Chaos

An agent-based model of multi-dimensional opinion dynamics and opinion alignment

Cite as: Chaos **30**, 093139 (2020); https://doi.org/10.1063/5.0007523
Submitted: 20 March 2020 . Accepted: 24 August 2020 . Published Online: 22 September 2020

Simon Schweighofer, David Garcia, and Frank Schweitzer

An agent-based model of multi-dimensional opinion dynamics and opinion alignment

Cite as: Chaos 30, 093139 (2020); doi: 10.1063/5.0007523
Submitted: 20 March 2020 · Accepted: 24 August 2020 ·
Published Online: 22 September 2020

View Online Export Citation CrossMark

Simon Schweighofer,[1] David Garcia,[1] and Frank Schweitzer[1,2,a)]

AFFILIATIONS

[1]Complexity Science Hub Vienna, Josefstädter Str. 39, 1080 Vienna, Austria
[2]Chair of Systems Design, ETH Zurich, Weinbergstrasse 58, 8092 Zurich, Switzerland

Note: This article is part of the Focus Issue, Dynamics of Social Systems.
[a)]**Author to whom correspondence should be addressed:** fschweitzer@ethz.ch

ABSTRACT

It is known that individual opinions on different policy issues often align to a dominant ideological dimension (e.g., left vs right) and become increasingly polarized. We provide an agent-based model that reproduces alignment and polarization as emergent properties of opinion dynamics in a multi-dimensional space of continuous opinions. The mechanisms for the change of agents' opinions in this multi-dimensional space are derived from cognitive dissonance theory and structural balance theory. We test assumptions from proximity voting and from directional voting regarding their ability to reproduce the expected emerging properties. We further study how the emotional involvement of agents, i.e., their individual resistance to change opinions, impacts the dynamics. We identify two regimes for the global and the individual alignment of opinions. If the affective involvement is high and shows a large variance across agents, this fosters the emergence of a dominant ideological dimension. Agents align their opinions along this dimension in opposite directions, i.e., create a state of polarization.

We develop an agent-based model to study how the opinions of agents change. Each agent has an individual opinion on various issues; therefore, this is a multi-dimensional problem. Our model tests different assumptions about how the opinions of agents influence each other. For example, agents can adjust their opinions such that they become closer in the opinion space or deviate even more. We demonstrate under which conditions agents align their opinions; that is, their various individual opinions can be mapped to the same dominating ideological dimension. This is an important finding because such a dominating characterization, for example, left-wing vs right-wing, is observed in empirical studies of opinion dynamics. However, so far, concise models to explain how this dominating dimension emerges from interactions were missing.

I. INTRODUCTION

The famous economist Nicolas Kaldor in 1961 suggested that theorists "should be free to start off with a stylized view of the facts—i.e., concentrate on broad tendencies, ignoring individual detail."[39] His advice was certainly taken by the numerous physicists modeling opinion dynamics,[9,24,35] one of the most flourishing topics in the area of sociophysics.[59] In many of these models, opinions are treated as binary variables, $\{0, 1\}$, very much like spins, and changes in opinions follow rather simplistic rules. Despite their abstract nature, these models have generated interesting insights into the dynamics of disordered systems.[18,50,64,68] For example, voter models allowed studying under which conditions consensus, i.e., a large domain with aligned spins, can be obtained or how a minority and a majority can coexist.[1,7,61]

The question is how well such models fare with respect to real, empirically observed opinion dynamics.[22] To answer it in the spirit of Kaldor requires us to specify the stylized facts that shall be used as a ground truth or a reference for judging such models. While there is no common agreement on these stylized facts, we can certainly pick, from our everyday experience, two observations in the political space that most scholars would subscribe to: (i) Opinions have become increasingly polarized; i.e., there are two fractions of almost equal share in the population with opposite opinions.[5,10,17,19,25–27,34,38,46,57,62,65,70] (ii) Opinions on different policy issues tend to be correlated strongly.[3,4] For example, individuals with a positive stance on cannabis legalization more likely have a negative stance on nuclear energy.[2,15]

Political scientists call the correlation among opinions *issue alignment*.[2,11,16] Issue alignment implies that the number of independent dimensions to describe all opinions is effectively reduced; i.e., opinion dimensions are bundled into dominant ideological dimensions.[54] We call this property *global alignment*. In political science, the main ideological dimension is called the left–right dimension, in the US also the liberal–conservative dimension.[25,40,43,45,47] If global alignment is high, the opinion spectrum of individuals regarding different issues can be characterized sufficiently by assigning them a position on this continuous ideological dimension. We can then define *individual alignment* as the alignment of an individual to this dominant ideological dimensions. Studies have shown that political behavior and decision making, such as election choices, parliamentary voting, legislative decisions, coalition formation, and judicial decisions, can be explained to a large degree based on the ideological positions of political actors.[3,4,37,49,54]

The emergence of this dominant ideological dimension is addressed as a major open research question in political science.[54] Based on the fact that political elites usually exhibit much stronger issue alignment than the general population, Poole (Ref. 54, p. 211) believes that "part of the answer to these questions is that political elites are *passionate* about their beliefs." We want to investigate this proposed link between what we call *affective involvement* in politics and issue alignment—both global and individual alignment. We will do this by modeling affective involvement as an individual variable and coupling it to opinion dynamics.

