mean difference between pre-group individually made decisions and post-group decisions (Chen, Gustafson & Lee, 2002).

A large number of other scholars view group polarization as the phenomenon that group members have a tendency to become more extreme in their thinking (positions, decisions, or choices) after group discussions (Isenberg 1986; Myers and Lamm 1976). Therefore, group polarization presents the effects of extreme inclinations of people after deliberating with each other. In other words, in group polarization scenarios, groups often make more extreme decisions than would the typical or average individual in the group (where “extreme” is defined internally, by reference to the group’s initial dispositions) makes.

One negative consequence of group polarization occurs when individual make risky decisions. This may result because individuals in a group sometimes do not feel as much responsibility and accountability for the actions of the group as they would if they were making the decision alone.

## Polarization and irrational belief

Belief polarization is seen as an example of irrational behavior. Belief polarization occurs when two people with different prior beliefs observe the same data and subsequently strengthen their beliefs (Batson, 1975). For example two doctors make a judgment about the same patient. The patient has either Disease 1 or Disease 2 and the two doctors initially disagree about the probability of each disease. The doctors observe the same piece of evidence—a cholesterol test result—and subsequently update their beliefs in opposite directions, both be- coming more certain about their initial diagnoses. ). This is a form of b*elief divergence* in which the person with the stronger belief in a hypothesis increases the strength of his or her belief and the person with the weaker belief decreases the strength of his or her belief.

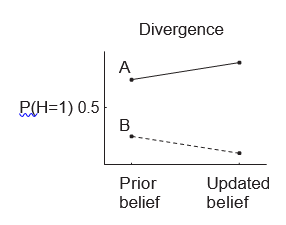


Fig 1. Belief update rationality

More formally, consider a situation in which two people observe data *D* that bear on some hypothesis *H*. Let *PA*(·) and *PB*(·) be probability distributions that capture the two people’s respective beliefs. Opposite updating occurs whenever one person’s belief in *H* increases after observing *D* and the other person’s belief in *H* decreases after observing *D*, or equivalently, when

[*PA*(*H* | *D*) — *PA*(*H*)][*PB*(*H* | *D*) — *PB*(*H*)] < 0. (0)

The preceding can therefore be employed as a trigger for occurrence of polarization with respect to a given event. Nevertheless, there are also other studies that showed that that belief divergence can result from normative inference if two people make different assumptions about factors that affect the relationship between hypothesis *H* and data *D*, such as which person is playing white in the chess game. Another source of divergence even in case of consistent belief update is related to the relationships between the variables in a situation, it is clear that this may draw widely divergence conclusions from the same data. This is also related to the topology of the underlined structure (or network).

# Polarization measurement viewpoints

Various contrasting, and sometimes complementary views about measurement of polarization can be found in literature, which seems rather natural given the wide scope and multidisciplinary aspects of polarization as already pointed out.

* One is designed to capture the formation of any arbitrary number of poles in population, sometimes called proper measures of polarization
* Another family sees polarization as the process by which a distribution becomes bi-polar, which are sometimes referred as measure of bi-polarization.
* Accordingly, measures of polarization have often been associated with measure of disagreement, dispersion, coverage range, divergence, among others.

In general, three properties seem to consider to be indispensable to any measure of polarization:

(i) polarization is a matter of groups so that when there is one group only there should be little polarization,

(ii) with two or more groups, polarization rises when “within-group dispersion” is reduced, and

(iii) polarization rises when “across-group” distances increase.

## Polarization measurement in economics studies

The strongly influenced Identification-alienation framework, introduced in Esteban and Ray (1994), suggests that polarization can be measured as the effect of two interrelated mechanisms: (1) alienation, which is felt by individuals from a given group (defined by income class, religion, race, education etc.) toward individuals belonging to other groups, and (2) identification, which unites members of any given group. This approach assumes that polarization requires that individuals identify with other members of their socioeconomic group and feel alienation to members of other groups.

Formally, suppose that an individual with income *x* feels group identification *I*(*x, F*) under the distribution *F*, and alienation *r* (*x, y*) with respect to some individual with income *y*. The *effective antagonism* that individual *x* feels towards *y* is modelled as some function *T*(*I, r*) strictly increasing in *r. Effective antagonism functional T* increases with alienation, which is fueled by some sense of identification as well. Polarization in view of Esteban and Ray model corresponds to the ‘sum’ of all effective antagonisms:

 (1)

For density f of F and an inter-personal alienation r that depends on income distance |x – y|, the preceding reads

 (2)

where the functional T fulfills T(i, a) = T(f(x), |x – y|) and with T(0, a) = T(i, 0) = 0.

Expression (1-2) have also been justified axiomatically through the following axioms 1-4

AXIOM 1. If a distribution is composed of a single basic density, then a squeeze of that basic density cannot increase polarization.

where a λ-squeeze of the density f is a transformation of this density as follows:



Especially, λ-squeeze is a very special type of second order stochastic dominance transformation that “globally contracts” the support of a distribution towards its mean

AXIOM 2. If a symmetric distribution is composed of three basic densities drawn from the same kernel, with mutually disjoint supports, then a symmetric squeeze of the side densities cannot reduce polarization.

This axiom argues that a particular “local” squeeze (as opposed to the “global” squeeze of the entire distribution in Axiom 1) must not bring polarization down.

AXIOM 2B. Let two distinct distributions F and G have the same mean and the same median and let F second order stochastically dominate G separately on [0;m] and on [m; ∞): Then F should be more bi-polarized than G.

AXIOM 3. Consider a symmetric distribution composed of four basic densities drawn from the same kernel, with mutually disjoint supports. An equal slide of the two inner densities outwards towards the outer densities makes polarization go up.

AXIOM 4. Let F and G be two distributions with possibly different, unnormalized populations such that P(F) ≤ P(G). Then, for all κ > 0, P(κF) ≥ P(κG), where κF and κG represent (identical) population scalings of F and G respectively.

AXIOM 4B. Let the distributions F and G have the same median income m: Then if G is more polarized than F, so is G’ relative to F’, where G’ and F’ are obtained from G and F respectively by dividing all incomes by their median.

Esteban and Ray (1991) and Duclos, Esteban and Ray (2004) refined (2) with further restriction such that the income density functions ( *f* ) whose integrals of these functions corresponding to various population sizes. It is assumed that identification at income *y* depends on the density at *x*, *f* (*x*), while alienation between two individuals with incomes *y* and *x* is given by |*y*−*x*|, then (2) boils down to

 (3)

where α∈ [0.25, 1] is an ethical parameter expressing the weight given to the identification part of the framework and K > 0 is some arbitrary constant which can be used for normalization.

Wang and Tsui (2000) derived an axiomatization bipolarization measure which reads as.

, r∈(0,1] (4)

Accordingly with this measure, polarization is captured by the average of a concave transformation of the distance with respect to the median income. It is essentially a measure of distance to the polarization minimizing distribution with all the population concentrated at the median income.

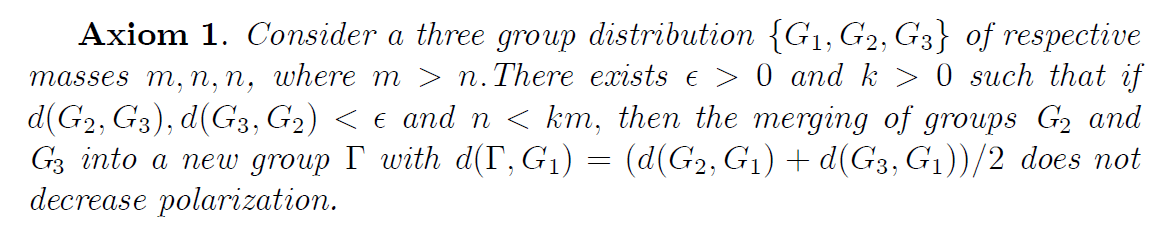
## Social polarization indices

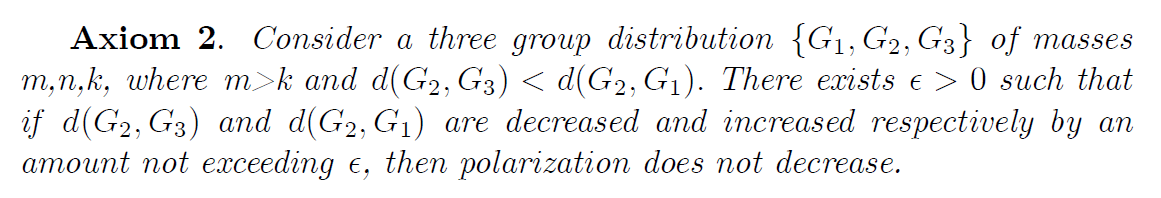
In her original contribution, Reynal-Querol (2002) adjusted the Esteban-Ray (1994) index of income polarization to the case of ethnicity and suggested an index of ethnic polarization, which is known in the literature as the *RQ* index. It records how far the actual ethnic distribution is from the bipolar case, which displays maximum polarization where the entire population is split equally between two ethnic groups only and each of the remaining groups has zero population size. Montalvo and Reynal-Querol (2005) demonstrated that RQ index is as a predictor for estimating the occurrence of civil wars.

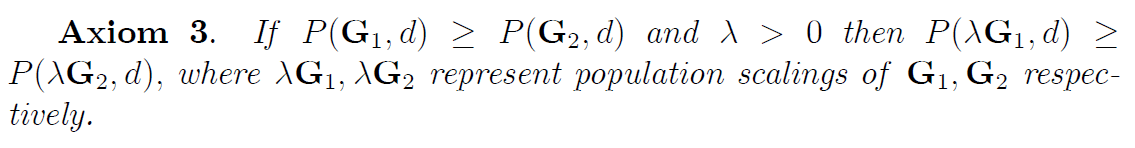
Assuming that the effective antagonism felt by an individual A belonging to the group Gi (of size Mi) towards individual B is measured by a non-negative function T(I, a), with I = I(Mi); a = a(d(A,B)). Besides, T is assumed continuous, increasing in its second argument and T(I; 0) = T(0; a) = 0. Total polarization is postulated to be the sum of all effective antagonisms as in ER model, the discrete formulating of ER model in the multi-group context, the social polarization index is given as

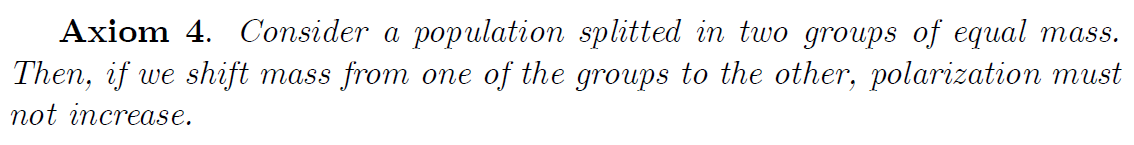
****  (5)

