JASSS
THE JOURNAL OF ARTIFICIAL SOCIETIES AND SOCIAL SIMULATION

An Argument Communication Model of Polarization and Ideological Alignment

Sven Banisch[1] **and Eckehard Olbrich**[1]

[1]*Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany*
Correspondence should be addressed to *sven.banisch@UniVerseCity.de*

Journal of Artificial Societies and Social Simulation 24(1) 1, 2021
Doi: 10.18564/jasss.4434 Url: http://jasss.soc.surrey.ac.uk/24/1/1.html

Received: 25-02-2020 Accepted: 02-11-2020 Published: 31-01-2021

Abstract: A multi-level model of opinion formation is presented which takes into account that attitudes on different issues are usually not independent. In the model, agents exchange beliefs regarding a series of facts. A cognitive structure of evaluative associations links different (partially overlapping) sets of facts to different political issues and determines an agents' attitudinal positions in a way borrowed from expectancy value theory. If agents preferentially interact with other agents that hold similar attitudes on one or several issues, this leads to biased argument pools and increasing polarization in the sense that groups of agents selectively believe in distinct subsets of facts. Besides the emergence of a bi-modal distribution of opinions on single issues that most previous opinion polarization models address, our model also accounts for the alignment of attitudes across several issues along ideological dimensions.

Keywords: Argument Communication Theory, Opinion Dynamics, Polarization, Ideological Alignment, Belief Systems, Cognitive-Evaluative Maps, Attitudes

Introduction

1.1 Understanding the dynamics of political opinions is a complicated issue as individual preferences are influenced by various psychological and social processes (Asch 1955; Kelman 1961; Friedkin & Johnsen 1990; Duggins 2017). The interplay of these processes creates a complex landscape of political preferences with a spectrum of positions, nuances and idiosyncracies whose dynamics is difficult to express in formalised models. While agent-based modelling of opinion dynamics has been very successful in identifying social and cognitive mechanisms that lead to consensus, plurality and polarization, their application to more specific cases remains a challenge (Flache et al. 2017). The main objective of this paper is to take agent-based modelling one step further towards the empirical application to political opinion dynamics.

1.2 The space of political opinions has a multilevel structure. On the one hand side, we distinguish beliefs about facts from the attitude on a political issue formed by considering a whole series of such beliefs speaking in favor or in disfavor of a certain position. In line with expectancy-value theories of attitude research and measurement (Fishbein & Raven 1962; Fishbein 1963; Ajzen 2001), we assume that an attitude regarding a political issue is formed through a structure of evaluations regarding the different aspects that are relevant for the topic. Also communication about those topics makes reference to these underlying argumentative dimensions. On the other hand, we present a model in which different political issues may be discussed at the same time. These topics are not independent but related through the underlying cognitive-evaluative meaning structures. With that, our model formalizes inter-attitudinal structures on the basis of the underlying thematic consistency or inconsistency which is closely related to the classical meaning of *ideology* (e.g. Eagly & Chaiken (1998, 281) or Converse (1964, 8)). Depending on the nature of the relations in these evaluative schemata, different patterns of political preferences evolve that lead to an organization along politico-ideological dimensions such as left versus right or liberal versus conservative.

1.3 The model proposed in this paper is an extension of the *argument communication theory of bi-polarization* (ACTB) that has been proposed in Mäs & Flache (2013). The ACTB attempts to explain the emergence of a bi-modal opinion distribution in the sense that small initial opinion differences are amplified in processes of social

influence so that two antagonistic groups covering the extremes of an opinion spectrum emerge. The argument-based approach by Mäs & Flache (2013) combines ideas from persuasive argument theory of group polarization (Burnstein & Vinokur 1975, 1977; Isenberg 1986) with the assumption that homophily with respect to the opinions (Byrne 1961; Huston & Levinger 1978; McPherson et al. 2001; Wimmer & Lewis 2010; Bakshy et al. 2015) guides interaction and communication behavior. Opinions are based on a series of pro- and con-arguments that are exchanged in an interaction process. While the argument exchange process affects believes at the lower level of facts, attitude or value homophily (Lazarsfeld & Merton 1954; Byrne 1961) acts at the aggregate level of attitudes. As a consequence, homophily creates a tendency that actors interact with like-minded others and therefore are likely to be exposed to arguments that further support their current attitudinal inclination. Analogous to the *law of group polarization* described in Sunstein (2002), this leads to biased argument pools that may reinforce group opinions in the direction of an initial inclination. Apart from the social bias of homophily no further psychological biases are needed to explain the emergence of a bi-polar opinion distribution. Corresponding to the experimental finding that deliberation has a larger effect for less salient topics (Farrar et al. 2010), one could argue that models implementing ACTB are describing a situation at the beginning of a discourse, because at later stages all arguments might be already known to the participants.

1.4 ACTB as well as quite a series of other recent polarization models focus on the formation of a bi-modal distribution of opinions regarding a single issue (e.g. Dandekar et al. 2013; Friedkin 2015; Mäs & Bischofberger 2015; Duggins 2017; Banisch & Olbrich 2019). However, such strong patterns of opinion divergence with respect to single issues cover only one aspect of polarization (DiMaggio et al. 1996; Bramson et al. 2016). Moreover, it has been empirically found only for certain morally charged *polarizing* topics such as abortion in US public opinion (DiMaggio et al. 1996). Besides different forms of social polarization such as, for instance, increasingly antagonistic references between two groups of different identity (Uitermark et al. 2016; Mason 2018), also more specific notions of issue polarization have been put forth (DiMaggio et al. 1996; Bramson et al. 2016). The sorting of opinions regarding diverse sets of political goals along ideological dimensions as recently identified in opinion data on American public opinion (Dimock et al. 2014) is of particular relevance in our context. In fact, an organization of the complex landscapes of political preferences along a few axes such as left versus right, liberal versus conservative or technocratic versus ecological (Leuthold et al. 2007) can only acquire meaning if political preferences regarding a multitude of political goals are correlated over a population and constrained in form of a belief system or ideology (Converse 1964).

1.5 The main aim of the paper is to show that ACTB is easily extended to account for this kind of opinion sorting or ideological alignment as well. For this purpose, we extend the model to multiple issues which are cognitively related because some arguments are relevant for more than one issue. While a series of multi-dimensional models of opinion formation that explain persistent opinion plurality is available in the literature (Axelrod 1997; Macy et al. 2003; Urbig & Malitz 2005; Fortunato et al. 2005; Baldassarri & Bearman 2007; Lorenz 2008; Huet et al. 2008; Flache & Macy 2011), the interrelatedness of opinions on various issues has not been addressed within these models. We think that the incorporation of issue interdependence on the basis of underlying arguments can contribute to a better understanding of the formation of complex but at the same time specifically organized opinion landscapes. For this purpose we assume the existence of cognitive-evaluative maps (cf. e.g. Rosa 2016, 214) which encode the evaluative meaning of different beliefs with respect to the different issues of discussion. To simplify the interpretation of the dynamical behavior of the model, we consider that these evaluation structures have been acquired in the same socio-cultural context and are collectively shared by all agents. At the same time, this can be seen as a way to incorporate some notion of cultural meaning (Berger & Luckmann 1970; Strauss & Quinn 1997; Schütz & Luckmann 2017) into models of opinion dynamics, and we will discuss the possibility that different *cultures* with their specific evaluative schemata engage in an argument exchange process and relate this to recent approaches to infer shared belief systems from attitudinal data (Baldassarri & Goldberg 2014; Daenekindt et al. 2017). Our structure is inspired by structural theories of attitudes (Fishbein & Raven 1962; Fishbein 1963; Ajzen 2001) in which a systematic distinction between beliefs and evaluations is made. In our model we consider the evaluations as externally given and model the exchange of arguments about a set of beliefs that are relevant in the thematic complex at question. As the same beliefs may contribute positively or negatively to different issues, the evaluative structure already imposes constraints on the admissible positions in the opinion space. Only certain combinations of opinions are possible and the argument exchange process of the ACTB can inform us about the likeliness of specific configurations especially in conditions of polarization.

1.6 In this paper, we concentrate on the capacity of the model to account for the emergence of coherent bundles of opinions. This is motivated by the fact that – implicitly or explicitly – spatial theories of political opinion and voting (Downs 1957; Leuthold et al. 2007; Laver 2014) rely on such an ideological ordering of political preferences. There exist already models that try to capture the opinion dynamics in such political or ideological spaces (Sznajd-Weron & Sznajd 2005; Laver 2005), but because the dynamical processes which may lead to such issue alignments were not explicitly modeled they had to rely on ad-hoc assumptions for the relationship between

the dynamics on the different dimensions of the space and were not able to incorporate semantic changes regarding the meaning of the axes. We propose a model that explains opinion alignment on multiple issues on the basis of the argumentative interrelatedness of different issues and the cognitive constraints this imposes. To analyze the principal effects that structured attitudes can bring about, we perform a series of simulation experiments for different prototypic settings in the case of two issues (Section 3). As attitudes in two-issue models are distributed in a two-dimensional opinion space (two judgements on a seven-point scale ranging from -3 (extreme disfavor) to +3 (extreme favor) in our cases) there are different possibilities to define *opinion distance* and consequently homophily. We explore four different modes including the case where different opinions with respect to a single *ideologically loaded* issue determines if arguments are adopted from an interaction partner or not. Also the case that certain beliefs operate as *identity signals* (Bacharach & Gambetta 2001) indicating to which ideological subgroup one belongs is considered. In this way, we obtain a systematic picture of the kind of opinion correlations that may be induced between two issues that are compatible or incompatible in the light of a series of facts.

1.7 At the same time, however, the model is part of a larger research agenda that aims to address real debates around specific topics using opinion dynamics approaches. We provide an example application related to climate change and electricity production inSection 4 to illustrate this potential of linking opinion dynamics more closely to empirical data on political opinion. This shows that ACTB is a very useful starting point for developing models that go beyond the reproduction of stylized facts for it allows to represent the content dimensions and arguments of real debates. Addressing the question of issue alignment with reference to the underlying argumentative dimensions, allows to relate coherent patterns of opinions to the ways different actors talk about the issues. Recent advances in precision language processing (Steels 2017; Van Eecke & Beuls 2018) and argument mining (Lippi & Torroni 2016) will afford new opportunities to develop empirically-grounded scenarios in which model assumptions can be tested (Willaert et al. 2020). Moreover, the structural and multilevel conception of opinions is generally compatible with common social science survey methods and closely resembles recent experimental designs to assess persuasiveness and effects of arguments (Kobayashi 2016; Shamon et al. 2019). The proposed framework, therefore, paves the way for completely new ways of model validation and micro foundation which have been repeatedly identified as the most important frontiers in model-based research on opinion dynamics (Sobkowicz 2009; Flache et al. 2017).