Can voter-type opinion dynamic models proposed by socio-physicists replicate the stylized facts of polarization and opinion alignment? Polarization is trivially built into voter-type models by the dichotomy of the two opposite opinions. Therefore, if we do not obtain consensus in the long run, i.e., the dominance of one opinion, we obtain polarization, i.e., the coexistence of two (by design) extreme opinions. In the absence of any alternative, these opinions already represent the ideological left and right positions. Model parameters allow us to adjust the fractions of the respective camps even to 50/50, i.e., a stalemate reminiscent of real political situations in quite a number of different countries.

We argue that such models have not passed the test for obvious reasons: They do not show the emergence of a polarized opinion state, and they also do not show the emergence of an ideological dimension along which polarization occurs. The term *emergence* refers to a process of self-organization that leads to a new systemic property as the result of the dynamic interactions between a large number of individuals. In our paper, these individuals are represented as agents with certain internal degrees of freedom, most notably their opinion. Interactions refer to the exchange of information about the opinions of others, which in turn results in an adjustment of the opinion of each individual. With respect to opinion dynamics, the emerging properties are polarization and global alignment.

In order to obtain these emerging properties, we have to change from binary opinions to continuous opinions. These opinions can still be mapped to a finite interval, e.g., $[-1, +1]$, but extreme opinions should be less frequent, at least initially, than moderate ones. Secondly, we have to change from one-dimensional opinions to multi-dimensional opinions. Each dimension represents a given policy issue about which an individual can have its own opinion. The dominating ideological dimension is not one of these policy dimensions, but will emerge through the correlation of these issues.

However, there is more to it. Experienced modelers would probably know how to obtain the requested outcome from simplistic assumptions. However, even a correct outcome on the macro level does not allow us to conclude that the respective assumptions for interactions on the micro level are correct as well. Because there are various ways of obtaining a given outcome, we need additional evidence for our micro-mechanisms. This means that we have to base our interaction model on theories or experiments that justify our assumptions. This is the most neglected problem of socio-physics models of opinion dynamics. To solve it would require to learn about works in sociology, psychology, and political science to consider how and why individuals change their opinions. These insights can still be formalized as shown, for example, in Ref. 28. The rules for interactions are then no longer *ad hoc* assumptions but backed up by additional disciplinary arguments.

The main goal of our paper is to provide an agent-based model of opinion dynamics that is able to reproduce the emergence of global alignment, starting from a random distribution of opinions. While the existence of alignment is already discussed particularly in political science, the challenge is to present a model that can generate global alignment as an emerging phenomenon, without encoding it in the setup of the model. Additionally, we are interested in the relation between alignment and individual affective involvement into politics.

In Sec. II, we will discuss the theoretical background of our modeling assumptions. We then continue with introducing the setup of our agent-based models in a multi-dimensional opinion space in Sec. III. Based on our theoretical assumptions, in Sec. III, we present three opinion dynamics models and analyze whether they lead to the emergence of global alignment. In Sec. IV, we then investigate the link between individual affective involvement in politics and issue alignment.

II. THEORETICAL BACKGROUND

A. Balance theory

Our first aim is to motivate our rules of opinion change from a plausible set of micro-mechanisms.[30] These micro-mechanisms are derived from established psychological theories, in particular, *cognitive balance theory*,[33] and its extension to social relations, *structural balance theory*.[8]

Cognitive balance theory focuses on the perspective of the individual, specifically the relation between its beliefs or opinions. It postulates that if an individual holds two or more beliefs that she judges as contradictory, she will experience this as unpleasant. To alleviate this unpleasantness, the individual will either adapt or drop one of these beliefs to re-establish accordance. Hence, an individual has the tendency to minimize cognitive dissonance, which can be seen as a micro-foundation of opinion formation.[28]

Structural balance theory extends cognitive balance theory to explain the relations between individuals. Two individuals i and j can have either a positive relation, $r^{ij} = +1$, or a negative one, $r^{ij} = -1$. Structural balance theory focuses on triadic relations $\{i, j, k\}$, i.e., relations between three individuals i, j, and k. It postulates that there are stable and unstable triads. Whether a triad is stable or unstable can be determined by multiplying the signs of the relations r^{ij}, r^{ik}, and r^{jk} in the triad. For example, if individuals i, j, and k have exclusively positive relations with each other, $r^{ij} = r^{ik} = r^{jk} = +1$, the product of their signs would be positive; i.e., the triad is stable. If, however, two relations are positive and one is negative, e.g., $r^{ij} = r^{ik} = +1$, $r^{jk} = -1$, the product would be negative and the triad is assumed to be unstable. Unstable triads have the tendency to transform themselves into stable triads. This means that either i convinces j and k to change their relations into a positive one, $r^{jk} = +1$, or i changes its own relation to either j or k to a negative one, $r^{ij} = -1$ or $r^{ik} = -1$. In the latter case, the triad would have two negative relations and one positive. Hence, the product of the signs becomes positive and the triad has become stable.