**Axiomatic characterization of social polarization index**

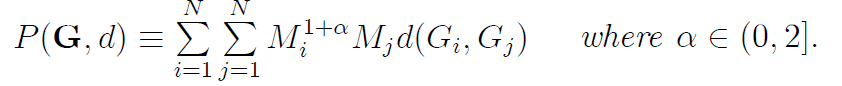
****

****

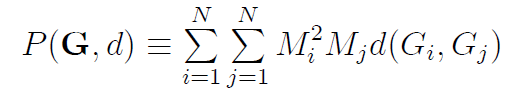
****

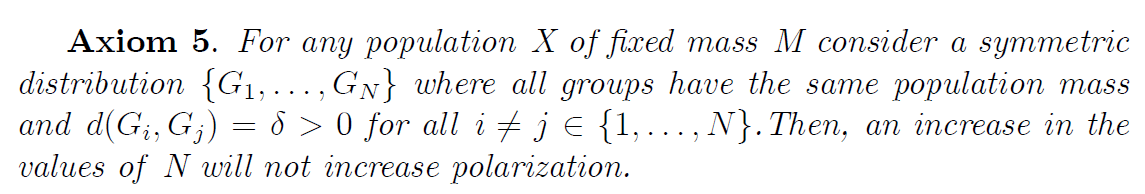
****

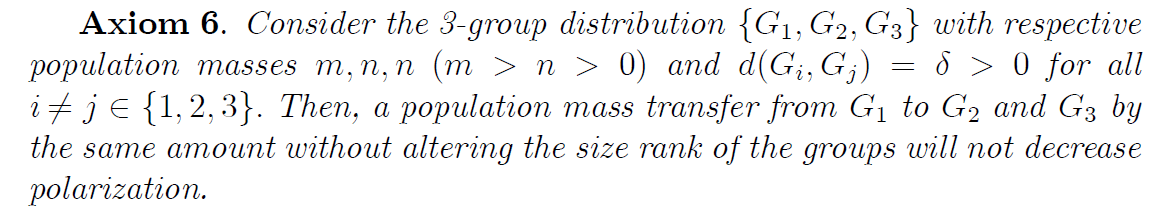
Especially, the fulfilment of Axioms 1-4 constraint the functional P(G,d) as

 (6)

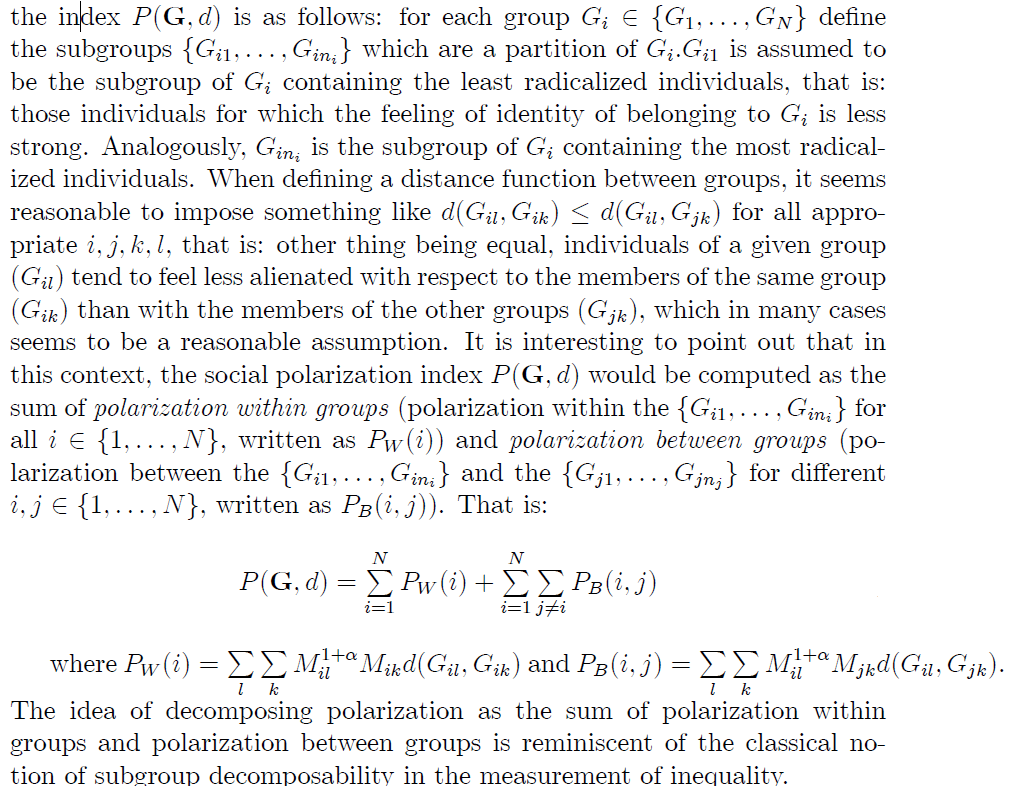
Further axiomatization through Axiom 5 and Axiom 6 below refines the polarization functional as

 (7)



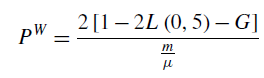


When considering more than two groups, it is important to distinguish the in-group and between-group polarization type. The Reynol-Queral index (RQ) has been used as an index of religious or ethnic polarization where one assumes that all groups feel equally alienated vis-à-vis each other. This is exemplified in the following statements by Permanyer (2008)

****

**Other measures of social group polarization**

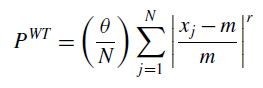
Wofson (1994) defined polarization as

**** (8)

where L(0,5) is the value of the Lorenz curve at the 50th percentile, *G* the Gini coefficient,

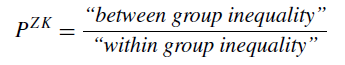
*m* the median, and *μ* the mean of the group.

Wang and Tsu (2000) proposed

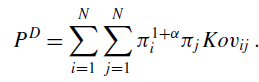
**** (9)

where *θ>* 0 and 0 *< r <* 1*. N* is the number of individuals in the population, *xj* the income of individual *j* for *j* = 1*,..N,* and *m* is the median income of the group

Zhang and Kanbur (2001) suggested the following ratio measure:

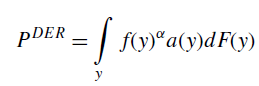
**** (10)

D’Ambrosio (2001) proposed a measure

**** (11)

*where πi* and *πj* are the relative sizes of groups *i* for i,j=1,..,N and *Kovij* is the Kolmogorov measure of variation distance between the density functions of the income distributions in groups *i* and *j*, which are derived using a kernel density estimator.

Duclos et al. (2014) proposed

**** (12)

For a distribution of non-negative data, the measure assigns a value describing the extent of the distribution’s polarization. Polarization itself depends on the number and the size of the groups within the distribution, the homogeneity within each group, and the heterogeneity between the groups

## Ethics Polarization

Consider a population constituted with k ethnic groups E1, E2,.., Ek and let be the proportion of individuals in Ei, yielding a probability distribution  referred to as “ethnic distribution” where

****

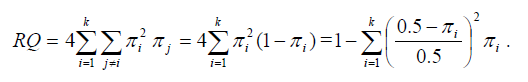
Then the Esteban-Ray index of income polarization forms the basis of RQ polarization, which reads

 (13)

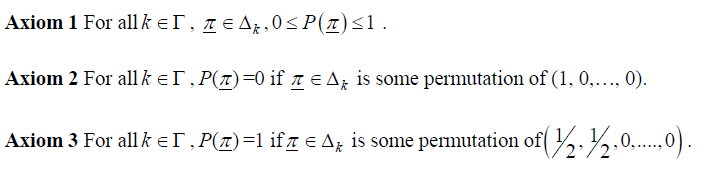
where *yi*  is the representative income of group *i* , λ > 0 is a constant and α∈(0, 1.6],

For α=0, *ER* in corresponds to the Gini index of income inequality.

The popular Reynal-Querol (2008), known as RQ index of ethnic polarization reads

 (14)

Axiomatic characterization of RQ index are the following



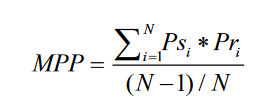
where stands for all discrete probability distributions of dimension k.

## Group polarization measurement

The approach to measuring group polarization is proposed from two different perspectives, at individual level as a preference change (Hinsz and Davis 1984) and at group level as a choice shift (Zuber et al. 1992). Choice shift refers to the difference between the individual’s average “pre-meeting” choice and the final collective choice. When the final group decision is more extreme compared with the average pre-meeting decisions, group polarization occurs. In measuring choice shift, group polarization is considered at a group level. Preference change, which is measured for group polarization at an individual level, takes into account the average difference between an individual’s pre-and post-meeting choices. After group discussions, if an individual’s post-meeting choice moves toward the group choice, then group polarization is claimed to have happened. In the setting of online forum communications, the preference change is considered more difficult to capture. This is attributed to two reasons. First, in a virtual community, online participants come and leave arbitrarily, which means in a group discussion a member could appear and join a discussion but leave in the next second, without waiting till after the discussion. Second, the use of aliases in online forums makes it harder to position each different individual, leading to the challenge of identifying individual decisions. Therefore in this study, group polarization is measured at the group level, the group’s choice shift, based on Whyte (1993) as well, which shows that the magnitude of choice shift is generally stronger than that of preference change.

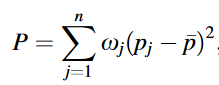
## Political party polarization measures

Sartori (1976) identified political polarization as the ideological distance between candidates, parties, and/or voters. As the distance grows and views become more polarized, the distribution of policy stances or ideological positions shifts from centripetal to centrifugal, taking a U-shape with a plunge in the middle. Most studies in this topic relied on weighted variance calculations to devise indices of political party system polarization (Datton, 2008; Ezrow, 2007; Pardos-Prado and Dinas, 2010). Ladner (2004) defines polarization as the average distance to party system’s ideological center of gravity referring to the variance-based measurement of Taylor and Herman (1971). One also distinguished polarization related to mass population and that of single party. An example of the former measure reads

(15)

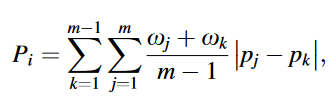
where MPP stands for *mass partisan polarization, Ps* for proportion of respondents within a country expressing intention to vote for a given party (*Party Support), Pr* for the proportion of respondents in a country indicating that they would never vote for a given party (*Party Rejection)* and *N* for *Number of Parties*. As the formula is a ratio of an observed value over the maximum possible value, it ranges from 0 to 1, which enables us to use it for comparisons across time and space. This formula demonstrates the distance of mean party positions on the Left–Right scale from the mean ideological (Left–Right) position of the legislature as a whole (Dalton, 2008)

Other measure of party polarization calculates the weighted sum of squared distances between each party’s position on a left-right scale and the center of gravity of the system (i.e., the weighted average of all the parties’ positions):

**** (16)

where  is the share of the vote received by party j, pj is the position of party j on the left-right scale, and  is the weighted average position of the parties. Using the CSES 11-point scale, this measure of polarization ranges from zero, if all the parties are in the same position, to 25, if the parties are perfectly polarized.

Another interesting measure is the individual perception of polarization given as the average distance the individual places between the parties

 (17)

where j and k are different parties, pj and pk are the positions the respondent assigned parties j and k, ωj and ωk are their vote shares, and m is the number of parties the respondent placed.

Some scholars link party polarization is also linked to societal conflict. Steiner and Martin (2012) argued that party dispersion should be measured on the conflict dimension of party system. Schmitt (2009) measures ideological conflict by a range-based indicator of polarization, while Pardos-Prado and Dinas (2010) claimed that the dispersion-based measure of polarization “indicate the level of political conflict”. Rodrik (1999) argued that the chance of finding a compromise between two groups is less in polarized contexts, or rather “it is difficult to coordinate on a “fair” distribution of resources”. In consequence, polarization causes conflicts, because groups perceive non-cooperative strategies as a more promising approach. The variety and interconnection of common measures of party polarization are summarized in Figure below that highlights the five formula types – variance, standard deviation (SD), mean absolute difference (MAD), range and counting the number of extreme parties – where each of them can further be based on three different weighting functions – equal, vote or seat share weighting.