1.8 The paper is organized as follows. We first introduce the model in Section 2. In Section 3 we analyze its basic behavior by looking at 12 cases differing in the type of evaluative structure and homophily mechanism for two issues. Two cases are explored with some more detail in this section. In Section 4 we look at an example with three issues that illustrates how meaningful arguments can in principle be integrated into the structure. Finally, we draw a conclusion on the paper and discuss the model from a broader perspective in Section 5. Notice that an online implementation of the model accompanies the paper and is briefly described in the end of Section Section 3.1.

The Model

2.1 We model a population of agents that exchange arguments about different political issues. Our model is based on three different ingredients: (1) agents exchange their beliefs regarding a series of facts; (2) different (partially overlapping) sets of facts are associated with different political issues and an agent's attitude towards these issues is a function of the evaluative relevance of the facts for the different issues; and (3) agents preferentially interact with other agents that hold similar attitudes on the issues. These combined processes give rise to polarization and *constraints or functional interdependencies* (cf. Converse 1964, 3) in the configuration of attitudes towards multiple political issues.

Argument strings

2.2 Consider a population of N agents that exchange arguments about different political issues. There are N_A argument dimensions related to facts which an agent (say s) either believes to be true $a_{sk} = 1$ or not $a_{sk} = 0$. That is, each agent holds a binary string $\vec{a}_s \in \{0,1\}^{N_A}$ representing her current beliefs in a number of facts. Within our setting argument communication corresponds to the exchange of beliefs in facts, and we do not further distinguish between the two. As a convention, we shall index the argument dimensions by k ($1 \leq k \leq N_A$) and the agents by s and r for sender and receiver ($1 \leq s, r \leq N$). Consequently, the argument strings of the entire population can be represented by an $N \times N_A$ matrix A in which single rows represent the argument strings

of the agents and the element a_{sk} denotes the belief of agent s with respect to the kth factual dimension. As shown in Figure 1 we assume that the string of beliefs a_s underlies an agent's opinion.

Evaluative structure

2.3 We consider that opinions are multi-level constructs. Agents hold beliefs regarding a series of factual dimensions (encoded in A) and opinions on the set of issues are determined by specific configurations of beliefs. One of the main assumptions of this work is that the link from beliefs to attitudes is realized by a cognitive-evaluative map that encodes how different factual dimensions contribute to an attitudinal judgement. This map is modeled as a bipartite graph $(\mathcal{I}, \mathcal{A}, \mathcal{C})$ which represents the relation between a set of issues $\mathcal{I}$ and the set of argumentative dimensions $\mathcal{A}$ related to facts. We denote the cardinality of these two sets as N_I and N_A respectively. The $N_A \times N_I$ evaluation matrix C assigns values $c_{ki} \in [-1, 0, 1]$ to the set $\mathcal{C}$ of associations which represent (i.) whether an attitude object i (a political issue in our case) is positively or negatively evaluated if a certain fact k is believed to be true (sign(c_{ki})), and (ii.) the extent to which that fact k contributes to the evaluation of i ($|c_{ki}|$). If a belief k is not relevant for an issue i (no link) then $c_{ki} = 0$. This structure is inspired by the expectancy value model of attitudes following Fishbein & Raven (1962) and Fishbein (1963) and one may think of the set of factual beliefs as beliefs on the presence or absence of attributes. The resulting cognitive architecture is illustrated in Figure 1 for two issues (squares) and 6 facts (circles). The evaluative structure contains four positive links (solid lines) and four negative links (dashed lines). Notice that first two beliefs are relevant only to the first issue. Their contribution to the second issue is zero in this example. Likewise, the last two beliefs are only relevant for the second issue ($c_{51} = c_{61} = 0$). The third and forth belief are relevant to both issues and speak in favor of one but against the other issue.

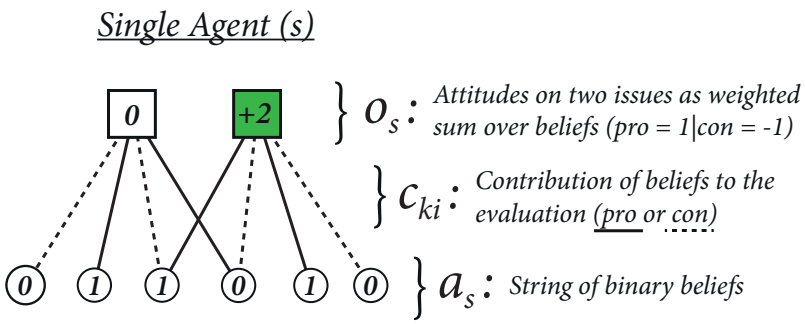

Figure 1: Cognitive architecture of the agents. Agents form attitudes on two different issues based on their beliefs in a number of facts which may contribute positively (pro, solid lines) or negatively (contra, dashed lines) to the attitudes.

2.4 In the computation of the evaluative judgement – that is, the attitude – regarding the different issues, we follow the algebraic model of expectancy value theory in a straightforward manner. Namely, the single argument-evaluation contributions are additively combined in the determination of the valence (degree of favor/disfavor) assigned the issue i:

$$o_s(i) = \sum_{k=1}^{N_A} a_{sk} c_{ki}. \tag{1}$$

Consequently, the attitudes on a whole set of N_I issues for a single agent s are given by the product $o_s = \vec{a}_s \cdot C$. For the entire population we can then write in matrix form

$$O = A \cdot C. \tag{2}$$

The element $o_{si} = o_s(i)$ in the resulting $N \times N_I$ matrix O represents the attitude or opinion of agent s towards issue i. Consider the agent shown in Fig. 1 as an example. It has a neutral attitude with respect to the first issue as one belief (2) supports a positive judgement, another belief (3) a negative one, and the third belief (5) is not relevant. The attitude regarding the second issue is rather positive because two beliefs support a positive stance.

2.5 Generally, the evaluative structure C may vary across individuals and every individual s could be represented by its own structure C_s. This would allow for inter-individual differences regarding the interpretation and relevance

of facts. Furthermore, while we model an exchange of arguments and assume that agents understand them equally, political discourse is actually often about the meaning of concepts involved in argumentative statements. That is, attitude change may actually often come about by changes in the perception, evaluative meaning and relevance of arguments. We shall discuss both issues in Section 5. As explained in the Introduction, in most parts of this paper, we assume a shared and time-homogeneous evaluative structure $C_s = C$ for all agents which is motivated by the fact that individuals within the same socio-cultural context have internalized similar evaluative schemata. We show that attitudes may polarize along ideologically coherent lines even if agents interpret facts in the same way because groups of agents selectively belief in distinct subsets of facts.

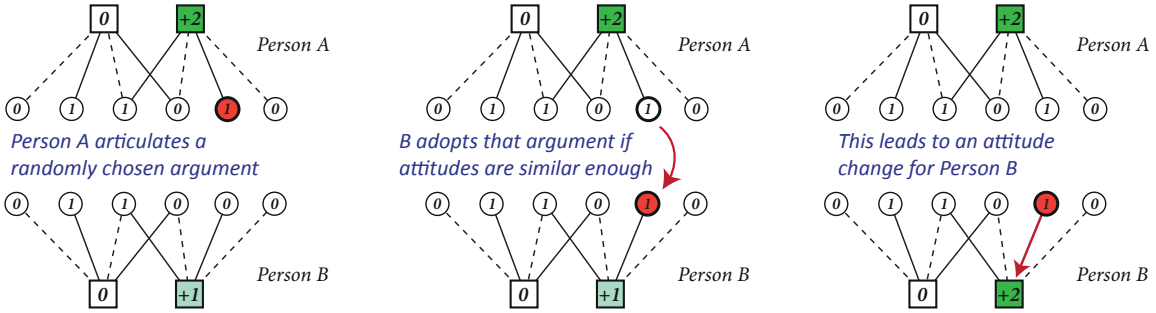

Figure 2: Summary of the argument exchange process performed during each step of the simulation. The sender (person A) chooses a random belief and presents it to its interlocutor. If the attitudes are similar enough (homophily, see below), the receiving agent (B) adopts that belief. As a result agent B has to recompute its attitudes (by Equation (1)). In this example the agent received an argument relating positively to the second issue and a weakly positive opinion is further reinforced by this.

Argument exchange

2.6 We assume a very simple mechanism for the argument exchange process and implement it as model of dyadic interaction. When running the model an agent pair (sender s and receiver r) with attitudes o_s and o_r is randomly selected at each time step. They engage in argument exchange activity if their opinion vectors are similar enough as explained below. The sender s, randomly choses an argumentative dimension k and articulates her belief a_k that fact k is true or not. The second agent r receives that argument and adopts the respective belief, i.e. $a_{rk} = a_{sk}$. With this simple procedure we follow previous approaches in opinion dynamics modeling with multidimensional opinions (Axelrod 1997; Banisch et al. 2010) and depart slightly from the more complex implementation in Mäs & Flache (2013) where the activation of one argument entails the deactivation of another. The argument exchange mechanism is visually summarized in Fig. 2

Attitude homophily and biased argument pools

2.7 Notice that in the argument exchange process implemented in the ACTB (Mäs & Flache 2013) as well as in our implementation no forms of biased argument processing are integrated. While the integration of such biases could be an interesting extension of the models, it is not needed to explain the emergence of bi-polar opinion distributions and the inter-attitudinal constraints that we address in this paper. A *homophily bias* which creates a tendency that actors interact and exchange arguments with like-minded others is sufficient.

2.8 In our case, homophily is assumed to play out at the level of attitudes towards the N_I issues and not on the level of the underlying arguments. Besides ACTB (Mäs & Flache 2013), this idea has another important predecessor in the biological model of sympatric speciation developed by Kondrashov (1986) and Kondrashov & Shpak (1998) where the phenotypic expression of genes (and not the genes themselves) become functional in terms of natural selection. In our context this means that attitudes (and not underlying cognitions) become functional regarding the selection of peers (Eagly & Chaiken 1998, 269) such that agents with opposing attitudes do not engage in constructive interaction. Just as phenotype assortativity leads to reproductively isolated gene pools and hence to the splitting of species (Kondrashov & Shpak 1998), homophily constraints at the attitude level lead to biased argument pools and attitude polarization in a sense closely related to Sunstein (2002).

2.9 There are different ways to integrate homophily assumptions into models of opinion dynamics. While a threshold mechanism referred to as bounded confidence is very common in continuous models of opinion dynamics

(Deffuant et al. 2000; Hegselmann & Krause 2002; Fortunato et al. 2005; Urbig & Malitz 2005; Lorenz 2007, 2008), a frequency-dependent interaction probability that takes into account the relative similarity of opinions with respect the entire population is used in Mäs & Flache (2013). In this work, we rely on a threshold mechanism by which agents influence one another only if their attitudes are similar enough. More precisely, let us denote by o_s and o_r the opinions of the sender and the receiver respectively and let $d(o_s, o_r)$ denote the distance between these opinions according to some distance measure on the space of opinions. Opinion or attitude homophily means that interaction is more likely if $d(o_s, o_r)$ is small accounting for the well-established observation that people are more likely to interact if they hold similar views (Byrne 1961; Huston & Levinger 1978; McPherson et al. 2001). The model studied here implements this by an influence threshold β on the opinion distance between two agents s and r such that exchange only takes place if $d(o_s, o_r) \leq \beta$. While this implementation of homophily strongly simplifies the decision taken in Mäs & Flache (2013), our results show that it leads to similar qualitative behaviour (see also Section 5). Notice also that such a threshold mechanism has been motivated in terms of self–categorization assuming that people feel in a group with others who have similar opinions (Lorenz 2008).