We combine the assumptions of cognitive and structural balance theory to explain how individuals' opinions interact with the social relations between individuals. An alternative approach is presented in one of our previous works on *weighted balance theory*,[57] which combines elements of cognitive balance theory with an emotional factor of evaluative extremeness that was calibrated through the empirical analysis of an electoral survey.

Figure 1 considers two individuals i and j and their opinions on three policy issues x, y, and z. For the moment, we assume that each individual can only have a positive or negative stance on each issue. Their opinions can be expressed by the opinion vectors $\mathbf{o}^i = \{o^i_x, o^i_y, o^i_z\}$. In the example shown in Fig. 1(a), for i, we find that $o^i_x = o^i_z = -1$, $o^i_y = +1$, while for j, we find that $o^j_x = -1$, $o^j_y = o^j_z = +1$. This means that both individuals have a negative stance on the issue x and a positive stance on y, but on the issue z, their opinions contradict each other.

Regarding the relation between individuals i and j, we now make the following assumption: Since i and j agree on two issues and disagree only on one, they have a positive relation; i.e., $r^{ij} = +1$. Together with the relation between individuals i and j, the three opinions now form three different triads, each of which contains i, j, and one of the three policy issues, x, y, or z. Based on the rules explained above, we see that the triads $\{i, j, x\}$ and $\{i, j, y\}$ are stable because the product of the signs is positive, whereas the triad $\{i, j, z\}$ is unstable because the product of the signs is negative. In other words, cognitive dissonance is produced if (i) i likes j but disagrees with j on any issue or (ii) if i dislikes j but agrees with j on an issue.

In the first case, we assume that i and j will reduce dissonance by assimilating their opinions to each other. In other words, there will be a positive (attractive) social influence between i and j. But what happens in the second case, where the opinions of i and j are so dissimilar that there is a negative relation between them? We consider two options regarding how individuals resolve the dissonance stemming from a negative relation, *bounded confidence* and *repulsion*.

B. Bounded confidence

First, one can assume that two individuals i and j do not interact anymore if differences in their opinions are larger than a certain threshold ε. This seems to be a reasonable argument because, without interaction, they are no longer confronted with the cognitive dissonance resulting from their negative relation. This argument also underlies the much-discussed filter bubbles and echo chambers,[51] which emerges when users of online media ignore information and opinions that do not fit their own. This option corresponds to the class of *bounded confidence models*,[14,32] and we will see whether it is sufficient to generate global alignment in our first two opinion dynamics models, presented in Secs. III C and III D.

The interaction threshold ε is usually assumed to be constant and equal across individuals. The bounded confidence model, for example, simply treats this as a tunable parameter that impacts the possibility of reaching consensus. However, it is very important for the opinion dynamics to what extent an individual may be affected by the respective policy issues. Psychological research shows that opinions with a stronger emotional component are more resistant to change.[56] In other words, beliefs or opinions associated with strong emotional reactions are more stable. Hence, this individual emotional level may have an impact on the interaction threshold, i.e., whether or not opinions change. Thus, it is reasonable that an individual's *affective involvement* in politics influences how soon she will experience a critical level of cognitive dissonance that makes her abort or avoid interaction. In Secs. III D and III E, we will discuss how affective involvement can be related to the interaction threshold ε.

C. Repulsion

Instead of avoiding the interaction, if individuals already disagree on most issues and only agree on very few, they also have the possibility to adjust their opinions but into the negative direction. This means that they resolve their cognitive dissonance by also disagreeing on the few issues they previously still agreed on. In Fig. 1(b), we consider the example that individuals i and j disagree on issues x and y and only agree on z. Because they disagree on more issues than they agree, the overall relation between i and j in this case is negative; $r^{ij} = -1$. Considering this relation between i and j, the triads $\{i, j, x\}$ and $\{i, j, y\}$ are stable because i and j disagree on the issues x and y. However, the triad $\{i, j, z\}$ is unstable because i and j, even though they have a negative relation with each other, agree on issue z. This triad produces a cognitive dissonance for i and j, which has to be resolved in some way.

Instead of simply stopping their interaction, individual i can also change her opinion $o^i_z = +1$ to $o^i_z = -1$. This transforms the triad $\{i, j, z\}$ into a stable one, and the same would result from j changing her opinion o^j_z. As a result, i and j are now in disagreement on all three issues x, y, and z. However, because of the negative relation $r^{ij} = -1$, their cognitive dissonances are reduced to a minimum. This outcome, which is desirable for both individuals, postulates a repulsive force between the political positions of i and j.