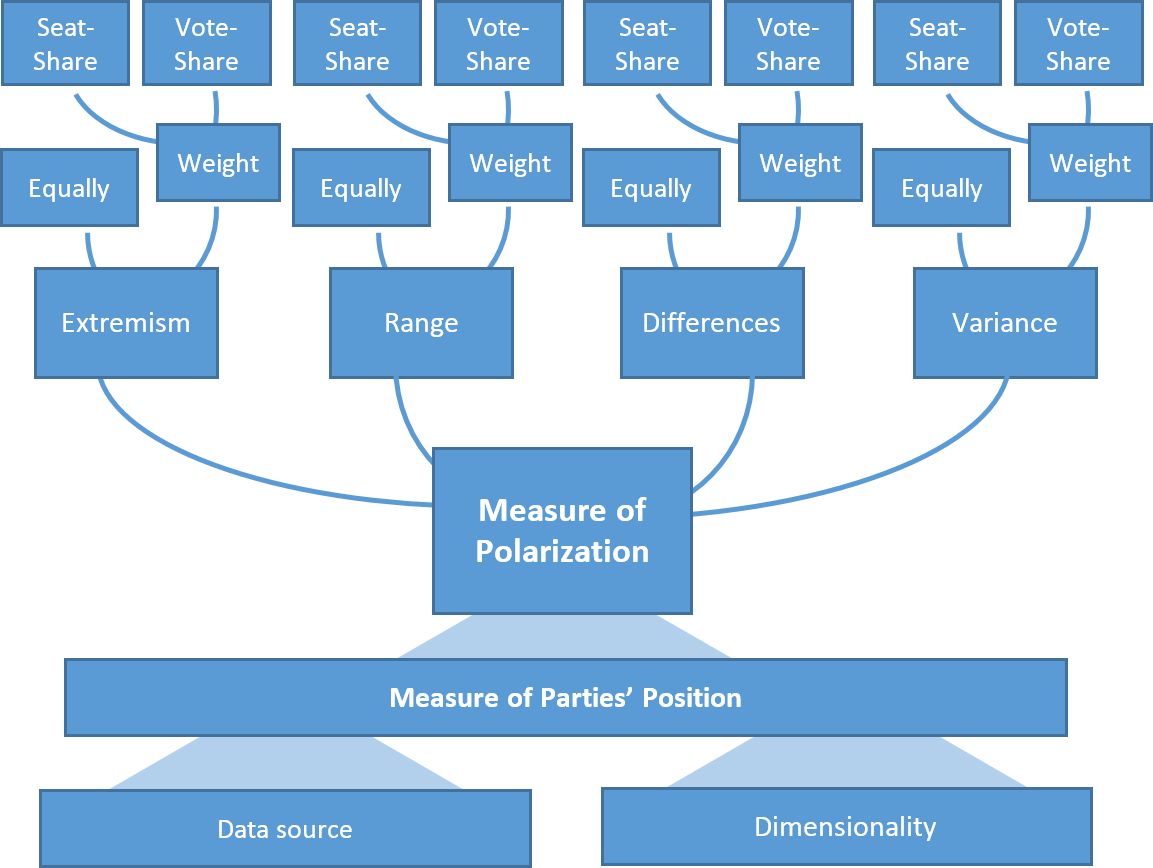


Fig.2. Illustration of variety of polarization measures

## Network polarization

Given a set of agents (individuals) together with their associated relationships that form a network G=(V,E,w) and a vector of opinions x∈[0,1]n, then disagreement index is measured by the amount of divergence of opinions of individuals of the network, which can expressed as

 (19)

Each term can be viewed as the cost of disagreement imposed upon i and j, where denotes the weight of the indirect edge (i,j) between node  and  that evaluates the importance of this link in the graph. This view that the social cost depends on the magnitude of the difference of opinions along edges is consistent with theories in social psychology according to which attitude conflicts in relationships are a source of psychological stress or instability.

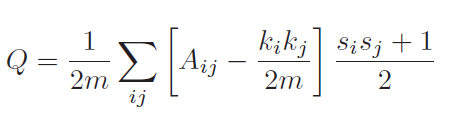
Other polarization index ignores the weighting of individual edges and yields for a total number of n edges

 (20)

Standard deviation like polarization reads for a graph of m nodes

 (21)

Other scholars re-utilized the standard Q-modularity measure, which quantifies the extent, relative to a random network, to which vertices cluster into community groups, and the higher its values, more modular the network is.

** (22)

where A is the adjacency matrix of the network, ki is the degree of the ith node, =1 if nodes i and j belong to the same group, otherwise, it takes -1 value.

*Effects of echo chambers*

The dynamic of social network and marketing strategies have motivated online platforms to boost the preferential attention to viewpoints similar to those already held by individual, as exemplified in platform like Twitter or Facebook to endorse friend’s recommendation, follow particular user, hashtag, raise the “like”, “happy faces” cases. By such recommendations, social network companies create “echo chambers" of similar-minded users. This, however, has also been found to be a crucial element in spread of fake news and polarization phenomenon across online communities. Therefore, special metrics that indicate the proportion of like / dislike counts issued by individuals and overall statistics with respect the underlined case can provide insights into another quantification of polarization measure.

# Reconciling Polarization measures

## Senses of polarization

The multiplicity of actors and disciplines in polarization research would certainly impact the conceptual boundaries and characterization of (social) polarization. One shall summarize below the nine senses of polarization reported by Bramson et al. (2016). The authors distinguished polarization measures in terms of individual distribution characteristics observable from the whole population (spread, dispersion, coverage and regionalization measures) and those defined in terms of group characteristics (community fracturing, distinctness, group divergence, group consensus, size parity) where the groups are either identified directly through histograms or through predefined categories, e.g., region, ethnicity, education level or community structure in a network

1. *Polarization Type 1: Spread.* An obvious way to measure polarization is in terms of the breadth of opinions; that is, how far apart are the extremes?

Polarization in the sense of spread can be measured as the value of the agent with the highest belief value minus the value of the agent with the lowest belief value (sometimes called the ‘range’ of the data); namely, for a distribution X (see Fig. 3)



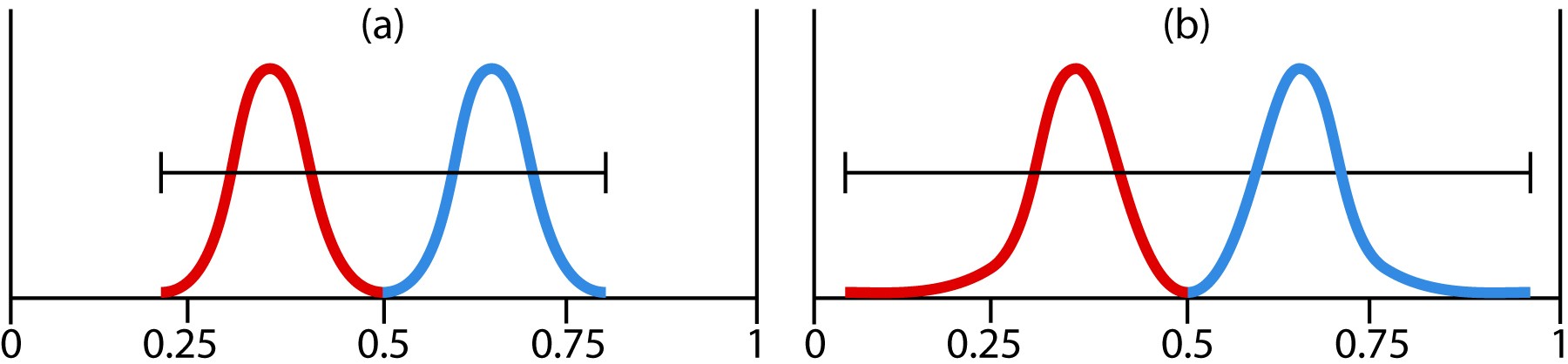


Figure 3. Belief distribution *b* shows greater polarization in the sense of spread than does belief distribution *a*. Two separate groups are shown, but that is irrelevant to polarization in the sense of spread.

1. *Polarization Type 2: Dispersion*

Another simple way to measure polarization is statistical dispersion (or statistical variation). Unlike spread, which considers only the extremes of the population, dispersion considers only the overall shape of the distribution. Any of various measures of statistical dispersion might be used: mean difference, average absolute deviation, standard deviation, coefficient of variation, or entropy, etc. Bimodality is frequently mentioned as a feature of polarized distributions and sometimes as part of the definition

For a population of N individuals, the dispersion given as average absolute deviation from mean corresponds to (see also Fig 4):

,  is the mean value of the distribution X

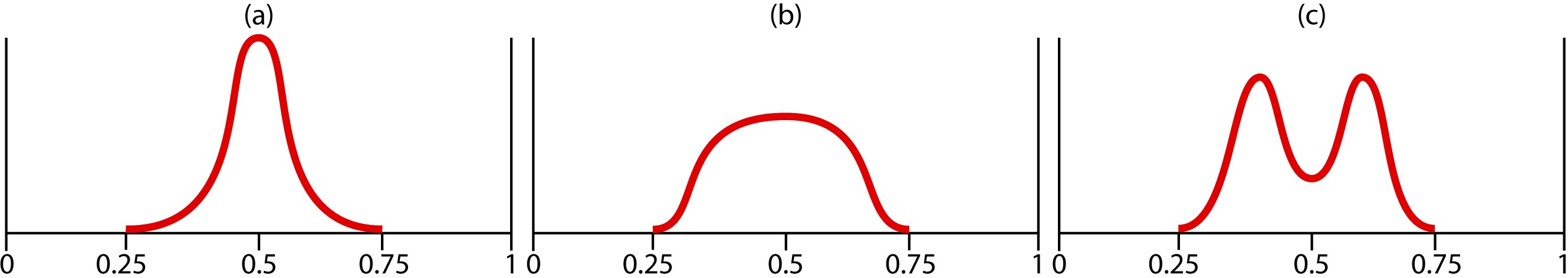


Figure 4. Distribution *c* shows greater polarization in the sense of dispersion than does belief distribution *b*, which is greater than distribution *a*.

1. *Polarization Type 3: Coverage.*

A polarized society is thought of as one with little diversity of opinion, one in which only narrow bands of the opinion space are occupied. A simple way to envisage polarization in this sense is to think of the spectrum of possible beliefs as divided into small bins. The proportion of empty bins will then constitute a measure of polarization as *coverage* (see Fig. 5).

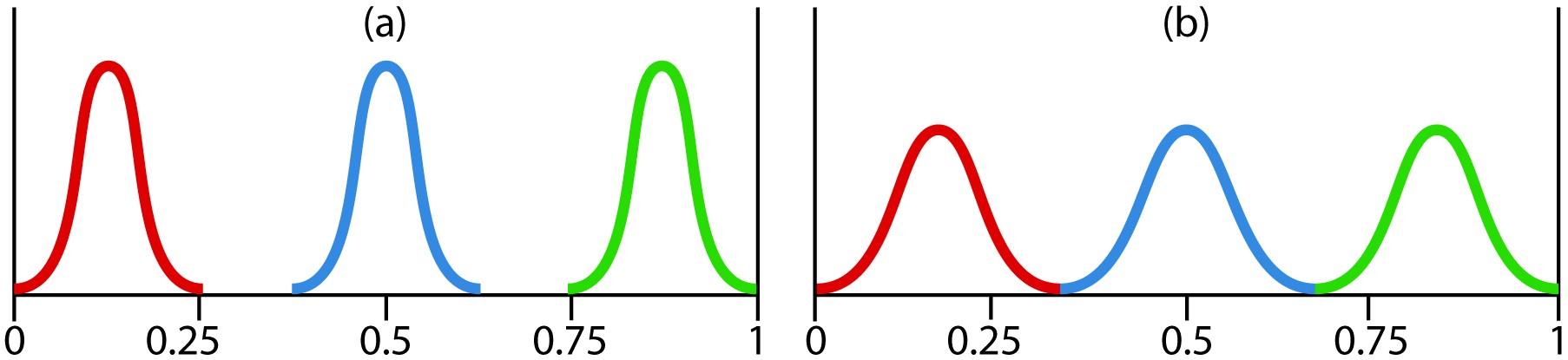


Figure 5. Distribution *a* is more polarized than *b* in the sense of representing less coverage on the spectrum of potential belief.