2.10 In our case, the position of agents in opinion space is determined by their belief strings A and the attitude structure C through Equation 2. In a model aiming at describing the evolution of attitudes on various issues, several options for computing the distance become available. For instance, it might be that positions on one issue are much more salient with respect to the decision of whether or not to engage in communication with another agent. In this case, we take into account only the positions on this issue in the distance computation. It might also be an asymmetric relation. Furthermore, even the situation that a single argument signals an unacceptable stance of the interlocutor might be plausible in some cases. We will explore some of these options and provide an overview of their impact in the next section.

2.11 Notice finally that the interaction behavior is only determined by opinion homophily, that is, by similarity in the attitude space (Byrne 1961). We do not consider homophily related to status or socio–demographic characteristics such as age, gender or religion (Lazarsfeld & Merton 1954) and the random pairing of agents is not mediated through a specific interaction network. Demographic attributes and interaction structures have been integrated with ACTB in Mäs et al. (2013) to analyze the effect of demographic faultlines in group discussion processes. In this paper, however, we focus on extending ACTB to multiple interrelated issues to understand how *ideological faultiness* may come about.

Implementation details

2.12 The analyses presented in the paper are based on MatLab (R2018b) implementations of the model. The basic version of the model with two issues (in the setting of Section 3.2) is provided in Figure 3. Notice that the model is implemented such that $N/2$ random pairs are drawn without replacement in the single iterations which means that each agent is either sender or receiver during each step. That is, if a simulation with $N = 1000$ agents runs for 1000 steps, each agent has been paired 1000 times, 500 times on average as sender and receiver. In addition to the analyses presented in the next two sections, an interactive online implementation (Javascript) is available under `www.universecity.de/demos/IssueAlignmentModel.html` (see Banisch (2019) and Section 3.1).

Results: Two Issues

Simulation settings and approach

Cognitive-evaluative maps

3.1 In this section we provide a general overview of the model behavior for the case of two issues. All the results are based on simulations with $N = 1000$ agents. In all the cases we consider that 6 factual dimensions are relevant for each of the issues and that three of them contribute positively with $c_{ki} = 1$ and three negatively with $c_{ki} = -1$ to the respective attitude. Following Equation (2), this means that the attitudinal judgements lie on a seven point scale ranging from -3 (extreme disfavor) to +3 (extreme favor) and are neutral $o_i = 0$ if all facts are believed. We look at three different conditions concerning the evaluative overlap with respect to the two issues:

1. the two issues are independent and there are 6 arguments relevant to the first and 6 other arguments relevant to the second issue (total number of 12 arguments),

```
function [O,A] = AMTwoIssuesShort(steps,beta)

% 1. Parameterization and initialization.
% 1.1. Basic numbers.
    N = 1000;    % number of agents
    M = 8;       % number of beliefs
    I = 2;       % number of issues

% 1.2. Shared cognitive-evaluative maps (strongly incongruent case).
    C(:,1)=[1;-1;1;-1;1;-1;0;0];    % belief evaluations issue 1
    C(:,2)=[0;0;-1;1;-1;1;-1;1];    % belief evaluations issue 2

% 1.3. Initialization.
    A = randi(2,N,M)-1; % random binary beliefs
    O = A*C;            % resulting attitudes/opinions for the population

% 2. Iteration loop.
    for step = 1:steps

% 2.1. Form N/2 random pairs of agents without replacement.
        pairs = randperm(N);

% 2.2. Interaction only if pairs are similar enough (influence threshold beta).
        interacting =   abs(O(pairs(1:2:N),1) - O(pairs(2:2:N),1)) + ...
                        abs(O(pairs(1:2:N),2) - O(pairs(2:2:N),2)) <= beta;

% 2.3. Argument exchange for each interacting pair.
        for pair = find(interacting)
            arg = randi(M);                                 % choose random argument
            A(pairs(2*pair),arg) = A(pairs(2*pair-1),arg);  % copy from sender to receiver
        end

% 2.4. Compute new attitudes for the population.
        O = A*C;
    end
end
```

Figure 3: Runable MATLAB-code of the basic model.

2. the two issues are weakly compatible in terms of the evaluative structure such that 2 factual dimensions contribute equally (one positively and one negatively) to the evaluation of the two issues (total number of 10 arguments),
3. the two issues are strongly incompatible by assuming that 2 arguments contribute positively to the first and negatively to the second issue and another 2 arguments contribute negatively to the first and positively to the second (total number of 8 arguments).

3.2 The respective bipartite graphs are shown in Fig. 4. Notice that there are 12 arguments in the first but only 8 in the last condition. We have chosen this setup to make sure that the range of opinions is equal (seven point scale from -3 to +3) in all the three conditions.

Modes of homophily

3.3 Furthermore, we analyze the effect of four different homophily mechanisms by considering four different measures of distance upon which agents are assumed to judge whether or not they engage in effective communication with one another:

1. the Manhattan distance $d(o_s, o_r) = |o_s(1) - o_r(1)| + |o_s(2) - o_r(2)|$ taking into account the two issues,
2. the Euclidean distance $d(o_s, o_r) = \sqrt{(o_s(1) - o_r(1))^2 + (o_s(2) - o_r(2))^2}$ that takes into account both issues as well,
3. the distance (or opinion difference) with respect to one issue i^*, $d(o_s, o_r) = |o_s(i^*) - o_r(i^*)|$,
4. and the difference with respect to a single belief k^*, $d(o_s, o_r) = |a_{sk^*} - a_{rk^*}|$.

3.4 A summary of the 12 different combinations of homophily conditions and the three different evaluative structures is provided by Fig. 10.

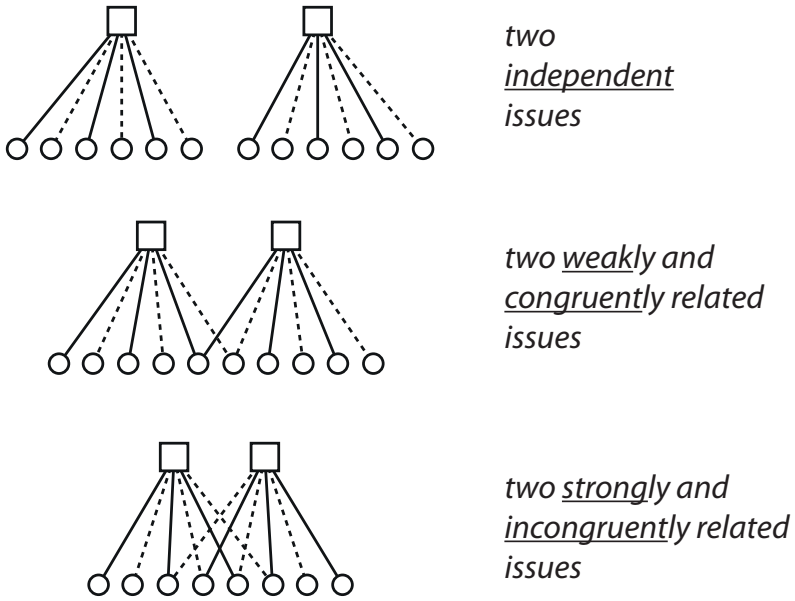

Figure 4: Three different evaluative maps are used that differ in the strength and congruency of the evaluative overlap. Solid lines indicate positive links and dashed line negative associations.

Initial conditions

3.5 In all the simulations we consider throughout this paper, the population is initialized with random binary argument strings. This means that for each argument independently there is a fifty-fifty chance to be activated ($a_k = 1$) in the beginning. Consequently, after projection onto the attitude scale by C, the initial distribution of attitudes with respect to the two issues is a binomial distribution with expected mean zero. Depending on the evaluative structure and in particular the evaluative overlap, the initial attitudes on the two issues are already correlated. As shown in the next section, this initial correlation is enforced in the argument exchange process especially under conditions of polarization.

Simulation approach

3.6 One of the main purposes of this paper is to prepare opinion dynamics models and the argument-based approach in particular for empirical application to real instances of debate. Going into the specialities of empirical cases with models that include expert knowledge or, potentially, argument networks inferred from real debate data (Willaert et al. 2020) brings along a shift of focus for the model analysis to more specific instances of empirical relevance. We approach the question of how to best provide an intuitive understanding of what the model entails in a way that could be called model phenomenology. The objective is not so much to statistically characterize it in the space of free parameters which is huge due to the combinatorial increase of possible cognitive-evaluative maps (but see Camargo (2020) for a very instructive attempt). The aim is much more to collect together the different kinds of phenomena the model may give rise to in conditions that are plausible in terms of application.

3.7 The main aim of this section is hence to provide an understanding of the kind of processes the model can give rise to. For this purpose, we adopt a phenomenological approach that goes deeper into two out of the twelve cases and provide an overview of all cases in the end of the section. In the two cases where we go into more detail, we first look at the time evolution of the actual opinion distribution of a single realization to provide intuition about the dynamical behavior of the model. We show that polarization in terms of a bi-modal opinion distribution on the issues and alignment across issues can be a stable outcome of the argument exchange process if homophily is strong, or a transient pattern if homophily is less strong. Secondly, we compare the opinion distribution after 1000 iterations for different β. Especially when population size increases, one can argue that transient opinion landscapes are empirically more relevant than the final absorbing states of the model (Banisch & Araújo 2010). Comparing the resulting opinion distributions for single-issue homophily (option 3 above) and the Manhattan distance on both issues (option 1 above) highlights the effect of different distance measures on the specific patterns of polarization and issue alignment.

Interactive model exploration

3.8 We hence concentrate on specific patterns in the opinion distribution of exemplary model runs that represent what to our experience are typical realizations for the different cases, and provide in this way a mesoscopic picture of model behavior. To complement this approach, an online implementation of the model available under Banisch (2019) accompanies the paper[1]. In this demonstration, users can interactively explore the effects of different distance measures and influence thresholds on the argument communication model with two issues. A wide range cognitive–evaluative maps (far beyond the three cases studied here) can be created by setting the number of congruent and incongruent evaluative connections between arguments and issues. While this allows to explore the cases of different argumentative overlap presented here, the online model provides additionally the possibility to include heterogeneity in the cognitive–evaluative maps, and the generation of different evaluative maps for different subpopulations in particular.