This seems counter-intuitive only if we assume that, if two individuals interact, they should end up agreeing on more issues than before and not on less. Alternatively, there is empirical evidence[36] that when individuals are confronted with positions on alcohol prohibition that they fundamentally disagree with, they move away from these positions. Similarly, in an experiment,[19] partisan voters were confronted with the information that leaders of the opposite

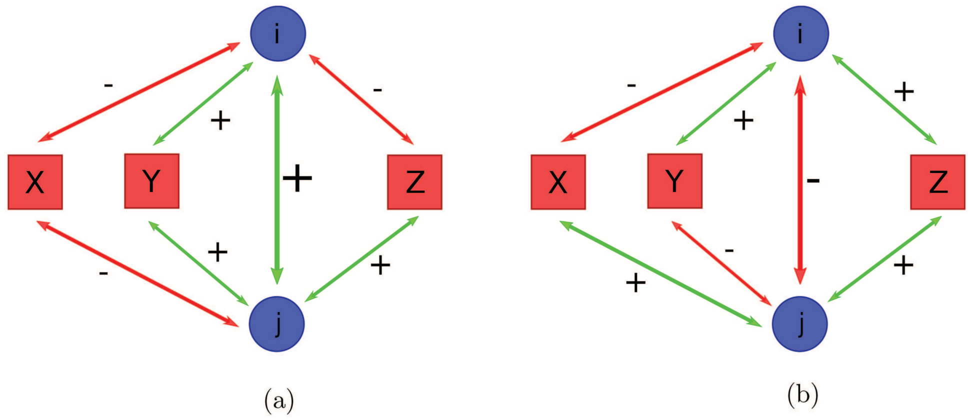

FIG. 1. Two agents i and j agreeing on the policy dimension x and y, but disagreeing on z (a), and disagreeing on x and y, but agreeing on z (b).

party endorse a certain policy and then adjusted their political position so as to contradict this policy.

The existence of a negative (repulsive) social influence is still debated in the literature. A recent study[66] could not find any evidence for a negative influence but only tested the social influence on a single issue dimension. However, this might not be applicable to our scenario with several opinion dimensions. Hence, in Sec. III E, we test the modeling assumption that if two agents disagree on too many issues, they modify their respective opinions such that they increase their overall disagreement.

We discussed bounded confidence and repulsion as two possible reactions to the cognitive dissonance stemming from encounters between individuals with dissimilar issue positions. But how do individuals determine their similarity in a multi-dimensional, continuous opinion space?

D. Proximity voting vs directional voting

Political science offers two paradigms to how individuals perceive similarity in a multi-dimensional, continuous opinion space: *proximity voting* and *directional voting*. Under proximity voting, it is assumed that individuals determine their similarity by the Euclidean distance between their positions in the opinion space. It has been pointed out, however, that proximity voting is based on rather weak empirical evidence and that its widespread usage in political science is rather due to its convenient mathematical properties.[4]

Proximity voting is challenged by a different paradigm, called *directional voting*.[55] This means that voters do not vote for the candidate that is closest to them in the opinion space but for the candidate that is on the "right side" of most issues.

We can illustrate the difference between the two paradigms with a simple calculation. Let us consider three agents with the following opinion vectors in a two-dimensional policy space, $\mathbf{o}^i = (0.1, 0.1)$, $\mathbf{o}^j = (0.3, 0.3)$, and $\mathbf{o}^k = (-0.1, -0.1)$. From Eq. (5), we know the Euclidean distances with respect to the agent i; i.e., $d^{ij} = \sqrt{0.08}$ and $d^{ik} = \sqrt{0.08}$. Thus, if the perception of similarity is based on proximity in the Euclidean space, i would perceive the opinion of j exactly as similar as the opinion of k because it has the same Euclidean distance to both. But is this assumption realistic? Certainly not, if positive values on a given policy issue, e.g., marijuana legalization, represent a positive stance and negative values a negative stance. Both i and j are on the "pro" side of both policy dimensions, $o_1^i > 0$, $o_2^i > 0$, $o_1^j > 0$, $o_2^j > 0$, while k is on the "contra" side, $o_1^k < 0$, $o_2^k < 0$. The difference between i and j only lies in the strength of their approval to both issues. From this perspective, it would make sense if i and j would perceive each other as very similar and k as very dissimilar. This is exactly what the directional voting paradigm postulates. The question whether it is more realistic is still debated.[42]

In Sec. III, we will implement our theoretical considerations in the form of three opinion dynamics models. The first, presented in Sec. III C, is based on the assumption of proximity voting and bounded confidence. Second, we will explore a model based on directional voting and bounded confidence (Sec. III D) and, finally, a model based on directional voting and repulsion (Sec. III E).

III. MODELING MULTI-DIMENSIONAL OPINION DYNAMICS

A. Model setup

In all of the following models, we consider a multi-dimensional opinion space, in which each dimension $m = \{1, 2, \ldots, M\}$ refers to a specific political issue. About each political issue m, an agent i has an opinion $o_m^i(t)$, which can change in discrete time steps $t = 1, \ldots, T$. These opinions shall be expressed as real numbers that can be normalized such that they always fall into the interval $[-1, +1]$. A strong opposition of i to the policy issue m corresponds to $o_m^i = -1$, strong approval to $o_m^i = 1$, and a neutral position to $o_m^i = 0$. The political position of each agent i in this multi-dimensional opinion space is characterized by an opinion vector $\mathbf{o}^i(t)$ composed of the M opinions $o_m^i(t)$. Considering a multi-agent system with $i = 1, \ldots, N$ agents, the multi-dimensional opinion space is populated with N opinion vectors $\mathbf{o}^i$.

Each agent is further characterized by an individual level of *affective involvement* in politics, e^i. This scalar value does not change over time and can be expressed as a real number from the interval $[0,1]$, with $e^i=0$ corresponding to nonexistent, and $e^i=1$ to extremely strong emotional involvement.