To quantify this polarization measure, consider a halos of radius r around each agent so that any portion of the attitude space within r of an agent is considered covered, with the rest being uncovered, so calculating the proportion of the spcace covered by those halos leads to



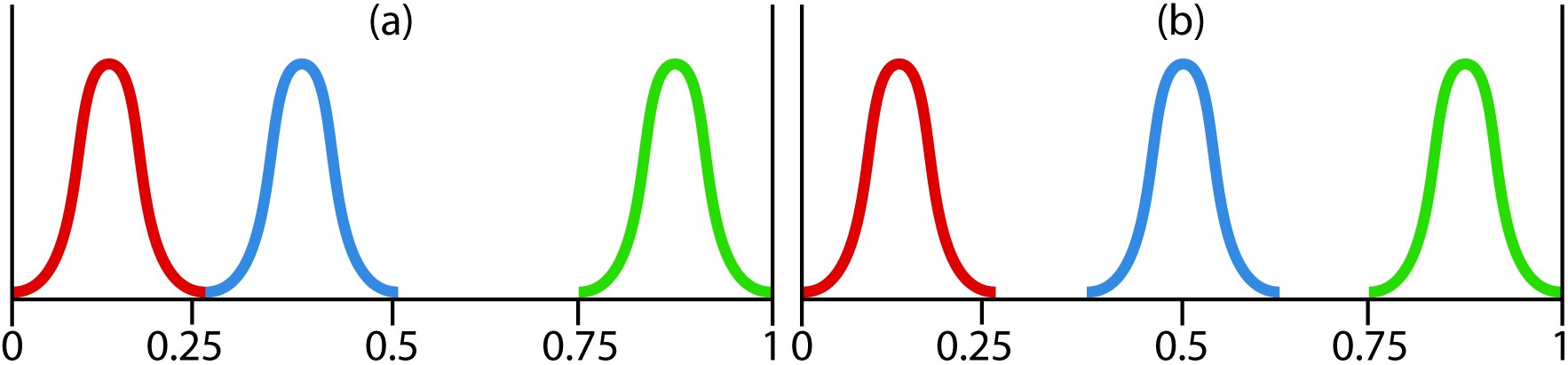
1. *Polarization Type 4: Regionalization.*

Polarization can be used to indicate belief belief regionalization, without attending to the total area covered over all. Regionalization counts the completely distinct clusters in the distribution, related to but distinct from the number of groups.

In order to measure the regionalization on a discrete spectrum it is sufficient to count the regions of contiguous empty bins (including the initial space and trailing space), or use the concept of halos in case of continuous case. For a discrete spectrum with ordered bins R, we can calculate the number of gaps as



Figure 4 shows two cases with the same coverage but in which counting empty regions between occupied areas gives us a measure of regionalization polarization in which *b* is more polarized than *a*.



**Fig. 6.** Distributions with equal coverage and spread, but in which (b) shows a larger number of empty spaces between occupied areas, indicating greater polarization and diversity in the sense of regionalization.

1. Community fragmentation

As a sense of polarization the more groups there are the more difficult it tends to be to get the individuals in those groups to agree and so the greater polarization is. The measure of community fragmentation boils down to quantify the degree to which the population can be broken into sub-population. This includes histogram like techniques where groups are identified directly from the histogram as collections of individuals included within the basin of attraction around peaks; use of exogenous variables (i.e., ethnicity, nationality, location); topology of the network related to relationships among individuals of the population.

1. Distinctiveness

Distinctness expresses the degree to which the groups are separate and distinguishable from each other, regardless of the distance between them. A simple measure for two groups A and B is:



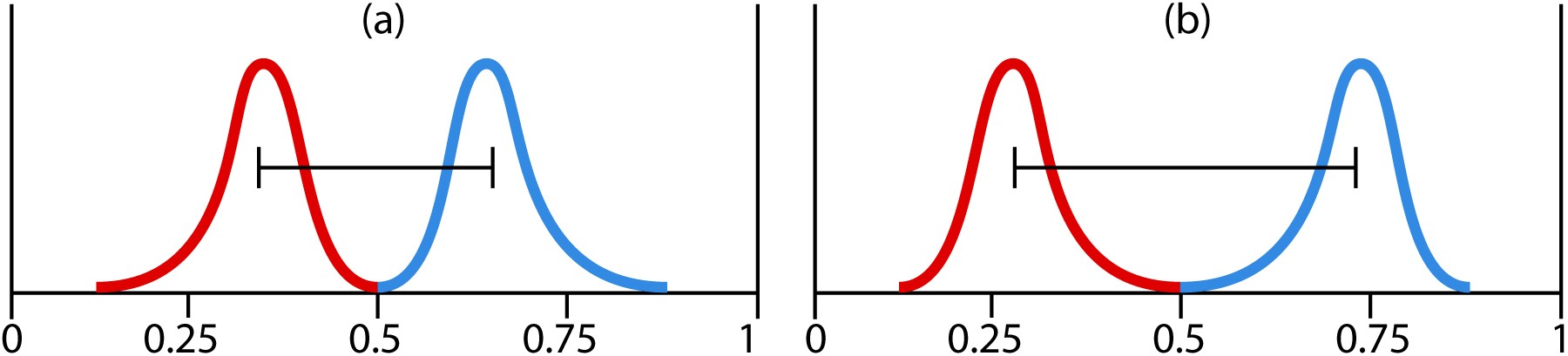
When there are more than two groups in the population, some aggregation of the pairwise comparisons must be made.

1. *Group divergence*

Group divergence captures how distant the groups' ideas are without accounting for their shapes. This contrasts with group distinctiveness which captures how different groups are with respect to being separate regardless of how far away those groups are in their beliefs. The more the groups diverge the greater the polarization and diversity of the population are.

For a set of G belief sub-distributions (groups) {g1 …. gG}, the average over pairwise differences in group means is:





**Fig. 7.** Attitude distribution (b) shows greater polarization and diversity than (a) in the sense of group divergence.

*iix) Group solidarity*

Group solidarity for either polarization or diversity can be measured via any measures of population dispersion: aggregated in-group absolute deviations, variances, pairwise distances, or even aggregated group spreads.

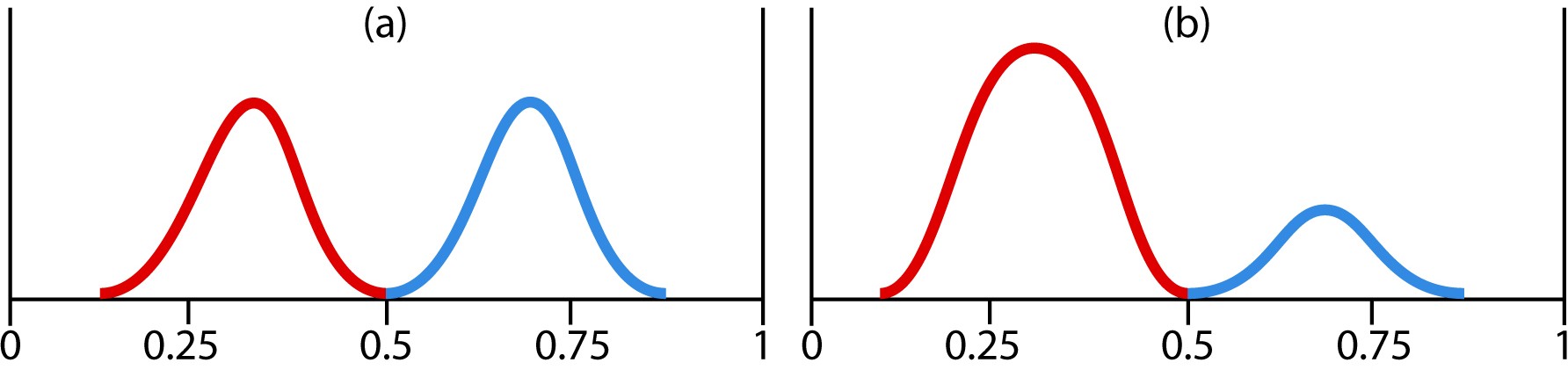
For a set of G belief sub-distributions (groups) {g1 … gG} with with population {n1 … nG}, an example of aggregated group solidarity measure is:



1. *Size disparity*

To measure size disparity independently of the other sense requires a formalism that is sensitive to the sizes of the groups -reaching maximum value when they are all the same - while being simultaneously insensitive to the number of groups. A normalized entropy measure fits this description exactly. For a set of G groups each with population proportion {p1 … pG},





**Fig. 8.** Groups with comparable sizes are more polarized and diverse than a large group with smaller outlier groups; therefore in the sense of size disparity, attitude distribution (a) shows a more polarized and more diverse population than (b).

## Major polarization models and the nine senses

In the majority of polarization models, polarization is viewed as a property of an opinion formation process instead of a property of a state of the network of agents constituting the population where opinion formation dynamics are polarizing if they result in an increased divergence of opinions.

A pioneer work in this field is DeGroot’s model of opinion formation. In this model, individuals are connected to each other in a social network whose edges are associated with weights representing the extent to which neighbors influence each other’s opinions. Individuals update their opinion as a weighted average of their current opinion and that of their neighbors.

An individual i has an opinion xi(t) ∈ [0, 1] at time step t=0, 1,2.... The extreme opinions 0 and 1 represent opposing points of view on an issue. In DeGroot’s model, the opinion of individual i is updated as a weighted sum of the opinions of its neighbors N(i), so, for weight matrix W that quantifies the relationship strength for each pair of individuals where , the opinion of the ith individual at time t+1, while interacting with its neighbors, is given by



Especially, opinion update under DeGroot’s process is equivalent to taking a weighted average of the total support for 0 and that for 1.

For an opinion vector b, the network disagreement index ND, which plays key role in polarization is provided by



We shall consider four main models of social polarization: Axelrod family of cultural diffusion models, bounded conﬁdence, relative agreement models, and structural balance models layered with belief transmission.

*Axelrod’s cultural diffusion model*

In Axelrod (1997), the authors explains the diversity of cultures as a consequence of homophily, which is the tendency to interact more frequently with individuals which are more similar. In his model, actors are characterized by a finite number of cultural features. In Axelrod’s own words, the more similar an actor is to a neighbor, the more likely that actor will adopt one of the neighbor’s traits.

In this model, individuals are located on the nodes of a network (or on the sites of a regular lattice) and are endowed with F integer variables (σ1, σ2,… σF ) that can assume q values, σf = 0, 1, ..., q− 1. The variables are called cultural features and q is the number of the possible traits allowed per feature. They are supposed to model the different “beliefs, attitudes and behavior” of individuals. In an elementary dynamic step, an individual i and one of his neighbors j are selected and the overlap between them



where is the kronicker’s delta. So with the probability the interaction occurs if one of the features for which traits are different (σf(i) ≠ σf(j)) is selected and the trait of the neighbor is set equal to σf(i), otherwise nothing happens.

*Bounded confidence model*

The original model considers a population of N individuals, where each individual i holds a continuous opinion xi ∈ [0, 1]. This opinion can be considered the degree by which an individual agrees or not to a certain position. According to bounded confidence, two agents must not have significantly different opinions in order to build up a successful interaction. Individuals are connected by a complete social network, and interact pairwise at discrete time steps. The interacting pair (i, j) is selected randomly from the population at each time point t (i.e., at each time step, a randomly selected agent discusses with one of its neighbors on the social graph, also chosen at random). After interaction, the two opinions, xi and xj may change, depending on a so called bounded confidence parameter ε ∈ [0, 1]. This can be seen as a measure of the open-mindedness of individuals in a population where large ε produces consensus and small ε yields polarized opinions. It defines a threshold on the distance between the opinion of the two individuals, beyond which communication between individuals is not possible due to conflicting views. The *bounded*nature of agents' confidence in this model is simply this: if an agent *i*is very confident in its own opinion, that is reflected in narrow bounds of tolerance on the difference in that agent's opinion and the opinion of another agent, for the other agent to influence the opinion of agent*i*; with less confidence, the bounds widen. If dij = |xi − xj| stands for the distance between opinions xi and xj, then information is exchanged between the two individuals only if dij ≤ ε, otherwise nothing happens. The model is based on attractive dynamics, i.e. the exchange of information results in the two opinions becoming closer to one another, modulated by a convergence parameter μ ∈ (0, 0.5] (Deffuant et al., 2000)

https://journals.plos.org/plosone/article/file?type=thumbnail&id=info:doi/10.1371/journal.pone.0213246.e001

Deffuant’s model is based on a compromise strategy: after a constructive debate, the positions of the interacting agents get closer to each other, by the relative amount μ.