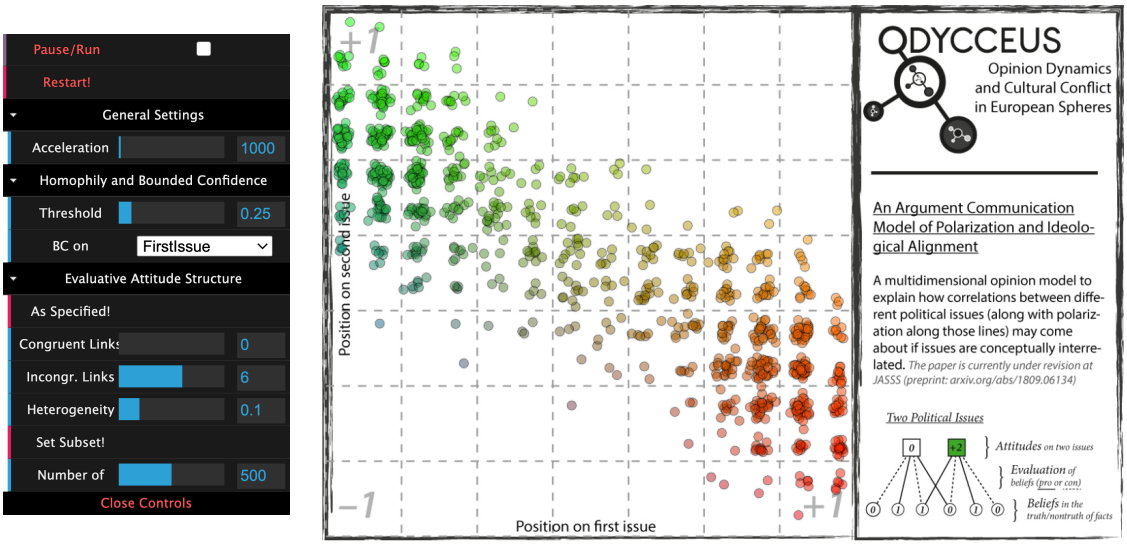

Figure 5: Screenshot of the online model explorer that accompanies the paper. The model is available at at the demo section of the first author's web page (Banisch 2019). Direct link: `www.universecity.de/demos/IssueAlignmentModel.html`. A brief guide to how to use it is provided under `www.universecity.de/demos/briefguideto/theargumentcommunicationmodel.pdf`

Two strongly coupled issues

3.9 To illustrate the dynamical behavior of the argument exchange model we first concentrate on a specific combination that highlights the two properties in the distribution of opinions our paper aims to address. Therefore, we consider the case of two issues that are strongly interrelated in an incongruent way. The respective evaluative structure is shown on the bottom of Fig. 4. There are 4 argument dimensions that are relevant for both issues. Two of those assign a positive weight to the first $c_{k1} = 1$ and a negative weight to the second issue $c_{k2} = -1$ and for the other two it is the other way around. This means that if one of these facts is believed $a_{sk} = 1$ by an agent s it will have a positive impact regarding its attitude on one and a negative impact on the evaluation of the other issue. With respect to homophily we assume in this section that the first issue *polarizes* meaning that agent with different opinions regarding the first issue are less likely to interact. The second issue plays no role in that. The influence threshold β is used to modulate the strength of homophily with respect to the first issue.

3.10 Fig. 6 shows the distribution of opinions on the two issues at different times of the process starting with the initial distribution on the upper left. As noted above, the random initialization of arguments leads to a binomial initial distribution of opinions for each of the issues. The extremes of the opinion spectrum are only rarely populated. The strong incongruent overlap encoded in the evaluative structure in this setting already induces a negative correlation between the initial opinions on both issues. Already after 100 steps this correlation pattern becomes considerably more pronounced, the spread of the distribution increases and the extremes become populated. Note that after this relatively short time the most extreme opinions on issue 1 (homophily-relevant) are adopted by the majority of agents but that also the second issue is polarized, albeit not that strongly. This strong pattern of bi-polarization is accentuated during subsequent steps of the simulation and agents with

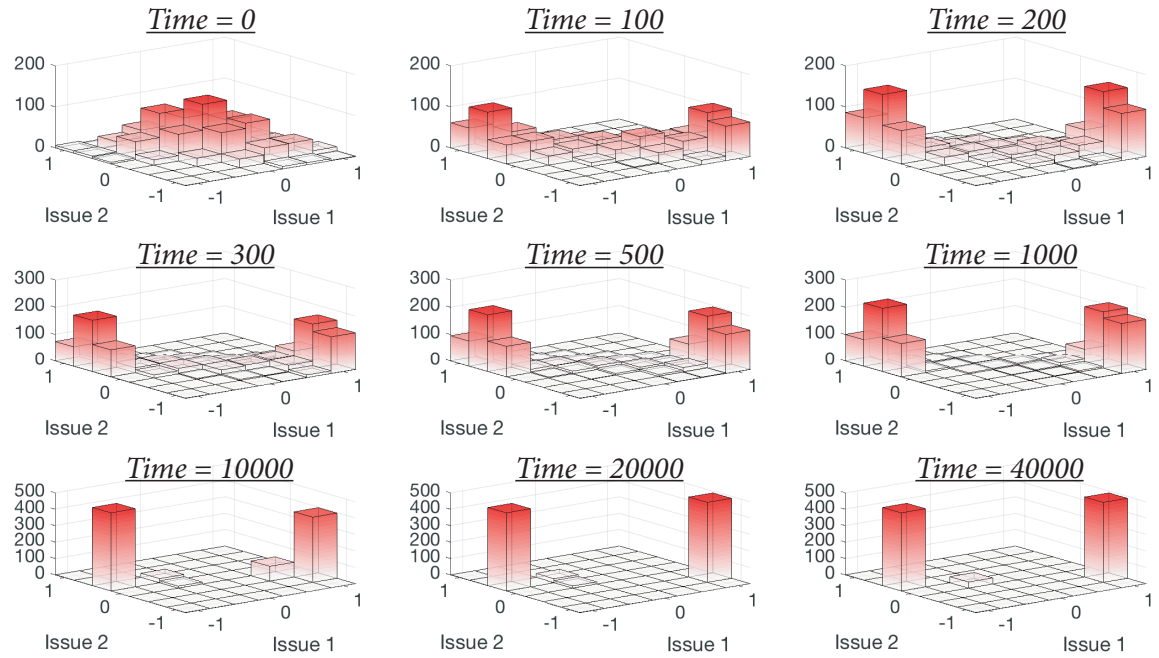

Figure 6: Attitude distribution of 1000 agents in the two-dimensional attitude space during different times of a single realization of the argument exchange process ($\beta = 1$).

intermediate attitudes become very rare. The strong negative coupling between the issues constrains the two different opinion groups to specific combinations of attitudes. The two groups develop opposing views not only with respect to the *polarizing* first issue but also with respect to the second one.

3.11 An example for such a setting might be two policy proposals that are discussed as competing alternative solution to the same problem. If one believes in coal as the future technology for electric power production and supports public investments into that area one probably disfavors subsidies in renewable energy production technology. Of course, our model provides only very stylized representations of such complex issues but it reveals a possible logic behind the formation of certain constellations of attitudes. It also shows how two groups that develop opposing attitudes tend to selectively believe and adopt facts that support their respective views if their interaction behavior is guided by their current attitudinal stance deciding about interlocutors whose arguments can be taken seriously.

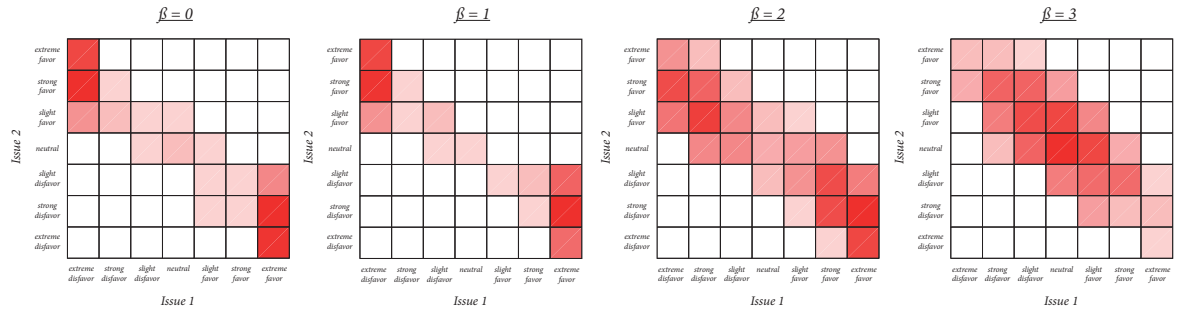

Figure 7: Attitude distribution in a model realisation with 1000 agents after 1000 iterations in the two-dimensional attitude space for different values of the influence threshold ($\beta = 0, 1, 2, 3$). A pattern of ideological coherent polarization emerges for a wide set of β.

3.12 In our model, the influence threshold β decides about the strength of this social interaction bias. For the given setting, Fig. 7 shows the opinion distributions after 1000 iterations for β ranging from zero to three (left to right). Notice that $\beta = 0$ indicates that only agents with exactly the same opinion on issue 1 engage into the argument exchange process. Likewise, for $\beta = 2$ agents that are differ at most 2 points on the 7 item attitude scale from extreme disfavor to extreme favor interact. This means, for instance, that a neutral agent will interact with strong

supporters but not with extreme ones and that agents that weakly favor one option will still interact with agents who weakly disfavor it.

3.13 As Fig. 7 shows, a polarization pattern emerges up to the case of $\beta = 2$. This effect is stronger on the salient first issue, but due to the strongly incongruent evaluative structure also rather pronounced on the second one. The population divides into two opposing groups of agents with a supportive stance regarding one and a negative stance regarding the other issue. This is a strong form of opinion alignment. If the threshold increases, homophily is not strong enough to lead to opinion bi-polarization and the population approaches a moderate consensus in the long run. Notice that the argument exchange process gives rise to only two outcomes in this mode of homophily: bi-polarization if β is small and consensus if β is large. Noteworthy, a bi-polarized organization of opinions is observed even for $\beta = 0$. This is in contrast to other models of bounded confidence (Hegselmann & Krause 2002; Deffuant et al. 2000; Lorenz 2007) with a typical transition from complete fragmentation for small influence thresholds to polarization for intermediate ones and consensus if β is large.

Two weakly congruent issues

3.14 Let us consider another example slightly more carefully and look at the case of two weakly coupled issues as shown in the middle in Fig. 4. The overlap in the evaluative structure indicates that a slightly positive relation between the opinions will result from such interdependencies. In terms of homophily, we consider now the Manhattan distance in the two-dimensional opinion space and assume that argument exchange takes place only if the $|o_s(1) - o_r(1)| + |o_s(2) - o_r(2)| \leq \beta$. The opinion distribution after 1000 iterations is shown in Fig. 8 for β ranging from zero to five.