To determine the initial state of the opinion vectors, for each agent and each opinion component o^i_m, a random number is sampled from a normal distribution, $\mathcal{N}(\mu_o,\sigma_o)$ truncated to the interval $[-1,+1]$. The mean of the initial opinion components is given as $\mu_o(t=0)=0$ and their standard deviation as $\sigma_o(t=0)<1$. This ensures that (i) all possible values have indeed a non-negligible probability to occur, but (ii) different from a uniform distribution, extreme opinions will not occur with the same probability as moderate opinions, but less frequent. Thus, the M-dimensional opinion space is initially populated with N opinion vectors $\mathbf{o}^i(0)$.

Similarly, for each agent, the affective level e^i is drawn from a truncated normal distribution $\mathcal{N}(\mu_e,\sigma_e)$, limited between 0 and 1, with mean μ_e and standard deviation σ_e.

B. General model dynamics

To specify the dynamics of the individual opinion vectors, we use the concept of *Brownian agents*.[58] In each of our models, the individual dynamics results from an additive superposition of deterministic and stochastic influences,

$$\mathbf{o}^i(t+1)=\mathcal{F}\left[\mathbf{o}^i(t),\mathbf{o}^j(t)\right]\,\mathcal{G}\left[\mathbf{o}^i(t),\mathbf{o}^j(t)\right]+\mathcal{Z}\left[\mathbf{o}^i(t)\right]. \quad (1)$$

The deterministic term is composed of two functions, $\mathcal{F}[\cdot]$ and $\mathcal{G}[\cdot]$, that depend on the opinion vector of the focal agent i, but also on the opinion vectors of other agents j, i could potentially interact with. Specifically, we consider asynchronous updating of the dynamics. This means, at every time step t, two agents i and j are selected uniformly at random from the agent pool of N agents. The term $\mathcal{F}\left[\mathbf{o}^i(t),\mathbf{o}^j(t)\right]$ then determines whether i and j will interact at all. This might not be the case if, for example, their opinion vectors $\mathbf{o}^i_t$ and $\mathbf{o}^j_t$ diverge too much, as we will discuss below. If i and j interact, then the term $\mathcal{G}\left[\mathbf{o}^i(t),\mathbf{o}^j(t)\right]$ determines how the opinion vector of i will change based on the influence from j.

The stochastic term, $\mathcal{Z}\left[\mathbf{o}^i(t)\right]$, represents random influences on the opinion vector of the agent i, specifically influences that do not originate from interactions with others individuals. For example, own thought processes may cause it to modify its opinions on various issues without external influences.

C. Bounded confidence model with proximity voting

1. Model description

Before we will discuss different functional forms of $\mathcal{F}[\cdot]$, $\mathcal{G}[\cdot]$, and $\mathcal{Z}[\cdot]$, we illustrate the dynamics by turning to the most simple case of a *one-dimensional* opinion space. Because there is only one policy issue, $M=1$, each agent i only has the opinion $o^i(t)\in\{-1,+1\}$. Using the linear transformation $x^i=(o^i+1)/2$, we can map these opinions to an interval $x^i\in\{0,1\}$. For the dynamics of continuous opinions $x^i(t)$, the bounded confidence model was proposed.[14,32,44] It assumes that two agents i and j will only interact if the difference between their opinions is smaller than a threshold value ε, denoted as the confidence interval, i.e., if the variable $z^{ij}(t)$ is larger than zero,

$$z^{ij}(t)=\varepsilon-\Delta x^{ij}(t)\geq 0,\quad \Delta x^{ij}(t)=\left|x^j(t)-x^i(t)\right|. \quad (2)$$

We note that, for the one-dimensional case, Δx^{ij} gives the Euclidean distance between the two opinions. If the two agents interact, both change their opinions toward the common mean; i.e.,

$$\begin{aligned}x^i(t+1)&=x^i(t)+\omega\left[x^j(t)-x^i(t)\right]\,\Theta\left[z^{ij}(t)\right],\\ x^j(t+1)&=x^j(t)+\omega\left[x^i(t)-x^j(t)\right]\,\Theta\left[z^{ji}(t)\right].\end{aligned} \quad (3)$$

Here, $\Theta[x]$ is the Heaviside function that gives $\Theta[x]=1$ if $x\geq 0$ and $\Theta[x]=0$ otherwise. The speed of change is determined by ω. If $\omega=0.5$, both agents immediately converge to the mean of their two opinions; i.e., $x^i(t+1)=x^j(t+1)=[x^j(t)+x^i(t)]/2$.

Whether or not the multi-agent system converges to a single opinion, denoted as consensus, depends on the value of ε. For $\varepsilon=0.5$, consensus is obtained; for $\varepsilon=0.2$, instead, two agent groups with distant opinions emerge. The smaller ε, the more different opinions coexist in equilibrium.[31,44] Various extensions of the bounded confidence model have been proposed,[13,23,52] also in combination with network dynamics.[29,41,48,69] For the bigger picture of this type of dynamics, see also Ref. 60.