*Relative agreement polarization model*

Deffuant et al. (2002) proposed the relative agreement model as an extension of bounded confidence model where pair of agents are selected at random to interact, in which, both agents present their opinions that are next updated according to the opinion of the other. While in bounded confidence model, an agent would only consider another agent’s opinion if it falls within the bounds of its own opinion.

More formally, each individual *i*possesses two variables; an opinion *xi*and an uncertainty*ui*, both of which are represented by a real numbers. Let the opinion of an individual be between 1.0 and -1.0 and its uncertainty be between 0.0 and 2.0. Randomly paired agent interactions are run until a stable population opinion state is reached. An agent interaction is calculated by first calculating the relative agreement between agents *i*and *j*by taking the overlap between the two agents' bounds*hij*, given by:

|  |  |
| --- | --- |
| *hij* = min (*xi* + *ui*, *xj* + *uj*) -max(*xi*-*ui*, *xj* -*uj*) |  |

Followed by subtracting the size of the non-overlapping part given by *2ui - hij*

So the total agreement between the two agents is given as:

|  |  |
| --- | --- |
| *hij* - (2*ui* - *hij*) = 2(*hij* - *ui*) |  |

Once that is calculated, the relative agreement is then given by:

|  |  |
| --- | --- |
| 2(*hij* - *ui*) / 2*ui* = (*hij*/ *ui*) - 1 |  |

Then if *hij* > *ui*, then update of *xj*and *uj*is given by:

|  |
| --- |
| *xj* := *xj*+ *μ*[(*hij*/ *ui*) - 1](*xi - xj*) |
| *uj* := *uj* + *μ*[(*hij*/*ui*) - 1](*ui - uj*) |

where *μ* is a constant parameter that is responsible for controlling the speed of population convergence. When *μ* is low the population is less responsive to the opinions of others and when *μ* is high, a population will reach a convergence very quickly.

*Structured balance model*

Structured balance theory (Feather, 1967) posits that the balance between trust and mistrust that dictate people’s opinions to become closer or further apart, respectively, plays a major role in the dynamical process of opinion separation. It uses signed graphs to describe the trust and mistrust relationships among the interacting social agents. An example of opinion update under structured balance model known as affine boomerang model is described below.

Given  a signed graph. Assume that each agent has an initial opinion  and self-weight 

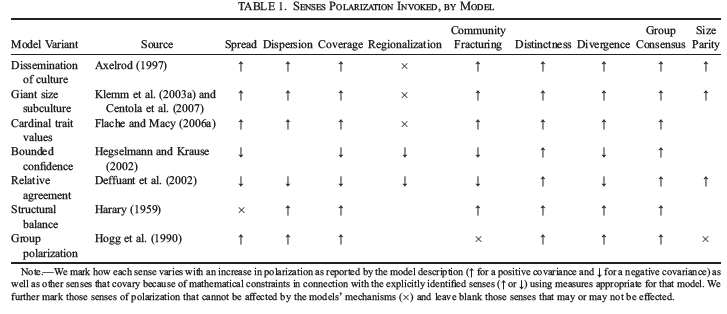
At each time step t, select randomly an edge of G assuming that each edge {i,j}, has a time invariant selective positive probability . Then the opinion of the two agents i and j are updated according to



*Hogg’s Group polarization*

Research on norm formation and group polarization shows that attitudes are responsive to social context. People use others’ attitudes, particularly when they share a social identity with those others, to construct a group norm that specifies what attitudes are normative. People use this normative information to configure their own attitudes. Group members deduce the content of a social identity from shared membership in a social category and the wider social context of intergroup relations (Postmes, Haslam, & Swaab, 2005).

According to Hogg et al. (1990), polarization is understood as a conformity phenomenon in which individuals who identify with a group conform, through the process of self-categorization, to the local norm which best represents the group. Whether polarization or convergence occurs, and the direction and extent of the polarization which occurs is governed by the salient social frame of reference which determines the position of the ingroup norm. Besides, self-categorization theory predicts a main effect for both independent variables: group polarization toward risk on risky items and cautious frames of reference, toward caution on cautious items and risky frames of reference, and no polarization at all on neutral items and neutral frames of reference. On the other hand, persuasive arguments theory would also predict an item-type effect in which there is polarization to risk on risky items and to caution on cautious items, but no effect at all, or an opposite effect, for the frame of reference manipulation. Because there are no actual informational differences between frame of reference conditions, persuasive arguments ought to predict no effect. Hogg et al. (1990) employed a simple indicator for polarization; namely, the degree of polarization of the prototype was defined as the ratio (prototype/mean discrepancy), and the degree of polarization of the estimated group consensus as the ratio (mean/consensus). This distinguishes it from other (more mathematical) polarization models while a rather a testing hypothesis like approach was employed.

****

## Other opinion dynamic based polarization models

The previous aforementioned opinion formation based models rely on continuous opinion value, nevertheless, there is also a growing interest into binary opinion representation. This includes Sznajd’s model (Sznajd and Weron, 2005) where the agents are assigned binary weights so that if the agents of the pair share the same opinion, they successfully impose their opinion on their neighbors. Otherwise, if the two agents disagree, each agent imposes its opinion on the other agent’s neighbor. While opinions are updated in a random sequential order. In essence, the dynamics of Sznajd’s model on a linear chain can be summarized by the simple sentence “just follow your next-to-nearest neighbor”.

Sznajd model has been applied in describing voting behavior in elections (Bernardes et al.,2002; Gonz alez et al., 2004); comprehending interaction of economic and personal attitudes of individuals, which evolve according to different rules but in a coupled manner (Sznajd-Weron, 2005).

(Hegselmann and Krause, 2002) (HK) describe another bounding box model where the update rule is modified such that agent i does not interact with one of its compatible neighbors, like in Deffuant’s model, but with all its compatible neighbors at once. Especially, Deffuant’s prescription is suitable to describe the opinion dynamics of large populations, where people meet in small groups, like pairs. In contrast, HK rule is intended to describe formal meetings, where there is an effective interaction involving many people at the

same time.

Other models focused on the importance of randomness in the process of opinion formation. (Jiang et al., 2007, 2008) proposed a model where an agent surrounded by a majority of neighbors with equal opinion will flip its opinion with the usual Metropolis probability exp(−ΔE/T), where ΔE is the increase of the Ising energy due to the flip and T the (social) temperature.

In (Bartolozzi et al., 2005) a model with binary opinions, evolving according to a heat bath dynamics, is proposed. The opinion field acting on a spin is given by a linear combination with random weights of a term proportional to the average opinion of its nearest neighbors on the social network, with a term proportional to the average opinion of the whole network.

In (Kuperman and Zanette, 2002), opinions are affected by three processes: social imitation, occurring via majority rule, fashion, expressed by an external modulation acting on all agents, and individual uncertainty, expressed by random noise.

Laguna et al. (2005) explored how the opinion dynamics is influenced by the hierarchical structure in societies/organizations. Bagnoli et al. (2004) investigated the interplay between opinions and personal taste.

As extension of Axelrod’s model, the possibility of one individual to change spontaneously one of his traits, independently of his neighborhood, is denoted as “cultural drift” in social science and corresponds to the addition of flipping events driven by random noise. In this respect, Klemm et al. (2003) showed that the inclusion of noise at rate r has a profound influence on the model, resulting in a noise-induced order-disorder transition independently of other parameters.

Friedkin and Johnsen (1990) opinion model allows for accounting of both the internal opinion and an expressed opinion of each agent, which enables accounting for any prior or non-rational opinion in the overall setting of Axerold’s opinion update.

In (Kuperman, 2006) two modified Axelrod-like dynamics have been introduced, where the interaction between individuals is also influenced by which trait is adopted by the majority of agents in the local neighborhood. Klemm et al., (2003) investigated the effects of complex interaction topologies like small-world and scale-free graph configurations. The incorporation of ideas from reinforcement learning in the context of opinion dynamics has been explored by Banish and Olbrish (2018).

Other models rely on the assumption of negative social influence and models that draw upon ideas from persuasion theory (cf. Mäs and Bischofberger (2015)). This seeks an explanation of polarization patterns by assuming a negative social influence in the interaction of distant agents such that the encounter and communication of two agents with very different views leads them to adopt even more distant positions in the opinion space.

Other persuasion models generally assume that communication partners exchange arguments about the object on which an attitude is formed and that new arguments are learned from an interaction partner if they are in support of an agents view. For instance, the model by Mäs and Flache (2013) is based on an explicit representation of attitudes borrowed from expectancy-value theory which treats an attitude as a weighted combination of a set of pro- and con-arguments regarding the underlying issue.

# Causes of polarization

## Social comparison versus persuasive argument theories

The majority of research evidence has come to favor two main explanations for the group polarization phenomenon, the social comparison and the persuasive arguments, which, according to Young and Chiu (2011), both types of processes are necessary to account for the explanation of the group polarization

phenomenon.

In **social comparison explanation**, individuals want to be perceived favorably by other group members, and also to perceive themselves favorably when they are in a group. Grodzki (2011) contend that group members are continuously concerned about their status in the group and are more motivated for approval by others. This means that once an individual hears what others believe or suggest during group discussion, they then adjust their positions in the direction of the dominant positions by the other group members (Boyer, 2012). From this perspective, the group discussion offers group members the opportunity to compare their positions and this creates group polarization.

The dynamic behind the social comparison explanation is that most people may want to take a position of a certain socially preferred sort takers, and their choice of position is partly a product of this desire. Although, it is acknowledged that one cannot know what such a position would be until the positions of others are revealed, which motivate individuals to take risk-free initiatives to move their judgments in order to preserve their image to others and their image to themselves.

The core affect theory (Russell and Barrett, 1999) through a set of empirical studies (.e.g., Gonzalez-Bailon et al., 2013) in computational social psychology suggest that emotions foster opinion polarization.

From the **persuasive argument** perspective, individuals freely exchange arguments that are available to them during group discussions (Mercier & Sperber, 2011). Novel arguments, bringing up fresh points, are especially likely to be persuasive. Therefore, group members make their final decisions based on the strength or persuasiveness of the revealed arguments so that the arguments are collectively shared during group interaction (Zhu, 2009). This based on the intuition that any individual’s position on an issue is partly a function of which arguments presented within the group seem convincing. The choice therefore moves in the direction of the most persuasive position defended by the group, taken collectivity. Because a group whose members are already inclined in a certain direction will have a disproportionate number of arguments supporting that same direction, the result of discussion will be to move individuals further in the direction of their initial inclinations.

## Factors influencing group polarization

* Personality of the group members, which include individual’s self-esteem, sociability, likeability, level of authoritarianism (Blamey, McCarthy & Smith, 2000).
* Gender, age, especially when linked to social relation and cultural norms where, for instance, male members are more likely to play prominent roles throughout deliberations than female members, while senior member are deemed more respected that youngsters.
* Background of the group members, which reveals the level of expertise, and therefore confidence in the underlying argument.
* Characteristic of messages in the sense that the characteristics of the message determine its persuasiveness. Essential characteristics of a persuasive message include whether the message is true, valid relevant or appropriate (Kaptein & Eckles, 2010). Another factor is the message framing, that is, whether the message is framed positively or negatively (White,MacDonnell & Dahl, 2011). It has been found that, messages that are negatively framed, are more persuasive than messages that use positively framed arguments in influencing people to change their decisions.
* Message source, which include expertise, trustworthiness, attractiveness and credibility as pointed out in Soliha and Dharmmesta, (2012)
* Effects of prior knowledge and experience. Especially, prior knowledge and values can also function as biases in the panel deliberations if the members are not prepared to be reflective in their deliberations

*Homophily impact*

Empirical studies have shown that homophily, i.e., greater interaction between like-minded individuals, results in polarization.