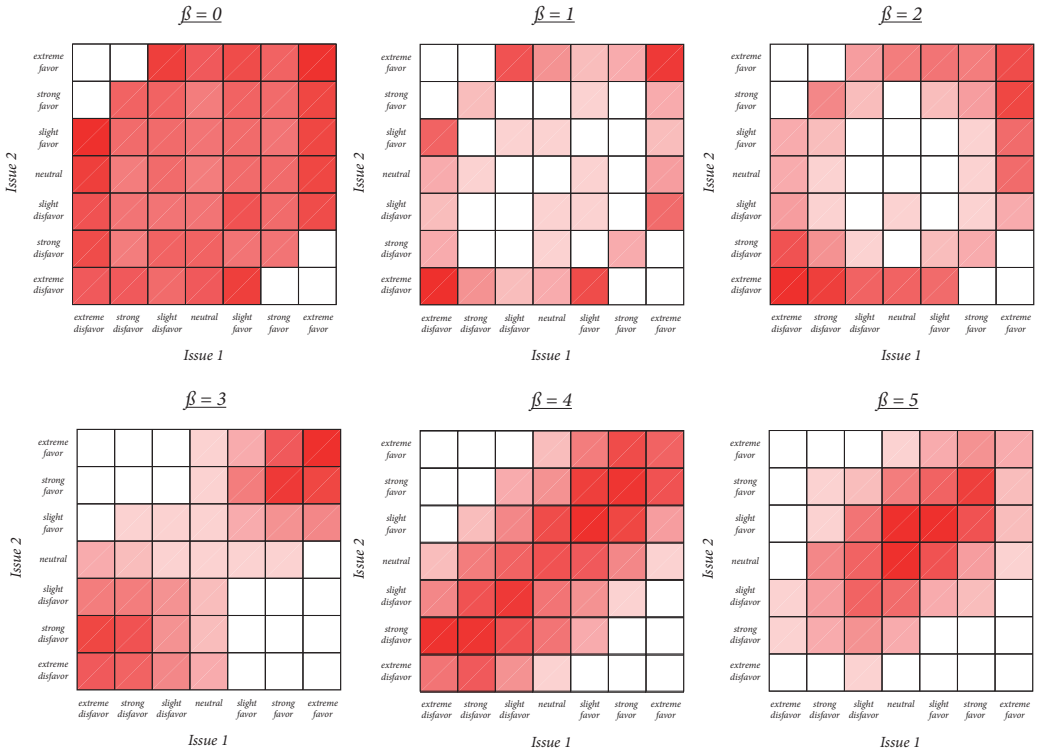

Figure 8: Attitude distribution in a model realisation with 1000 agents after 1000 iterations in the two-dimensional attitude space for different values of the influence threshold ($\beta = 0, 1, 2, 3, 4, 5$). The Manhattan distance is used in the homophily mechanism.

3.15 We observe two interesting transitions in this case. First, if β is small (here if $\beta \leq 2$), distributions emerge in which the neutral positions on the two issues are more and more sparsely occupied. This is compatible with the analysis of the 2D bounded confidence model by Lorenz (2003). Especially the case of $\beta = 1$ shows that there may be several highly populated configurations of opinions arranged elliptically around the center. As opposed to the previous case of homophily with respect to a single issue, a more fragmented opinion profile is observed for small β if the two issues are taken into consideration for homophily. This means that despite the

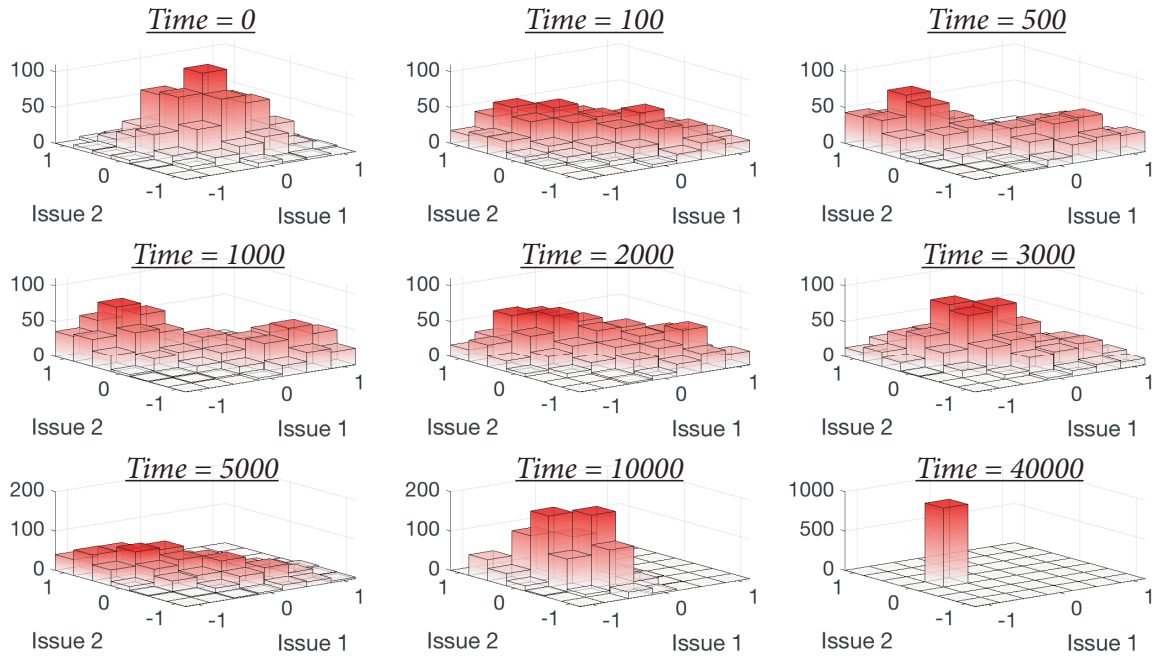

Figure 9: Time evolution of the attitude distribution in a model realisation with $\beta = 4$. The persistence of intermediate opinions resolves the short-term polarization in the long run.

positive coupling of the two issues, groups of agents may adopt a negative stance on one and a positive opinion on the other issue. As shown in the last row of Fig. 10 the shape of this depends on the strength of the evaluative overlap between the two issues.

3.16 As the influence threshold increases beyond $\beta = 2$ a different pattern emerges which is more closely related to a typical bi-polarized opinion distribution. Two groups emerge that develop consistently opposing standpoints on the two issues with negative or positive opinions regarding both issues. Notice that for $\beta = 4$ the distribution along this negative/negative to positive/positive dimension is much more flat and intermediate neutral opinions are still present after 1000 iterations. While a polarized state is stable for $\beta = 2$ it becomes a transient phenomena for $\beta = 4$.

3.17 This is shown in Fig. 9 where the time evolution of the opinion distribution is considered. Although a clear pattern of bi-polarization is visible after 500 steps, intermediate moderate opinions can be sustained in this parameter setting. These intermediate agents help to maintain a *flow of arguments* between the two groups which circumvent the emergence of isolated argument pools that would lead to persistent inter-group polarization. As a result of this persistent exposure to diverse arguments, agents with extreme opinions become more moderate again. In fact, the similarity between the initial opinion distribution and the distribution after 3000 steps is remarkable. The fact that very similar attitude distributions can lead to increasing polarization in one case and to moderation and long–term consensus in the other under argument communication indicates that a very interesting reorganization of beliefs has taken place in this initial phase of increasing polarization. It also highlights that the cognitive–evaluative layer – beliefs, evaluations, arguments, etc. – underlying patterns of public opinion may be of crucial importance for a better understanding of opinion change.

Summary of the model behavior

3.18 In Fig. 10, an overview of the model behavior in the different settings described in Section 3.1 is provided. Here we show the distribution of opinions after 1000 iteration for a relatively small $\beta = 1$. The three columns of this figure correspond to the three different evaluative maps shown in Fig. 4. In the first row the two issues are completely independent, in the second column the issues are weakly interrelated as in the previous section, and the third column represents the strongly coupled case considered in Section 3.9.

3.19 The rows correspond to four different ways to take into account opinion homophily. In the first row, we consider the case that two agents engage in the argument exchange process only if their belief regarding a single fact ($k = 5$ in this case) is equal. The idea behind this is that a single belief may signal an unacceptable stance of the

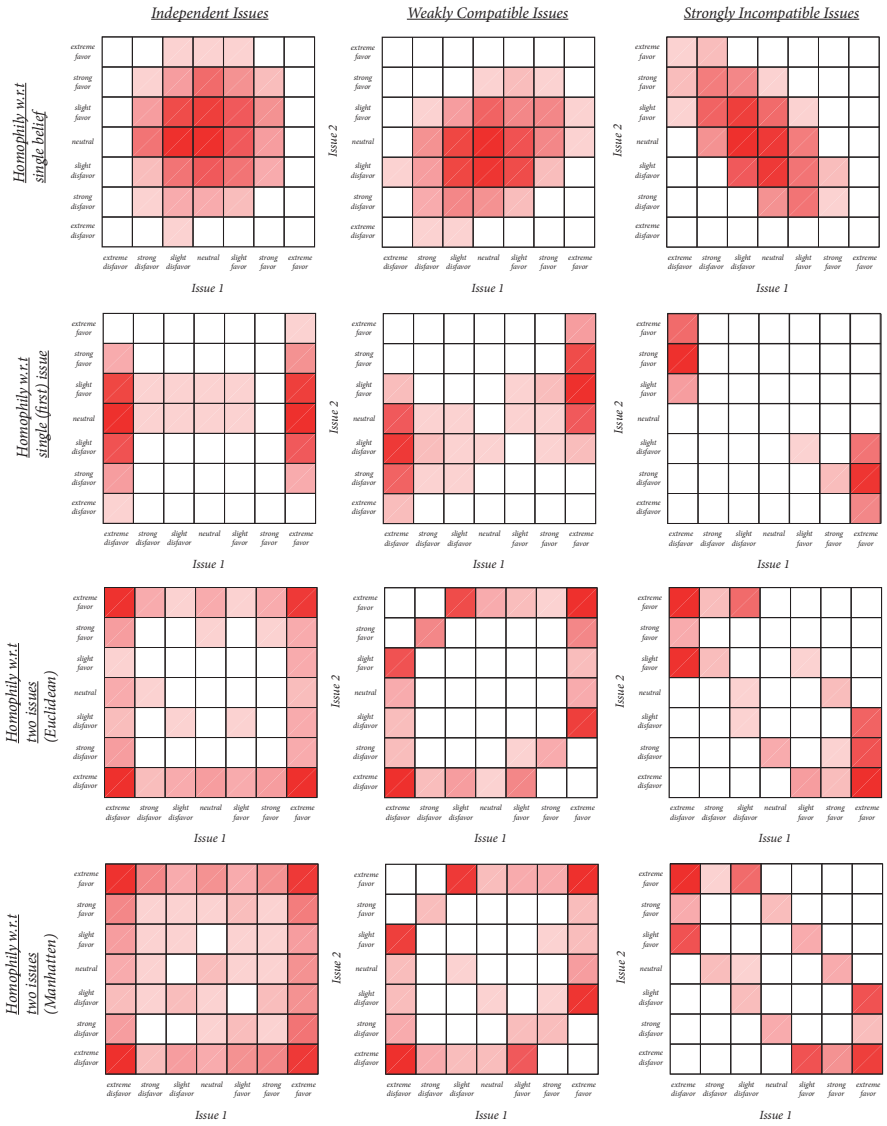

Figure 10: Overview over the model behavior in the different conditions described in Section 3.1. The distribution of single model realisations with $N = 1000$ after 1000 iterations is shown with $\beta = 1$. The reader is invited to explore this variety in the online tool that accompanies the paper (Banisch 2019).