The bounded confidence model assumes deterministic dynamics. Hence, it is expressed by the general opinion dynamics of Eq. (1), if we choose the different functions as follows (with $\mathbf{o}^i=o^i$ for the one-dimensional case),

$$\begin{aligned}\mathcal{F}\left[o^i(t),o^j(t)\right]&=\Theta\left[2\varepsilon-\left|o^j(t)-o^i(t)\right|\right],\\ \mathcal{G}\left[o^i(t),o^j(t)\right]&=o^i(t)+\omega\left[o^j(t)-o^i(t)\right],\\ \mathcal{Z}\left[o^i(t)\right]&=0.\end{aligned} \quad (4)$$

The bounded confidence model can be formally generalized toward multi-dimensional problems, but the results are not trivial, as we will show. As an illustration, we first use a two-dimensional opinion space shown in Fig. 2. There, each agent i has an opinion about issues 1 and 2, denoted by the opinion vector $\mathbf{o}^i(t)=o^i_1(t)\mathbf{s}_1+o^i_2(t)\mathbf{s}_2\equiv\{o^i_1(t),o^i_2(t)\}$. Here, $\mathbf{s}_1$, $\mathbf{s}_2$ denote the unit vectors (versors) of the respective coordinate axes. This means, in order to decide whether two agents interact, we have to determine the similarity of their opinion vectors at a given time step t.

One possible measure to determine the difference between the two opinion vectors is the Euclidean distance $d^{ij}(t)=d\left[\mathbf{o}^i(t),\mathbf{o}^j(t)\right]$, which is defined for the two-dimensional opinion space as

$$d^{ij}(t)=\sqrt{\left[o^i_1(t)-o^j_1(t)\right]^2+\left[o^i_2(t)-o^j_2(t)\right]^2}. \quad (5)$$

Indeed, as discussed in Sec. II D, the Euclidean distance is applied by political scientists to measure the similarity in the opinion space between two political actors. Greater distance then corresponds to less similarity.[4] To normalize the Euclidean distance to values between 0 and 1, it has to be divided by the length of the diagonal of the opinion space, $\sqrt{4M}=2\sqrt{2}$.

We now assume as in the one-dimensional bounded confidence model that two agents i and j interact if their normalized Euclidean distance is less than a given threshold value 2ε. As a

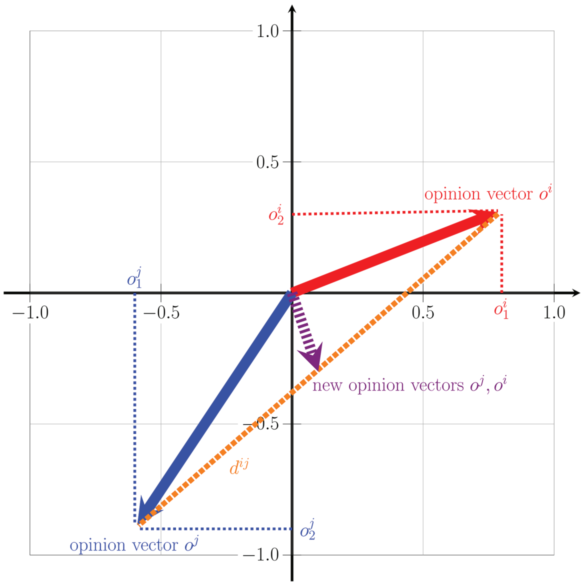

FIG. 2. Opinion vectors of two agents *i* and *j* in a two-dimensional opinion space. d^{ij} denotes the Euclidean distance [Eq. (5)]. The new opinion for both agents follows from Eq. (4).

result of this interaction, both agents adjust their opinions component wise to the common mean. With $\omega = 0.5$, we choose for the two-dimensional case the different functions in the general opinion dynamics of Eq. (1) as follows:

$$\mathcal{F}\left[\mathbf{o}^i(t), \mathbf{o}^j(t)\right] = \Theta\left[4\sqrt{2}\,\varepsilon - d^{ij}(t)\right],$$
$$\mathcal{G}\left[\mathbf{o}^i(t), \mathbf{o}^j(t)\right] = \frac{\mathbf{o}^i(t) + \mathbf{o}^j(t)}{2} = \frac{o_1^i(t) + o_1^j(t)}{2} + \frac{o_2^i(t) + o_2^j(t)}{2},$$
$$\mathcal{Z}\left[\mathbf{o}^i(t)\right] = \xi_i(t). \tag{6}$$

At difference with the deterministic bounded confidence model, here, we have added a random vector drawn from a truncated normal distribution, limited between −1 and 1, with mean $\mu_\xi = 0$ and standard deviation $0 < \sigma_\xi < 1$. This shall account for stochastic influences on the opinion formation not related to the interactions and is similar to the noise introduced into one-dimensional bounded confidence models, e.g., by Refs. 6, 20, and 53.

2. Model outcomes

We illustrate the dynamics of the two-dimensional bounded confidence model, Eq. (6), by means of stochastic simulations of the multi-agent system. The final results are shown in Fig. 3 for two different values of the bounded confidence interval ε. The details of the dynamics are presented in Fig. 14 of Appendix A.