In social psychology, polarization is related to the phenomenon of biased assimilation, according to which individuals process new information in a biased manner whereby they readily accept confirming evidence while critically examining disconfirming evidence. Individuals are more likely to carefully consider disconfirming evidence if it is presented by a source that is otherwise similar to them. Centola (2011) empirically showed that individuals are much more likely to adopt health behaviors when they are a part of more homophilous networks.

## Impact of Computer mediated Communication

As the progress of information technologies, the use of computer-mediated communication has shown to be able to make a difference in group polarization (El-Shinnawy and Vinze 1998). With computer-mediated communication, people engage in a group discussion with less social presence, compared with traditional face-to-face communications (Rice 1993). Seigel et al. (1986) showed that dispersed group communications in an electronic form have greater group polarization than face-to-face group communications. Their studies have shown that when group members are in a distributed or anonymous environment, group discussions can lead to even higher levels of group polarization compared to traditional face-to-face meetings. Two motivation grounds have reported for this purpose. First, greater numbers of novel arguments are generated in computer-mediated communications, and second, there is higher incidence of one-upmanship behaviors.

Computer-mediated communication is shown to impact group polarization (El-Shinnawy and Vinze 1998) with reduced social presence (Rice 1993). Siegelet al. (1986) show that in computer-mediated communications, where groups might communicate in a more dispersed environment, group polarization is intensified compared with the traditional face-to-face communications. Two explanations are given in prior work. First, when a group of members work together via computer-mediated communications, it is more likely that high-quality and interesting thoughts could be produced (Valacich et al. 1994). Moreover, in computer-mediated communications people are able to communicate in an anonymous way, which leads to a stronger group polarization because the exchange of social cues is restricted (Connoly et al. 1990) and social presence is again reduced. While there has been a growing interest in studying computer-mediated communications and group polarization, little is investigated about how one particular type of computer-mediated communication, online forums or message boards, affects group polarization.

## Impact of online discussion forums

The innovation of the Internet technologies has encouraged more and more online forum and message board communications nowadays, changing how information is presented to people and the ways in which people can respond to that information. For instance, stock message boards, or stock forums, provide an unprecedented opportunity for investors to invest, debate, and exchange stock information and personal opinions with reduced or even no previous social connections. Statistics show that the level of forum activity keeps growing: during the year 1999 to 2001, more than 35 millions messages about public American firms were posted and discussed on Yahoo! Finance (Antweiler and Frank 2002).

In contrast, to face-to-face or computer mediated communication settings, relatively little research is conducted to explore polarization or group polarization in such online forums and virtual communities.

Nevertheless, the challenges and difficulties of exploring online forum communication cannot be ignored, Research has identified three aspects of such difficulties: noise, timeless interaction and complex topical dynamic. First, message board postings are known to have significant noise, contradictory opinions, rumors, and manipulations. The interactions among forum participants and online messages hence result in even more complicated communications. Second, the activity and scope of online forum communication are free of time and space constraints, leading to an enormous volume of information that can overwhelm forum participants. This also leads to the difficulty of identifying the form and boundary of communications among people. Third, the dynamic and complicated natures of online forums make it more difficult to identify factors and to develop measures for group discussions and communications, compared with other types of traditional or computer-mediated communications in prior work, where experiments can be more controlled.

Another important aspect of source of polarization fueled by the intense exposure to social media and online forums is the effect of echo-chamber that makes the user interacting only with those he/she likes, which often hide the overall picture and raises the possibility of misguided by rumors and false news.

A generic graphical representation of the polarization causes is highlighted in Fig. 9

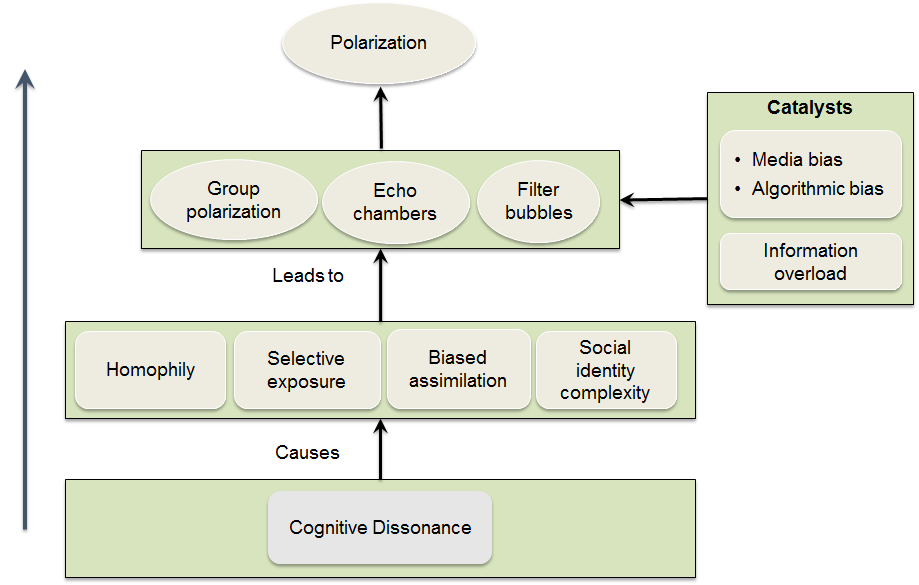
****

Fig. 9. Graphical illustration of some polarization causes

# Open questions, challenges for YOUNGRES community and ways forward

## Mathematical constructions of polarization measures

So far the axiomatization of polarization measures have dominantly been performed in econometric, while one notices full absence of such trend in social network analysis or party polarization measures. This calls for further research in order to narrow the gap between the various disciplines involved in polarization science. In this respect, bringing aspects of graph theory, topology and possibly random graph (the latter is deemed interesting for robustness and noise sensitivity) would enable us to identify relevant properties that can be turned into axiom (s) for characterizing some of the social network like polarization measures and/or party polarization measures.

It is for instance of interest to investigate the underlying algrebraical properties and any asymptotic behaviors of the current polarization measures, especially, with respect to increase of population, group size, repetitive behaviors, seldom behaviors, among others.

Such analysis can benefit from functional analysis and asymptotic system behaviors in control systems.

## Relevancy of Statistical models

The challenges of using statistical physics models for modeling social phenomena cannot be ignored. Especially, the tractable models are usually oversimplified and too general and thus lack flexibility required to reflect the features of a particular social phenomena. Moreover, while computational studies are rigorous in investigating the specific mechanisms and dynamics of polarization they often lack empirical foundation and thus it remains often unclear to what extent these often abstract models accurately represent phenomena we see in real world.

Therefore, statistical physics model should be complemented by both intensive simulations and empirical evidence through real testing and longitudinal infield observations

## Cascade effects

The social influence explanation invokes factors similar to those that underlie the reputational cascade. A major difference is that the social influence explanation concerns presentation to self as well as presentation to others. Note also that group polarization may or may not be a result of any cascade effect; the question is whether the accumulation of views from others operates in the form of a cascade. Existing work on group polarization does not answer such question.

Although, the phenomenon of cascade as route to disseminate information in social networks is well-known, its impact on polarization is strategically very interesting. Especially, it provides insights to both the exponential increase of (group) polarization, and, at the same times, provide clues for the equally interesting phenomenon of depolarization. Therefore cascade effects can be employed to develop appropriate approach for depolarization.

## Iterated Polarization Games

The logic of group polarization suggests that if participants engage in repeated discussions – if, for example, they meet each month, express views, and take votes – there should be repeated shifts toward, and past, the defined pole. In such iterated “polarization games,” deliberation over time should produce a situation in which individuals hold positions more extreme than those of any individual member before the series of deliberations began.

Although, it is acknowledged that iterated polarization games seem to be an important real-world phenomenon, it is still open why and when do groups stop polarizing? How they can shift in the opposite direction? Adequate answer to such questions is still missing in literature. For instance, social cascades can lead to change in direction as a result of external shocks, as through the release of new information, possibly, the same processes seem to terminate or to reverse group polarization.

The multiple interaction effect may also ultimately be linked to Nash equilibrium in game theory and can provide insights into predicting the phenomenon of polarization in society with regard to the impact of time evolution and external factors.

## Rhetorical Asymmetry

Group polarization may be aggravated or attenuated if one or another side has a systematic advantage in rhetoric. Often existing cultural norms are the primarily source of any rhetorical asymmetry that can be observed with the process of group decision making, and thereby, polarization. Besides, such norms vary over space and time, producing shifts in rhetorical asymmetry. For instance, legislative judgments about criminal punishment can involve such asymmetry. In other group organization, the presence of members that systematically favor lower taxes, more aid for scholarship students, greater funding for environmental protection may yield similar asymmetry. Much remains to be explored in this context as pointed out by social psychology scholars. Perhaps the most striking implication is that when there is an initial distribution of views in a certain direction, and when a more extreme movement in that direction has a rhetorical advantage, quite extreme shifts can be expected.

More formally, such asymmetry can be captured using appropriate prior function that allows for such information in subsequent inference like reasoning.

## Role of communication

Over the past few years, concerns about the negative societal consequences of the online spread of misinformation and propaganda have become widespread. New technological tools that allow anyone to easily broadcast political information to large numbers of citizens can lead to a more pluralistic public debate, but they can also give a platform to extremist voices and actors seeking to manipulate the political agenda in their own financial or political interest (Tucker et al. 2017). Attention to this problem spiked after the 2016 U.S. presidential election, during which “fake news” was widely shared on social media and reached large numbers of citizens, propagated at least in part by foreign actors (see e.g., Shane 2017). Although there is broad scholarly agreement regarding the high prevalence of misinformation and propaganda in online platforms regardless whether it has an impact on political outcomes, trust in overall system remains an open question.

Tucker et al. (2018) identified 8 research gaps in this matter.

(1) better estimates of the effects of exposure to information and disinformation online;

(2) cross- and multi-platform research;

(3) disinformation spread through images and video;

(4) the generalizability and comparability of governmental findings;

(5) the role of ideological asymmetries in mediating the effect of exposure to disinformation and polarization; (6) the effects of new laws and regulations intended to limit the spread of disinformation;

(7) better understanding of the strengths and weaknesses of different methods of bot detection and analysis;

(8) the role of political elites in spreading disinformation.

The illustration of Fig. 10 highlights some of the interaction among the key patterns of political polarization and various facets of the communication channels.