interlocutor (such as, for instance, neglecting that CO2 emissions cause global warming). However, under the argument exchange dynamics incorporated into our model, this is generally not sufficient for polarization to emerge. The second row considers the homophily mechanisms studied in Section 3.1. As expected, a strong pattern of bi-polarization with respect to the *polarizing* issue emerges. The strength of interrelatedness of the two issues encoded in the cognitive-evaluative map governs to what extent the other issues (issue 2 in this case) polarizes *along with* the first issue. Noteworthy, in this case the bi-polarized outcome as opposed to fragmentation in other bounded confidence models (Lorenz 2007) is observed even if β is very small. The last two rows correspond to two different distance measures that take into account the agents' positions on both issues. In the third row the Euclidean distance is used and in the forth one the Manhattan distance. The figure shows that the behavior in terms of the distribution of opinions is very similar. Opinions are arranged around the center with strength and direction of argumentative overlap governing the shape of this pattern. Notice that a similar effect of a circular pattern around an unpopulated center has been observed in multi–dimensional continuous opinion dynamics with bounded confidence (Lorenz 2003, 53). When issues are interdependent the arrangement of admissible opinions around the center shows that certain opinion configurations are impossible due to the evaluative structure. Namely, under weak congruent coupling an extremely positive stance on one issues implies that an extremely negative stance with respect to the other is not admissible. These constraints imposed by the evaluative map are even more substantial under strong coupling where the upper right and the

lower left corner of the opinion space cannot be occupied.

3.20 The main purpose of this section has been to highlight two essential properties of the argument exchange model when extended to multiple issues. For this purpose, we have concentrated on a set of stylized settings. We have shown that the bi-polarizing dynamics of the ACTB (Mäs & Flache 2013) is recovered by our version. By showing that the polarization with respect to one issue may force agents to take a specific viewpoint on another issue if they are cognitively related we provide a possible explanation of a further important aspect of opinion polarization: the alignment of opinions across different issues within the two opposing groups of agents. This aspect of polarization – that is, an increasing *ideological uniformity* – is at the core of recent empirical studies on political polarization in American public opinion (Dimock et al. 2014). The comparison of modes of attitude homophily that take into account only one or respectively the two issues suggests that a single particularly *loaded* issue may be an important driver towards more pronounced bi-polarization into two ideological aligned camps.

An Example with Three Issues

4.1 One of the objectives of the model presented throughout this paper is to work towards a framework that allows to connect opinion dynamics modeling to real data on political statements and opinions. The evaluative structure is compatible with structural expectancy-value models of attitudes (Fishbein & Raven 1962; Fishbein 1963; Ajzen 2001) that are still widely used in survey-based attitude research. We will exploit this potential in the future but shall conclude this paper with a stylized example of some empirical plausibility.

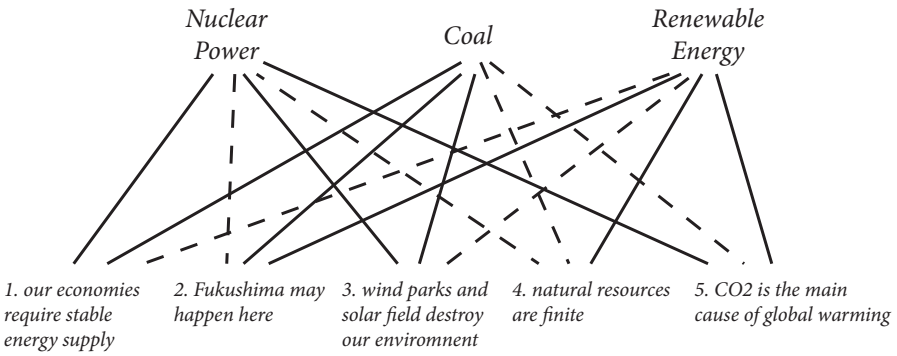

Figure 11: An example with five factual dimensions and three issues. The structure is highly constrained as all arguments are negatively (dashed) or positively (solid) related to all the three issues.

4.2 We consider a population of agents that debates on three different technologies for electric power production: nuclear power (i_1), coal (i_2) and technologies based on renewable sources (i_3). We allow only five arguments that are related by different degrees to economic stability (k_1, k_4), security (k_2) and sustainability (k_3, k_4, k_5):

- k_1 The growing economies of the world need reliable and scalable energy sources.
- k_2 An accident like in Fukushima may happen here.
- k_3 Large wind parks and vast solar fields impair the appearance of our environment.
- k_4 Natural resources are finite.
- k_5 Human-made CO2 emissions are the main cause of global warming.

4.3 The choice of some of these statements has been inspired by the debates on the web page ProCon.org.[2] Furthermore, recent analyses on the comments to Guardian articles on climate change have revealed a common use of arguments that interrelate the pros and cons of different energy technologies in real debates (Willaert et al. 2020). As shown in Fig. 11, all the arguments are linked to all the issues as they are most often used as pro- or con-arguments for one technology against others. Notice that the arguments have been chosen with some care so that each issue has two negative and three positive connections.

4.4 We run this model using the Manhattan distance in the homophily condition and with $\beta = 3$. Fig. 12 shows the opinion configurations after 20000 steps. In this case, the population has converged and no further argument exchange is possible due to the influence threshold. In this simulation three different groups emerge. The first one with almost one half of the population (471 agents) develops an opinion that is extremely supportive of renewable energy technology (+3) and has a negative opinion regarding the other two (-1). The string of beliefs of this group is (01011), meaning that they belief that nuclear accidents may happen again (2nd argument), that resources are finite (4th argument) and that CO2 causes global warming (5th argument). That is, the process of repeated argument exchange has led this group to believe only in those three arguments that are positively related to renewable energy. The second group of 359 agents strongly support nuclear power (+3) while downgrading renewables (-1) but not coal (+1). The associated belief string of this subpopulation is (10101). As for the first group, the interaction process has led to selective beliefs in arguments supporting nuclear power while neglecting the second and fourth argument concerning nuclear accidents and finite resources. A third smaller group of agents (170) supports coal (+2) while being neutral (0) with respect to the other two technologies. The beliefs in this group have settled on (11000), that is, the first argument related to stable supply and the second one related to nuclear accidents.

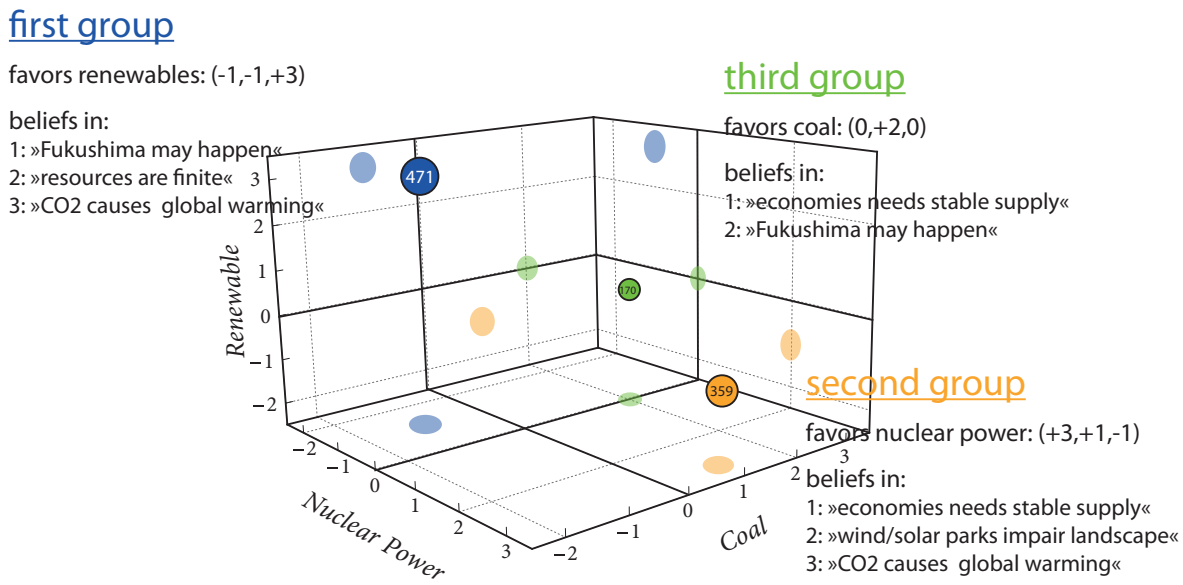

Figure 12: Result of a realization of the model with the evaluative structure shown in Fig. 11. Three groups emerge in the argument exchange process. One (blue, 471 agents) strongly supports renewable energy opposing coal and nuclear power. A second one (yellow, 359 agents) strongly supports nuclear power but is also slightly positive with respect to coal. A third group (green, 170 agents) favors coal and is neutral with respect to nuclear power and renewables.

4.5 Although we will discuss implications and limitations of the model in the next section, a short comment on the significance of this example shall be made here. We do certainly not want to claim that this restricted argument set is suited to accurately represent the real debate around those issues. But the example is informative about how our argument exchange model can be applied and what kind of results we can expect if real argument data is fed into it. Very often the intertwining of different arguments about different issues is rather hard to disentangle and its consequences for deliberative communication are unclear. The proposed model can improve our understanding how underlying structures (of beliefs, of argumentation, of evaluative connotations) affect debates like the one caricatured here.

4.6 For instance, further simulations have revealed that the emergence of a blue group (strong renewable supporters) is very robust and that the distance in attitude space between this group and the supporters of the two other technologies is generally larger than the distance between these other two. In the example sketched here this primary dimension of polarization between renewable supporters and opposers is very stable. Also an approximately fifty-fifty division of the population along this axis is a robust feature. The reason for this is an asymmetry in the evaluative structure (Fig. 11) in which the argument 4 (*natural resources are finite*) is the only one which contributes negatively to two of the issues (namely, coal and nuclear power). In such circumstances, an alliance between the supporters of coal and nuclear power (e.g. to gain majority) becomes considerably more

likely than the other coalitions. This is also consistent with the model outcomes for larger β (i.e. $\beta = 4$) where in virtually all cases two groups of similar size emerge that resemble such a formation of coalitions.