Figure 3(a) shows the initial state for our simulations, where agents got randomly assigned an opinion vector in the two-dimensional opinion space, as described in Sec. III B. Figure 3(b) shows the outcome of the opinion dynamics if a rather large confidence interval ε is chosen, while Fig. 3(c) shows the outcome for a rather small value of ε. The results are in line with insights from the one-dimensional bounded confidence model [Eq. (3)]. If ε is large enough, we see the emergence of consensus in the middle of the opinion space. For the classical bounded confidence model, this would be $x^{\text{stat}} = 0.5$, whereas it is here $\{o_1^{\text{stat}}, o_2^{\text{stat}}\} = \{0, 0\}$. If ε is too small to reach consensus, we observe the formation of different clusters in the opinion space, i.e., groups of agents converging to the same opinion. More specifically, most agents converge to a big cluster in the center of the opinion space, while agents at the fringe of the original, spherical random distribution are left behind and form smaller clusters. This indicates the long-term coexistence of different opinions in the multi-agent system. We note that, in this case, still the majority of vectors converge to the origin, which shows the largest cluster of agents, whereas in the periphery, some agents are left behind in smaller clusters.

While the outcome reached is mainly determined by the value of ε, it also depends on the level of randomness, expressed by ξ. If the standard deviation σ_ξ is sufficiently large, random changes of opinions are able to bring agents sufficiently close in the opinion space such that they can continue to interact. This then fosters the emergence of consensus by destabilizing opinion clusters in the periphery. Whether the outcome of the simulation results in consensus or coexistence is certainly different from the agent perspective, the resulting average opinion over all agents in the long run is in both cases the same, namely, $\{\bar{o}_1, \bar{o}_2\} = \{0, 0\}$.

Thus, irrespective of the value of ε, the multi-dimensional bounded confidence model does not lead to the emergence of a main ideological dimension and global opinion alignment.

D. Bounded confidence model with directional voting

1. Model description

In order to apply the directional voting paradigm, we have to modify the way agents quantify distances between opinions. For this, we introduce a new measure, *directional similarity*, D^{ij}. While the Euclidean distance d^{ij}, Eq. (5), takes the full information from the opinion vectors $\mathbf{o}^i$ and $\mathbf{o}^j$ into account, the directional similarity only uses information about the angles ϕ^i and ϕ^j of the respective opinion vectors. Different from Ref. 63, we use the angle instead of the cosine distance in order to keep our similarity function consistent with our opinion change function (vector rotation; see below).

To formalize this step, we transform the opinion vector $\mathbf{o}^i(t) = \left\{o_1^i(t), o_2^i(t)\right\}$ into polar coordinates, $\mathbf{o}^i(t) = \left\{\left|o^i(t)\right|, \phi^i(t)\right\}$, where the length $\left|o^i(t)\right|$ and the angle $\phi^i(t)$ of the vector are, for a two-dimensional opinion space, defined as follows:

$$\left|o^i(t)\right| = \sqrt{\left[o_1^i(t)\right]^2 + \left[o_2^i(t)\right]^2},$$
$$\phi^i(t) = \begin{cases} \arctan\left[o_2^i(t)/o_1^i(t)\right] & \text{for } o_1^i(t) > 0, o_2^i(t) \geq 0, \\ \arctan\left[o_2^i(t)/o_1^i(t)\right] + 2\pi & \text{for } o_1^i(t) > 0, o_2^i(t) < 0, \\ \arctan\left[o_2^i(t)/o_1^i(t)\right] + \pi & \text{for } o_1^i(t) < 0. \end{cases} \tag{7}$$

The case analysis is needed because $\arctan(x)$ is not an injective function, but it is convenient to implement. This always returns a

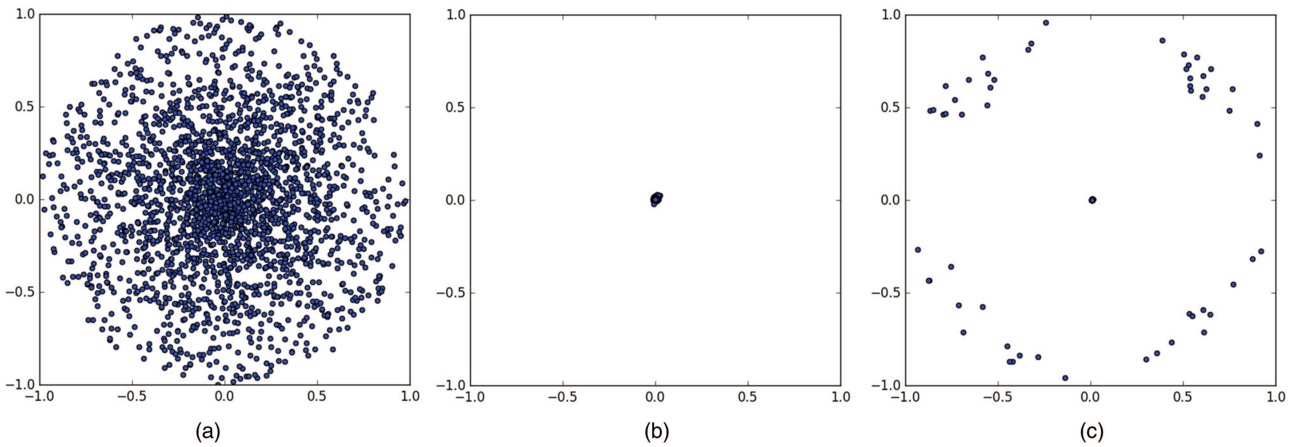

FIG. 3. Opinions of $N = 10\,000$ agents in a two-dimensional opinion space. (a) Initial distribution at time $t = 0$. (b) Long-term outcome ($t = 60\,000$) if $\varepsilon = 0.5$. (c) Long-term outcome ($t = 210\,000$) if $\varepsilon = 0.25$. For other parameters, see Appendix A.