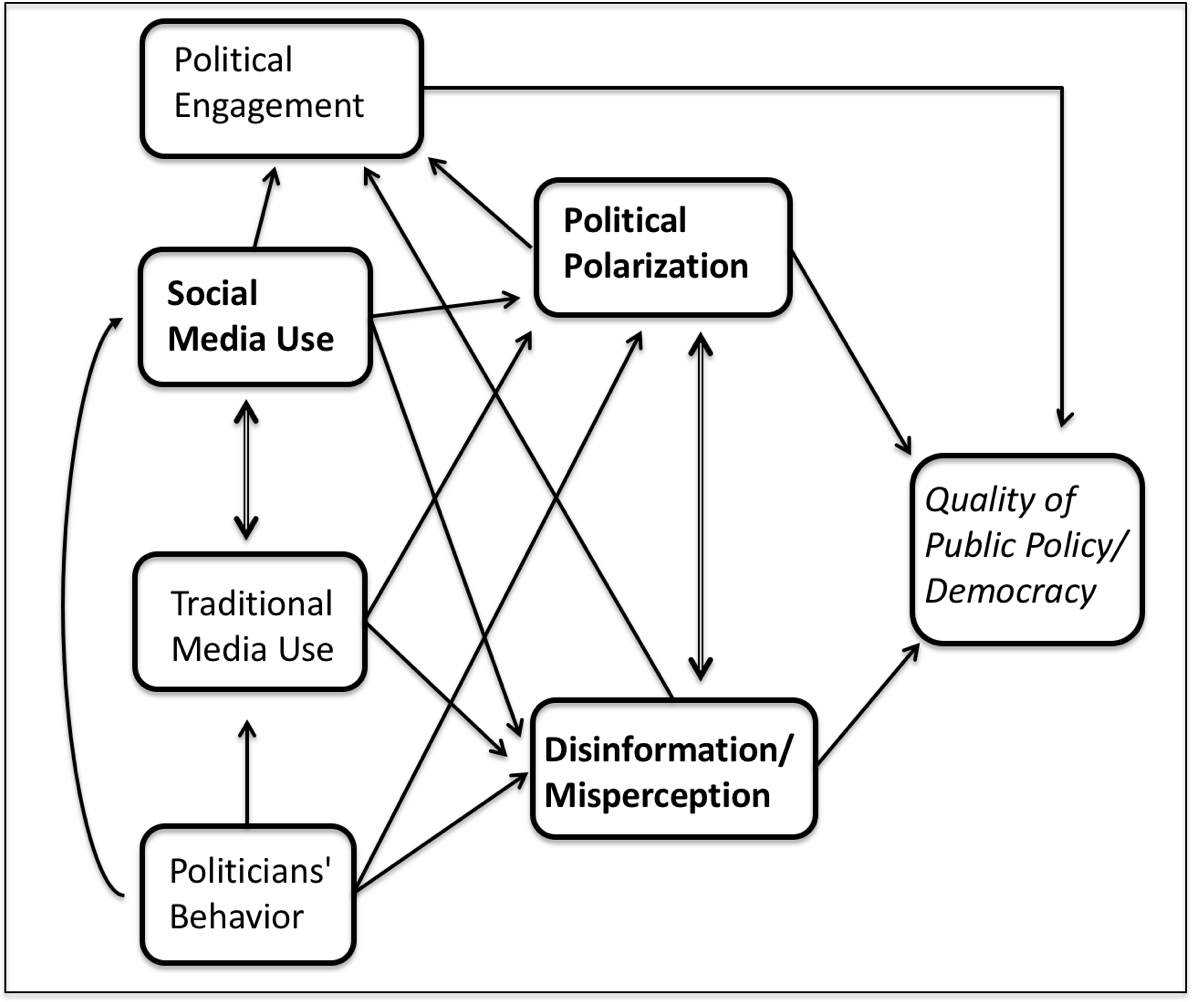


Fig. 10. Communication facets in the polarization

## Social media and online forum polarization

Often users communicate online with those they share values and thoughts. Therefore, the extent to which people are regularly exposed to disagreement, either cross-partisan interactions or some other mechanism, remains an open question, which is worth investigating by itself.

Likewise, it is scientifically interesting to identify same users who move across partisan boundaries in order to comprehend the polarization phenomenon in social network. Do these cross-cutting discussions occur primarily via existing relationships or via “weak ties”—for example, friends of friends? Such investigations are still on their infancy because of lack of relevant dataset and the anonymity paradigm that spread over social media community.

Another relevant aspect is related to the quality of the underlined online discussion, if any. For instance, Papacharissi (2004) questioned the extent to which online conversations are civil. A comprehensive analysis of Reddit found a marked increase in incivility since 2016 (Nithyanand et al. 2017), which may question any conclusions drawn from online discussion analysis.

Nevertheless, one should point out the fact that online Platforms are constantly changing their algorithms and business models, with effects on user behavior that can sometimes be large and close to offline behavior.

On the other hand, it has been reported in (Tucker et al. 2018) that the consumption of political information through social media increases cross-cutting exposure, which has a range of positive effects on civic engagement, political moderation, and the quality of democratic politics, but also facilitates the spread of misinformation.

## Level of interactions

Scholars (e.g., Franzmann, 2011) distinguishes four types of interactions among individuals: cooperation, negotiation, contest and conflict that can occur at several dimensions, see Fig 11.

Contest as well as cooperation presuppose a common goal, but the principal of contest is individualistic and other’s interests are not considered. In conflicts “actors enter into a social relationship in which they inflict damage on each other”. In contrast, negotiation characterizes actors’ diverging goals, but the own interests are partially subordinated. Therefore, the principle is still solidarity which is the common feature of negotiation and cooperation, whereas negotiation can be the starting point of cooperation.

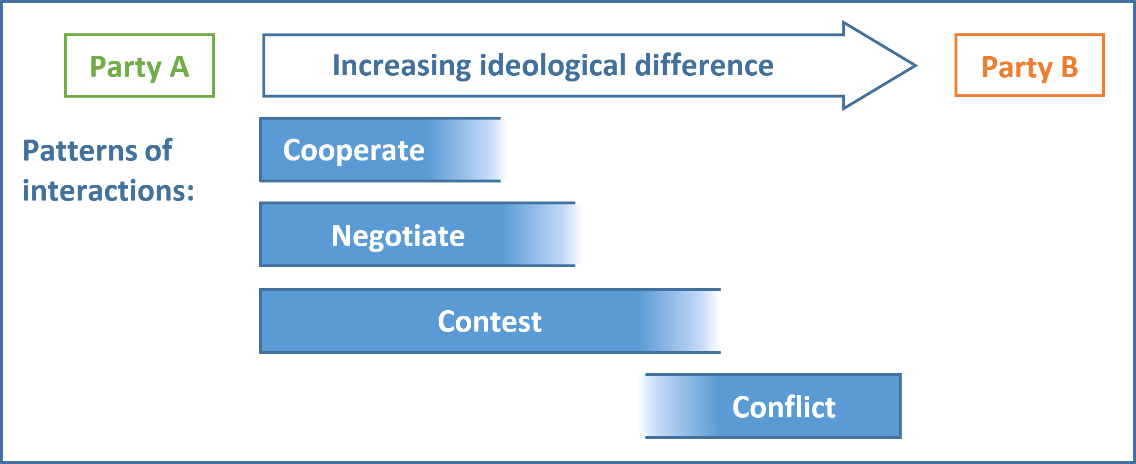


Fig. 11. Ideological differences and patterns of interaction

It is therefore ultimately interesting to investigate the effect of these four interaction patterns on group polarization. This is of interest to both economical studies as well as social psychology where the patterns govern the bargaining process (economics) and individual/group behavior formation process (social psychology), providing another angle to view express the polarization measure.

## Brandsma’s polarization model promoted by EU RAN specialists

In the context of individual and/or group radicalization resulted from polarization process, the EU RAN specialists promote the Brandsma’s framework promoted by Dutch philosopher Bart Brandsma (Brandsma, 1988), see also, <https://www.polarisatie.nl/eng-home>, as an efficient tool to depolarize the group and /or individual. The essence of the polarization model is built upon *three rules* and *five roles*:

**Three rules:**

*1.* Polarization is a *thought construct*, ‘us-and-them’ thinking, based on identities and groups. Polarization is the spin, the story exploiting an incident or conflict.

*2.* Polarization *needs fuel*; it thrives on talking about identities in combination with judgment. If there is no communication, no energy put into the polarization, it will die out.

*3.* Polarization is about *feelings and emotions*. Facts and figures won’t get the job done in managing polarization.

**The five roles**

1. The *Pushers, positioning themselves at two opposing poles,* are trying to create polarisation, they are the instigators acting from the poles. They claim 100% truth and are more in the ‘send mode’ than the ‘listen mode’. They don’t want a real dialogue.

2. The *Joiners* have chosen sides and moved towards the pushers; this is polarisation unfolding.

3. The *Middle ground*, the nuanced, have not chosen sides, and not taking part in polarisation. They could be neutral, scared or indifferent. They are targeted by the pushers; the game of polarisation is won or lost in the middle ground.

4. The *Bridge-builder* is trying to bring peace and moderation by reaching out to both opposing poles. But, by doing so, he or she underlines the existence of the two poles, and as a result might even be adding fuel to the polarisation.

5. The *Scapegoats* are being blamed or attacked, these could be the non-polarised ‘in the middle ground’ or the bridge-builders.

**Four game changers:**

1. Change *the target audience*. Pushers portray an enemy in the other pusher but target the middle , ground, that is where the actual is taking place . So, target the middle ground for depolarisation.

2. Change *the topic*. Move away from the identity construct chosen by the pushers and start a conversation on the common concerns and interests of those in the middle ground.

3. Change *position*. Don’t act above the parties, but move towards the middle ground;

4. Change *the tone*; this is not a question of facts being right or wrong. Use mediating speech, try to engage and connect with the diverse middle ground.

**Polarisation – conflict**

Brandsma stresses the importance of making a distinction between polarisation and conflict. Polarisation is an artificial construction of identities. For conflict, there must have been an incident, perpetrators and victims. Something has been broken or stolen, or there are people wounded. People are directly involved and are fighting over something. They are problem owners.

Polarisation is about people who are being targeted by narrow identity communication and pressured to choose sides. Pushers try to lure them into polarisation. The definition of the problem and problem ownership is not very clear. Are all Muslims part of a conflict if an attack takes place that was claimed by Daesh? Or are some actors trying to pull people into a ‘black and white identity construct’? When conflict occurs, we can rely on a long tradition of mediation and peace building. The dynamics of polarisation are however different. Familiar elements from conflict mediation, like *mediating speech* and *mediating behaviour,* are helpful, but polarisation also needs other approaches. This is where the above mentioned three rules, five roles and four game-changers come into play.



Fig. 12. Brandsma’s polarization model

**Ways forward**

Despite the practical popularity of Brandsma’s framework as a tool and a set of recommendations that help in detecting and dismantling polarization in groups or individuals, its mathematical and theoretical foundations with respect to social psychology and well established metrics for polarization measures are ill-known and still to be investigated. This can constitute part of future work within YoungRes project.

**References**

Ackoff, R. L. (1967). Management misinformation systems. *Management Science*, 14(4), 18-21.

Antweiler, W. & Frank, M. (2004). Is all that talk just noise? The information content of Internet Stock Discussion Boards. *Journal of Finance*, 59(3), 1259-1295.

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. The Journal of Conflict Resolution, 41(2):203–226.

Banisch, S., Araujo, T., and a, J. L. (2010). Opinion dynamics and communication networks. Advances in Complex Systems, 13(1):95–111.

Banisch S., and Olbrish E. (2018), Opinion polarization by learning from social feedback., arXiv:1704.02890v3 [physics.soc-ph]

Bartolozzi, M., D. B. Leinweber, and A. W. Thomas (2005), Stochastic opinion formation in scale free network, Physical Review E. 72(4), 046113

Batson, C. D. (1975). Rational processing or rationalization? The effect of disconfirming information on a stated religious belief. *Journal of Personality and Social Psychology, 32,* 176–184. doi:10.1037/h0076771

Bernardes, A. T., D. Stauffer, and J. Kert´esz (2002),Election results and the Sznadj model on Barabasi network, European Physics J. B 25, 123-127

Boyer, L. (2012). Managing conflict: What to do and how to do it. Bradenton, Sarasota FL Leadership Options, LLC

Bramson et al. (2016), Disambiguation of social polarization concepts and measures, Journal of Mathematical Sociology, 40 (2), 80-111.

Brockner, J. (1992). The escalation of commitment to a failing course of action: Toward theoretical progress. *Academy of Management Review*, 17(1), 39-61.

Brown, R. (1965). Social psychology. New York: Free Press.

Burnstein, E. (1982). Persuasion as argument processing. In H. Brandstatter, J. H. Davis, & G. Stocker-Kreichgauer (Eds.), *Group Decision Making* (pp. 103-124). New York: Academic Press.