Concluding Discussion

5.1 This paper presents a model of argument exchange dynamics that extends the *argument communication theory of bi-polarization* (ACTB) proposed in Mäs & Flache (2013, 2). Our main contribution is to show that an argument exchange account of social influence dynamics can provide a useful framework for modeling processes of issue alignment (Baldassarri & Goldberg 2014; Dimock et al. 2014) by which attitudes on a set of issues become correlated along (ideological) dimensions. While the main focus in recent model-based studies of polarization (e.g. Dandekar et al. 2013; Friedkin 2015; Mäs & Bischofberger 2015; Duggins 2017; Banisch & Olbrich 2019) has been to explain the emergence of a bi-polar opinion distribution on a single issue, empirically motivated studies of polarization (e.g DiMaggio et al. 1996; Baldassarri & Goldberg 2014; Dimock et al. 2014) indicate that the alignment of attitudes across several issues is at least an equally important signature of polarization. To our knowledge this is the first modeling account that addresses these inter-issue dependencies and constraints in an explicit way.

5.2 With this paper, we contribute to a relatively new development to introduce a cognitive level into models of opinion dynamics. While most previous models operate on rather abstract opinion spaces ranging from binary states (Sznajd-Weron & Sznajd 2000; Galam 2005; Banisch 2014) to multi-dimensional (Axelrod 1997; Baldassarri & Bearman 2007; Banisch et al. 2010) and continuous state spaces (Deffuant et al. 2000; Hegselmann & Krause 2002)[3], the incorporation of belief systems (Friedkin et al. 2016; Parsegov et al. 2016), structural representations of attitudes (Urbig & Malitz 2005; Mäs & Flache 2013) and cognitive networks (Van Overwalle & Heylighen 2006; Wolf et al. 2012, 2015) is a relatively new development (see also Conte & Paolucci 2014). By allowing to integrate sometimes very specific information (see esp. Wolf et al. 2015) on the issues addressed by the model, this is a promising development to answer the fundamental critique that opinion dynamics lacks links to real data (Sobkowicz 2009), see also Flache et al. (2017). In this paper, we contribute to this development using the model by Mäs & Flache (2013) as a starting point. The multilevel opinion structure assumed in the original paper is extended to multiple cognitively related issues by assuming a cognitive-evaluative map which models the evaluative meaning of different arguments with respect to the different issues. This conceptualization of attitudes is closely related to structural theories of attitudes (Ajzen 2001), generally compatible with standard survey techniques and conceptualizes attitudes and arguments in a way closely related to recent experimental designs to measure the influence of arguments (Kobayashi 2016; Shamon et al. 2019). It therefore provides a promising framework to establish the connection between models and empirical cases. Moreover, it is also compatible with the above-mentioned connectionist models that employ cognitive networks (Van Overwalle & Siebler 2005, 273) and therefore an optimal trade-off between model complexity and parsimony, rich enough to incorporate relevant aspects of the conceptual structure underlying a real debate.

5.3 We illustrate this potential by the example studied in the last section, but the main focus of this paper has been to understand the basic properties of such a model. The model is already complex and involves quite a number of parameters and design choices, most importantly: the parameterization of the evaluative structure, the conceptual overlap and different forms of homophily at the attitude level. It is important to understand the impact of these modeling choices before the model can be used in more applied scenarios. We have followed a computational approach here by systematically analyzing 12 combinations of 4 homophily modes and 3 paradigmatic evaluative structures, but also analytical strategies developed in theoretical biology can be applied to simple cases (Banisch et al. 2018).

5.4 Most sociologist agree that one should include psychological complexity into models addressing social aggregations and collective phenomena only to the extent that they substantially contribute to the explanation of the phenomenon at stake (Lindenberg 1992; Kron 2004; Kroneberg 2005). On the one hand side, we follow this tradition by developing a very simple model of argument communication that simplifies the original model of ACTB by Mäs & Flache (2013) in two regards. Namely, we assume a process of argument exchange in which beliefs are directly transmitted from a sender to a receiver (*copied*) and we simplify the homophily mechanism by relying on the concept of an influence threshold. While the psychological precision of these assumptions at the inter-individual level can be disputed (and sometimes is, Mueller & Tan 2018) our results show that, at the aggregate level, the essential dynamical properties of ACTB are preserved under the parsimonious design choices made here. Notice also that in finite size systems occasional deviations from the sharp threshold function may lead to homogeneity in the long run (Mäs et al. 2010). In another regard, we take a step towards a more complex model by focusing on different cognitive-evaluative maps that link beliefs to attitudes. This generates a

considerable degree of freedom in our model and, in fact, would lead to a combinatorial explosion of the number of free parameters if these structures were purely individual.

5.5 However, in our case we consider that these evaluation structures have been acquired in the same socio-cultural context and are collectively shared by all agents. While from the methodological point of view this simply reduces the number of parameters and makes systematic computational analysis possible, there is also a tradition in Sociology to view culture as shared meaning structures (Berger & Luckmann 1970; Schütz & Luckmann 2017) and their integration in form of cognitive networks seems a viable approach (cf. Kroneberg 2005, 359, and see also Strauss & Quinn 1997). Understood in this way, as *ideal-typical* cognitive representations of evaluative meaning acquired in long processes of socialization, these structures become very relevant to the analysis of public opinion formation and are, in our model, essential for the explanation of issue alignment. Furthermore, the model entails the possibility to consider different sub-population or sub-cultures with differing cognitive maps and is therefore suited to explore the impact of cultural differences (operationalized in this way) on deliberative argument exchange process (a first attempt is made in Banisch et al. 2018). This also allows to link the model to recent empirical work on the identification of different belief systems within different social strata or sub-cultures (Baldassarri & Goldberg 2014; Daenekindt et al. 2017, see also Converse 1964) and the qualitative differences with respect to issue alignment in particular (cf. Goldberg 2011, Figure 7). The online implementation of the model (Banisch 2019) entails the possibility to define subgroups with different evaluative structures.

5.6 Models are caricatures of real social processes and for opinion dynamics, where empirical measurement is in itself a hard task, this is even more evident. The model put forth here concentrates on a process of exchange of arguments and leaves many important things out. Identity, sometimes argued to play a pivotal role in opinion making (Achen & Bartels 2017) is not included. We have considered the case of a single argument that might signal identity, but a more appropriate incorporation of identity certainly requires the integration of different mechanisms to account for the complex interaction of attitude and identity. Well-established psychological effects relating motivated reasoning and biased argument processing to attitude polarization (Lord et al. 1979; Kunda 1990; Taber et al. 2009; Kobayashi 2016) are missing as well. On the other hand, the homophily mechanism by which agents avoid exchange with others that think differently can also be interpreted as a tendency to judge their arguments as not reliable or relevant giving favor to arguments that are coherent with the own attitude. Social structure beyond opinion homophily is also left out of the model. We have done some simulation experiments that include different social interaction structures and found no indication that it changes the results reported here, mainly due to the fact that very similar processes of opinion alignment and polarization are observed within different communities. Yet it might be very interesting to reconsider the effects of demographic crisscrossing addressed in Mäs et al. (2013) when *demographic faultliness* (716) mark differences in the evaluative maps such that arguments are interpreted differently.

5.7 Given that many different things could be – and have been – included into models of opinion dynamics, it is *a fortiori* important to devise scenarios in which models can be tested and different assumptions confronted. This is very difficult with most existing models as opinions are usually *void of meaning* and no correspondence to the thematic dimensions of real debates is sought. The model developed in this paper points out a possible direction to overcome this deficiency by mapping arguments used in real debates and the opinions they support. This allows to embed the model in *empirically informed scenarios* so that they can be validated on specific cases of discourse that involve opinions. The energy technology example sketched in the previous section provides an illustration of such a setting. Using survey data or argument data extracted from text with precision language processing techniques (Steels 2017; Van Eecke & Beuls 2018; Willaert et al. 2020) to inform the underlying opinion structure will open up completely new ways of model validation. The argument exchange mechanism studied here might be appropriate for some thematic complexes, others might teach us that different aspects such as group identity or morality have to be taken more seriously. In any case, research on opinion dynamics will greatly benefit from such a program.

Acknowledgment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 732942 (Opinion Dynamics and Cultural Conflict in European Spaces – www.Odycceus.eu).

Notes

[1]Notice that running the model requires a Browser with WebGL support. Notice also that the number of arguments is increased to 20 in order to allow more flexibility in defining cross-cutting cognitive-evaluative maps.

[2]https://alternativeenergy.procon.org in particular.

[3]See Castellano et al. (2009) and Flache et al. (2017) for more encompassing overviews.

References

Achen, C. H. & Bartels, L. M. (2017). *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton, NJ: Princeton University Press

Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology*, *52*(1), 27–58

Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, *193*(5), 31–35

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

Bacharach, M. & Gambetta, D. (2001). Trust in signs. In K. S. Cook (Ed.), *Trust in Society*, (pp. 148–184). New York, NY: Russell Sage Foundation

Bakshy, E., Messing, S. & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132

Baldassarri, D. & Bearman, P. (2007). Dynamics of political polarization. *American Sociological Review*, *72*(5), 784–811

Baldassarri, D. & Goldberg, A. (2014). Neither ideologues nor agnostics: Alternative voters' belief system in an age of partisan politics. *American Journal of Sociology*, *120*(1), 45–95

Banisch, S. (2014). From microscopic heterogeneity to macroscopic complexity in the contrarian voter model. *Advances in Complex Systems*, *17*(5), 1450025

Banisch, S. (2019). Demo Section: Interactive Exploration of Opinion Models. Retrieved from: `http://www.universecity.de/index.php?site=demos`

Banisch, S. & Araújo, T. (2010). On the empirical relevance of the transient in opinion models. *Physics Letters A*, *374*(31), 3197–3200

Banisch, S., Araújo, T. & Louçã, J. (2010). Opinion dynamics and communication networks. *Advances in Complex Systems*, *13*(1), 95–111

Banisch, S. & Olbrich, E. (2019). Opinion polarization by learning from social feedback. *The Journal of Mathematical Sociology*, *43*(2), 76–103

Banisch, S., Tran, T. D. & Olbrich, E. (2018). Argument exchange dynamics in a population with divergent mindsets. Paper presented at the Conference on Complex Systems 2018, Thessaloniki

Berger, P. L. & Luckmann, T. (1970). *Die gesellschaftliche Konstruktion der Wirklichkeit. Eine Theorie der Wissenssoziologie*. Frankfurt A. M.

Bramson, A., Grim, P., Singer, D. J., Fisher, S., Berger, W., Sack, G. & Flocken, C. (2016). Disambiguation of social polarization concepts and measures. *The Journal of Mathematical Sociology*, *40*(2), 80–111

Burnstein, E. & Vinokur, A. (1975). What a person thinks upon learning he has chosen differently from others: Nice evidence for the persuasive-arguments explanation of choice shifts. *Journal of Experimental Social Psychology*, *11*(5), 412–426

Burnstein, E. & Vinokur, A. (1977). Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology*, *13*(4), 315–332

Byrne, D. (1961). Interpersonal attraction and attitude similarity. *The Journal of Abnormal and Social Psychology*, *62*(3), 713

Camargo, C. Q. (2020). New methods for the steady-state analysis of complex agent-based models. *Frontiers in Physics*, *8*, 103

Castellano, C., Fortunato, S. & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of modern physics*, *81*(2), 591

Conte, R. & Paolucci, M. (2014). On agent-based modeling and computational social science. *Frontiers in Psychology*, *5*, 668

Converse, P. E. (1964). The nature of belief systems in mass publics. *Critical Review*, *18*(1-3), 1–74

Daenekindt, S., de Koster, W. & van der Waal, J. (2017). How people organise cultural attitudes: Cultural belief systems and the populist radical right. *West European Politics*, *40*(4), 791–811

Dandekar, P., Goel, A. & Lee, D. T. (2013). Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, *110*(15), 5791–5796

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

DiMaggio, P., Evans, J. & Bryson, B. (1996). Have American's social attitudes become more polarized? *American Journal of Sociology*, *102*(3), 690–755

Dimock, M., Doherty, C., Kiley, J. & Oates, R. (2014). Political polarization in the American public. Pew Research Center

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of Political Economy*, *65*(2), 135–150

Duggins, P. (2017). A psychologically-motivated model of opinion change with applications to American politics. *Journal of Artificial Societies and Social Simulation*, *20*(1), 13

Eagly, A. H. & Chaiken, S. (1998). Attitude structure and function. In D. Gilbert, S. Fiske & G. Lindzey (Eds.), *Handbook of Social Psychology*, (pp. 269–322). New York, NY: McGraw-Hill

Farrar, C., Fishkin, J. S., Green, D. P., List, C., Luskin, R. C. & Paluck, E. L. (2010). Disaggregating deliberation's effects: An experiment within a deliberative poll. *British Journal of Political Science*, *40*(2), 333–347

Fishbein, M. (1963). An investigation of the relationship between beliefs about an object and the attitude toward that object. *Human Relations*, *16*(3), 233–239

Fishbein, M. & Raven, B. H. (1962). The AB scales: An operational definition of belief and attitude. *Human Relations*, *15*(1), 35–44

Flache, A. & Macy, M. W. (2011). Small worlds and cultural polarization. *The Journal of Mathematical Sociology*, *35*(1-3), 146–176

Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S. & Lorenz, J. (2017). Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, *20*(4), 2

Fortunato, S., Latora, V., Pluchino, A. & Rapisarda, A. (2005). Vector opinion dynamics in a bounded confidence consensus model. *International Journal of Modern Physics C*, *16*(10), 1535–1551

Friedkin, N. E. (2015). The problem of social control and coordination of complex systems in sociology: A look at the community cleavage problem. *IEEE Control Systems*, *35*(3), 40–51

Friedkin, N. E. & Johnsen, E. C. (1990). Social influence and opinions. *Journal of Mathematical Sociology*, *15*(3-4), 193–206

Friedkin, N. E., Proskurnikov, A. V., Tempo, R. & Parsegov, S. E. (2016). Network science on belief system dynamics under logic constraints. *Science*, *354*(6310), 321–326

Galam, S. (2005). Local dynamics vs. social mechanisms: A unifying frame. *Europhysics Letters*, *70*(6), 705–711

Goldberg, A. (2011). Mapping shared understandings using relational class analysis: The case of the cultural omnivore reexamined. *American Journal of Sociology*, *116*(5), 1397–1436

Hegselmann, R. & Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, *5*(3), 2

Huet, S., Deffuant, G. & Jager, W. (2008). A rejection mechanism in 2d bounded confidence provides more conformity. *Advances in Complex Systems*, *11*(4), 529–549

Huston, T. L. & Levinger, G. (1978). Interpersonal attraction and relationships. *Annual Review of Psychology*, *29*(1), 115–156

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

Kelman, H. (1961). Processes of opinion change. *Public Opinion Quarterly*, *25*(1), 57–78

Kobayashi, K. (2016). Relational processing of conflicting arguments: Effects on biased assimilation. *Comprehensive Psychology*, *5*, 2165222816657801

Kondrashov, A. S. (1986). Multilocus model of sympatric speciation. III. Computer simulations. *Theoretical Population Biology*, *29*(1), 1–15

Kondrashov, A. S. & Shpak, M. (1998). On the origin of species by means of assortative mating. *Proceedings of the Royal Society B*, *265*, 2273–2278

Kron, T. (2004). General Theory of Action? Inkonsistenzen in der Handlungstheorie von Hartmut Esser. *Zeitschrift für Soziologie*, *33*(3), 186–205

Kroneberg, C. (2005). Die Definition der Situation und die variable Rationalität der Akteure. ein allgemeines Modell des Handelns. *Zeitschrift für Soziologie*, *34*(5), 344–363

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498

Laver, M. (2005). Policy and the dynamics of political competition. *American Political Science Review*, *99*(2), 263–281

Laver, M. (2014). Measuring policy positions in political space. *Annual Review of Political Science*, *17*(1), 207–223

Lazarsfeld, P. & Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. In M. Berger, T. Abel & C. H. Page (Eds.), *Freedom and Control in Modern Society*, (pp. 18–66). New York, NY: Van Nostrand

Leuthold, H., Hermann, M. & Fabrikant, S. I. (2007). Making the political landscape visible: Mapping and analyzing voting patterns in an ideological space. *Environment and Planning B: Planning and Design*, *34*(5), 785–807

Lindenberg, S. (1992). The method of decreasing abstraction. In J. S. Coleman & T. J. Fararo (Eds.), *Rational Choice Theory. Advocacy and Critique*, vol. 1, (pp. 3–20). Thousand Oaks, CA: Sage Publications

Lippi, M. & Torroni, P. (2016). Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology*, *16*(2), 10–25

Lord, C. G., Ross, L. & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109

Lorenz, J. (2003). Mehrdimensionale meinungsdynamik bei wechselndem vertrauen. Master's thesis, University of Bremen

Lorenz, J. (2007). Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, *18*(12), 1819–1838

Lorenz, J. (2008). Fostering consensus in multidimensional continuous opinion dynamics under bounded confidence. In *Managing Complexity: Insights, Concepts, Applications*, (pp. 321–334). Berlin: Springer

Macy, M. W., Kitts, J. A., Flache, A. & Benard, S. (2003). Polarization in dynamic networks: A Hopfield model of emergent structure. *Dynamic Social Network Modeling and Analysis*, (pp. 162–173)

Mäs, M. & Bischofberger, L. (2015). Will the personalization of online social networks foster opinion polarization? Available at SSRN 2553436

Mäs, M. & Flache, A. (2013). Differentiation without distancing. Explaining bi-polarization of opinions without negative influence. *PloS One*, *8*(11), e74516

Mäs, M., Flache, A. & Helbing, D. (2010). Individualization as driving force of clustering phenomena in humans. *PLoS Computational Biology*, *6*(10), e1000959

Mäs, M., Flache, A., Takács, K. & Jehn, K. A. (2013). In the short term we divide, in the long term we unite: Demographic crisscrossing and the effects of faultlines on subgroup polarization. *Organization Science*, *24*(3), 716–736

Mason, L. (2018). *Uncivil Agreement: How Politics Became Our Identity*. Chicago, IL: University of Chicago Press

McPherson, M., Smith-Lovin, L. & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415–444

Mueller, S. T. & Tan, Y.-Y. S. (2018). Cognitive perspectives on opinion dynamics: The role of knowledge in consensus formation, opinion divergence, and group polarization. *Journal of Computational Social Science*, *1*(1), 15–48

Parsegov, S. E., Proskurnikov, A. V., Tempo, R. & Friedkin, N. E. (2016). Novel multidimensional models of opinion dynamics in social networks. *IEEE Transactions on Automatic Control*, *62*(5), 2270–2285

Rosa, H. (2016). *Resonanz: Eine Soziologie der Weltbeziehung*. Berlin: Suhrkamp Verlag

Schütz, A. & Luckmann, T. (2017). *Strukturen der Lebenswelt*. München: UVK, Konstanz u. München

Shamon, H., Schumann, D., Fischer, W., Vögele, S., Heinrichs, H. U. & Kuckshinrichs, W. (2019). Changing attitudes and conflicting arguments: Reviewing stakeholder communication on electricity technologies in Germany. *Energy Research & Social Science*, *55*, 106–121

Sobkowicz, P. (2009). Modelling opinion formation with physics tools: Call for closer link with reality. *Journal of Artificial Societies and Social Simulation*, *12*(1), 11

Steels, L. (2017). Basics of fluid construction grammar. *Constructions and Frames*, *9*(2), 178–225

Strauss, C. & Quinn, N. (1997). *A Cognitive Theory of Cultural Meaning*. Cambridge: Cambridge University Press

Sunstein, C. R. (2002). The law of group polarization. *Journal of Political Philosophy*, *10*(2), 175–195

Sznajd-Weron, K. & Sznajd, J. (2000). Opinion evolution in closed community. *International Journal of Modern Physics C*, *11*, 1157–1165

Sznajd-Weron, K. & Sznajd, J. (2005). Who is left, who is right? *Physica A: Statistical Mechanics and its Applications*, *351*(2-4), 593–604

Taber, C. S., Cann, D. & Kucsova, S. (2009). The motivated processing of political arguments. *Political Behavior*, *31*(2), 137–155

Uitermark, J., Traag, V. A. & Bruggeman, J. (2016). Dissecting discursive contention: A relational analysis of the Dutch debate on minority integration, 1990-2006. *Social Networks*, *47*, 107–115

Urbig, D. & Malitz, R. (2005). Dynamics of structured attitudes and opinions. Third Conference of the European Social Simulation Association

Van Eecke, P. & Beuls, K. (2018). Exploring the creative potential of computational construction grammar. *Zeitschrift für Anglistik und Amerikanistik*, *66*(3), 341–355

Van Overwalle, F. & Heylighen, F. (2006). Talking nets: A multiagent connectionist approach to communication and trust between individuals. *Psychological Review*, *113*(3), 606–627

Van Overwalle, F. & Siebler, F. (2005). A connectionist model of attitude formation and change. *Personality and Social Psychology Review*, *9*(3), 231–274

Willaert, T., Banisch, S., Van Eecke, P. & Beuls, K. (2020). Tracking causal relations in the news. Available as arXiv preprint 1912.01252

Wimmer, A. & Lewis, K. (2010). Beyond and below racial homophily: ERG models of a friendship network documented on Facebook. *American Journal of Sociology*, *116*(2), 583–642

Wolf, I., Neumann, J., Schröder, T. & de Haan, G. (2012). The adoption of electric vehicles: An empirical agent-based model of attitude formation and change. Proceedings of the 8th Conference of the European Association for Social Simulation, Salzburg

Wolf, I., Schröder, T., Neumann, J. & de Haan, G. (2015). Changing minds about electric cars: An empirically grounded agent-based modeling approach. *Technological Forecasting and Social Change*, *94*, 269–285