Butler, J. & K., Crino, M. D. (1992). Effects of initial tendency and real risk on choice shift. *Organizational Behavior and Human Decision Processes*, 53(1), 14-34.

Castellano C., Fortunato, V. Loreto (2009), Statistical physics of social dynamics. Review of Modern Physics 81, 591-646. doi:10.1103/RevModPhys.81.591

Centola D (2011) An experimental study of homophily in the adoption of health behavior. Science 334(6060):1269–1272.

Chacoma, A. and Zanette, D. H. (2015). Opinion formation by social influence: from experiments to modeling. PLoS ONE, 10(10):e0140406.

Chen, C.K., Gustafson, D. H., & Lee, Y. D. (2002). The effect of quantitative decision aid-analytic hierarchy process-on group polarization. Group Decision and Negotiation, Volume,11, 329-344.

Coleman, J. S. (1957). Community conflict. New York: Free Press.

Connolly, T., Jessup, L. M., & Valacich, J. S. (1990). Effects of anonymity and evaluative tone on idea generation in computer-mediated groups. *Management Science*, 36(6), 689-703.

D’Ambrosio, C. (2001) Household characteristics and the distribution of income in Italy: an application of social distance measures. The Review of Income and Wealth **47**(1), 43–64

Dalton, R. J. (2008). The quantity and the quality of party systems: Party system polarization, its measurement, and its consequences. Comparative Political Studies, 41, 899-920

Das, S. R., & M. Y. Chen. (2001). Yahoo! for Amazon: Sentiment parsing from small talk on the web. In the *Proceedings of 8th Asia Pacific Finance Association Annual Conference*. July 22-25, Bangkok, Thailand.

Deffuant, G., Neau, D., Amblard, F., and Weisbuch, G. (2000). Mixing beliefs among interacting agents. Advances in Complex Systems, 3(01n04):87–98

DeGroot MH (1974) Reaching a consensus.J Am Stat Assoc 69(345):118–121

DiMaggio, P., Evans, J., & Bryson, B. (1996). Have Americans’ social attitudes become more polarized? American Journal of Sociology,102, 690–755. doi:10.1086/230995

Duclos, J.-Y., Esteban, J., Ray, D. (2004) Polarization: Concepts, measurement, estimation. Econometrica **72**(6), 1737–1772

El-Shinnawy, M. & A. S. Vinze. (1998). Polarization and persuasive argumentation: A study of decision making in group settings. *MIS Quarterly* 22(2), 165-198.

Esteban, J., Ray, D. (1994) On the measurement of polarization. Econometrica **62**(4), 819–851

Esteban, J., Ray, D. (1999) Conflict and distribution. Journal of Economic Theory **87**, 379–415

Feather N. T. (1967), A structural balance approach to the analysis of communication effects. Advances in Experimental Social Psychology, 3:99–165

Ezrow, L. (2007). The variance matters: How party systems represent the preferences of voters. Journal of Politics, 69, 182-192

Esteban, J., Grad´ın, C., Ray, D. (1999) Extensions of a measure of Polarization with an Application to the income distribution of five OECD countries. Working Paper No. 218, Maxwell School of Citizenship and Public Affairs, Syracuse University, Syracuse, New York

Festinger, L., H. W. Riecken, & S. Schachter. (1956). When prophecy fails. Minneapolis, MN: University of Minnesota Press.

Foster, J.E. and M.C. Wolfson (2010), Polarization and the decline of the middle class: Canada and the US, The Journal of Economic Inequality, 2010, 8 (2), 247, 273.

Friedkin NE, Johnsen E (1990) Social influence and opinions. J Math Soc 15(3–4):193–206

Fromkin, H. (1970). Effects of experimentally aroused feelings of undistinctiveness upon valuation of scarce and novel experiences. *Journal of Personality and Social Psychology*, 16(3), 521-529

Garimella K, Morales GDF, Gionis A, Mathioudakis M (2016) Quantifying controversy in social media. In: ACM international conference on web search and data mining, WSDM, pp 33–42

Garrett RK (2009) Echo chambers online? Politically motivated selective exposure among internet news users1. J Comput Mediat Commun 14(2):265–285.

Gonzalez, M. C., A. O. Sousa, and H. J. Herrmann(2004), Opinion formation on a deterministic pseudo fractal network, International Journal of Modern Physics, C 15(1), 45-57.

González-Bailón, S., Borge-Holthoefer, J., Moreno, Y.( 2013). Broadcasters and hidden influentials in online protest diffusion. American Behavioral Scientist 57, 943–965

Grodzki, J. S. (2011). Role identity: At the intersection of organizational socialization and individual sense making of new principals and vice-principals. Canadian Journal of Educational Administration and Policy, Issue 127

Guerra PHC, Jr, WM, Cardie C, Kleinberg R (2013) A measure of polarization on social media networks based on community boundaries. In: International conference on weblogs and social media, ICWSM

Gradın, C. (2000) Polarization by sub-populations in Spain, 1973–91. The Review of Income and Wealth **46**(4), 457–474

Gu, B., Konana, P., Rajagopalan, R., & Chen, H. M. (2007). Competition among virtual communities and user valuation: The case of investing-related communities. *Information Systems Research*, 18(1), 68-85.

Hinsz V. B., & Davis J. H. (1984). Persuasive arguments theory, group polarization, and choice shifts. *Journal of Personality and Social Psychology*, 10(2), 260-268.

Hogg, M. A., Turner, J. C., & Davidson, B. (1990). Polarized norms and social frames of reference:A test of the self-categorization theory of group polarization. Basic and Applied Social Psychology, 11, 77-100.

Isenberg, Daniel J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personal and Social Psychology*, 50(6), 1141-1151.

Janis, I. L. (1989). Crucial decisions: Leadership in policymaking and crisis management. New York: Free Press.

Jiang, L.-l., D.-y. Hua, and T. Chen (2007),Opinion Dynamics on directed small word network, European Physical Journal, B. 65, 251-255

Jiang, L.-L., D.-Y. Hua, J.-F. Zhu, B.-H. Wang, and T. Zhou (2008), eprint arXiv:0801.1896.

Johansen, R. D., Sibbet, S. Benson, A. Martin, R. Mittman, & P. Saffo. (1991). Leading business teams: How teams can use technology and process to enhance group performance. Reading, MA: Addison-Wesley.

Kaplan, M. F. (1977). Discussion polarization effects in a modified jury decision paradigm: Informational influences. *Sociometry*, 40(3), 262-271.

Kuperman, M. N., and D. Zanette (2002), Stochastic Reasonance in a model opinion formation on small-word network, European Physical Journal B 26, 387-391

Latane, B. (1981). The psychology of social impact. *American Psychologist*, 36(4), 343-356.

Maes, Michael and Bischofberger, Lukas, Will the Personalization of Online Social Networks Foster Opinion Polarization? (January 21, 2015). Available at SSRN: <https://ssrn.com/abstract=2553436> or [http://dx.doi.org/10.2139/ssrn.2553436](https://dx.doi.org/10.2139/ssrn.2553436)

McCauley, C. (1989). The nature of social influence in groupthink: Compliance and internalization. *Journal of Personality and Social Psychology*, 57(2), 250-260.

McGrath, J. E. (1984). Groups: Interaction and performance. Englewood Cliffs, NJ: Prentice Hall.

Montalvo, J.G., and Reynal-Querol M. (2005):.Ethnic polarization, potential conflict and civil wars., American Economic Review, 95, 796-815.

Mosler, K., Schmid, F. (2005) Beschreibende Statistik und Wirtschaftsstatistik. Springer, Berlin

Myers, D. G. & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin*, 83, 602-627.

Pardos-Prado, S., & Dinas, E. (2010). Systemic polarisation and spatial voting. European Journal of Political Research, 49, 759-786.

Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. New Media & Society, 6, 259-283.

Permanyer I. (2008), The measurement of Social Polarization, Tehnical report, Institute Analsi Economica, (CSIC) University of Barcelona.

Postmes, T., Haslam, S. A., & Swaab, R. I. (2005). Social influence in small groups: An interactive model of social identity formation. European Review of Social Psychology, 16, 1-42.

Pruitt, D. G. (1971). Choice shifts in group discussion: An introductory review. *Journal of Personality and Social Psychology*, 20(3), 339-360.

Reynal-Querol M. (2002), “Ethnicity, Political Systems, and Civil Wars,” Journal of Conflict Resolution 46, 29–54.

Rice, R. E. (1993). Media appropriateness: Using social presence theory to compare traditional and new organizational media. *Human Communication Research*, 19(4), 451-484.

Russell, J. A., & Feldman Barrett, L. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. Journal of Personality and Social Psychology, 76, 805–819

Rodrik, D. (1999). Where Did All the Growth Go?: External Shocks, Social Conflict, and Growth Collapses. Journal of Economic Growth 4(4):385–412

Sanders, G. S. & R. S. Baron. (1977). Is social comparison irrelevant for producing choice shift? *Journal of Experimental Social Psychology*,13(4), 303-314.

Sartori, G. (1976). Parties and party systems: A framework for analysis. Cambridge, England: Cambridge University Press.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, 27, 379-423, 623-656.

Short, J., E. Williams, & B. Christie. (1976). The social psychology of telecommunications. New York: John Wiley.

Sia, C., Tan, B. C., & Wei, K. (2002). Group polarization and computer-mediated communication: Effects of communication cues, social presence, and anonymity. *Information Systems Research*, 13(1), 70-90.

Siegel, J., Dubrovsky, V.. Kiesler, S., & McGuire, T. W. (1986). Group processes in computer-mediated communication. *Organizational Behavior and Human Decision Processes*, 37(2), 157-187.

Sznajd-Weron K. (2005), Mean field results for the two component model, Physics Review E, 71, 046110,

Steiner, N. and C. Martin (2012). Economic Integration, Party Polarisation and Electoral Turnout. West European Politics 35(2):238–65.

Stoner, J. A. F. (1961). A comparison of individual and group decisions involving risk. Cambridge, MA: Unpublished master's thesis, MIT, School of Industrial Management.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, LXII(3), 1139-1168.

Tucker J. A.(2018) Social Media, Political polarization and political disinformation: A review of the scientific literature. Technical Report William & Flora Hewelett Foundation

Valacich, J. S., A. R. Dennis, & T. Connolly. (1994). Idea generation in computer-based groups: A new ending to an old story. *Organizational Behavior and Human Decision Processes*, 57(3), 448-467.

Vinokur, A., & E. Burnstein. (1978). Novel argumentation and attitude change: The case of polarization following group discussion. *European Journal of Social Psychology*, 8(3), 335-348.

Wagner, C., & Majchrzak, A. (2007). Enabling customer-centricity using wikis and the wiki way. *Journal of Management Information Systems*, 23(3), 17-43.

Walton, R. E., & R. B. McKersie. (1965). A behavioral theory of labor negotiation. New York: McCraw Hill.

Wang, Y.-Q., Tsui, K.-Y. (2000) Polarization orderings and new classes of polarization indices. Journal of Public Economic Theory **2**(3), 349–363

Wolfson, M.C. (1994) When inequalities diverge. The American Economic Review **84**(2), 353–358

Whyte, G. (1993). Escalating commitment in individual and group decision making: A prospect theory approach. *Organizational Behavior and Human Decision Processes*, 54(3), 430-455.

Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin*, 84(5), 963-976.

Wysocki, P. D. (1999). Cheap talk on the web: The determinants of postings on stock discussion boards. Working Paper 98025, University of Michigan, Ann Arbor, MI.

Zhang, X., Kanbur, R. (2001) What difference do polarisation measures make? An application to China. Journal of Development Studies **37**, 85–98

Zuber, J. A,, Crott, H. W., & Werner, J. (1992). Choice shift and group polarization: An analysis of the status of arguments and social decision schemes. *Journal of Personality and Social Psychology*, 62, 50-61