value $\phi \in [0, 2\pi]$. We can then define the difference between the angles of the opinion vectors of agents i and j as

$$\Delta\phi^{ij}(t) = \begin{cases} \left[\phi^j(t) - \phi^i(t)\right] & \text{if } \left|\phi^j(t) - \phi^i(t)\right| \leq \pi, \\ 2\pi - \left[\phi^j(t) - \phi^i(t)\right] & \text{if } \left|\phi^j(t) - \phi^i(t)\right| > \pi. \end{cases} \tag{8}$$

This always returns a value $\Delta\phi \in [0, \pi]$, which can be mapped to an interval $[0, 1]$ by scaling $\Delta\phi/\pi$.

To specify the general opinion dynamics, Eq. (1), we make the following assumption for $\mathcal{F}\left\{\mathbf{o}^i(t), \mathbf{o}^j(t)\right\}$: two randomly chosen agents i and j will only interact if $\Delta\phi^{ij}(t)/\pi$ is less than a critical threshold ε^i. Different from the proximity voting model, we now follow studies such as Ref. 12 and model ε as an individual parameter. Specifically, we assume that it is coupled to the *affective involvement* e^i. As mentioned above, agents with a high level of affective involvement may become less tolerant to other opinions. Therefore, we define $\varepsilon \equiv \varepsilon^i = 1 - e^i$, where e^i initially is randomly chosen from the interval $[0, 1]$ and constant over time. This results in

$$\mathcal{F}\left\{\mathbf{o}^i(t), \mathbf{o}^j(t)\right\} = \Theta\left[D^{ij}(t) - e^i\right], \quad D^{ij}(t) = 1 - \frac{\Delta\phi^{ij}(t)}{\pi}. \tag{9}$$

We call $D^{ij}(t)$ the *pairwise directional similarity*. It becomes maximal, $D^{ij} = 1$, if both agents have perfectly aligned opinion vectors. In this case, even a maximal affective involvement, i.e., a minimal confidence interval, will not prevent them from interacting.

If the two agents interact, then they change their opinion such that they align their opinion vectors; i.e., the opinion vectors rotate to a new angle $\phi^i \to \theta^i$, but their absolute value does not change. The update function $\mathcal{G}\left\{\mathbf{o}^i(t), \mathbf{o}^j(t)\right\}$, therefore, reads in Cartesian coordinates as

$$\mathcal{G}\left\{\mathbf{o}^i(t), \mathbf{o}^j(t)\right\} = \left|\mathbf{o}^i(t)\right| \left[\cos\left\{\theta^i(t)\right\}\mathbf{s}_1 + \sin\left\{\theta^i(t)\right\}\mathbf{s}_2\right]. \tag{10}$$

Random influences now only affect the angle ϕ^i, i.e.,

$$\mathcal{Z}\left[\mathbf{o}^i(t)\right] = \xi(t)\phi^i(t). \tag{11}$$

Therefore, the updated angle $\theta^i(t)$ is determined both by the (deterministic) rotation and by random changes $\xi(t)$,

$$\theta^i(t) = \phi^i(t)\left[1 + \xi(t)\right] + \omega\Delta\phi^{ij}(t). \tag{12}$$

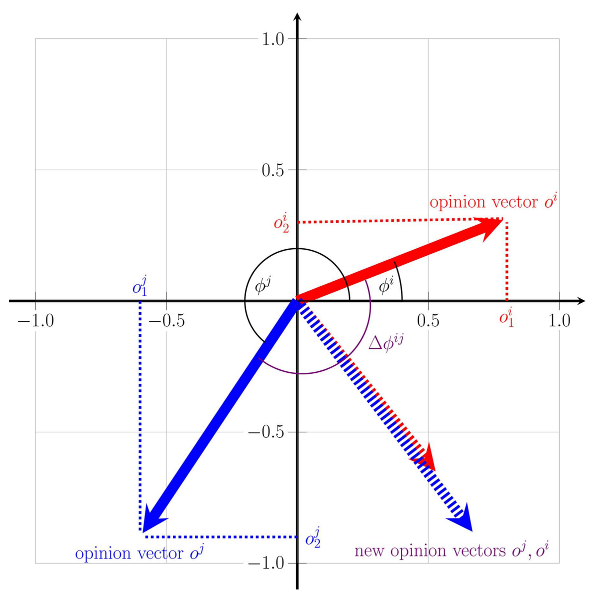

FIG. 4. Opinion vectors of two agents i and j in a two-dimensional opinion space. $\Delta\phi^{ij}$ is given by Eq. (8). The new opinion for both agents follows from Eq. (13).

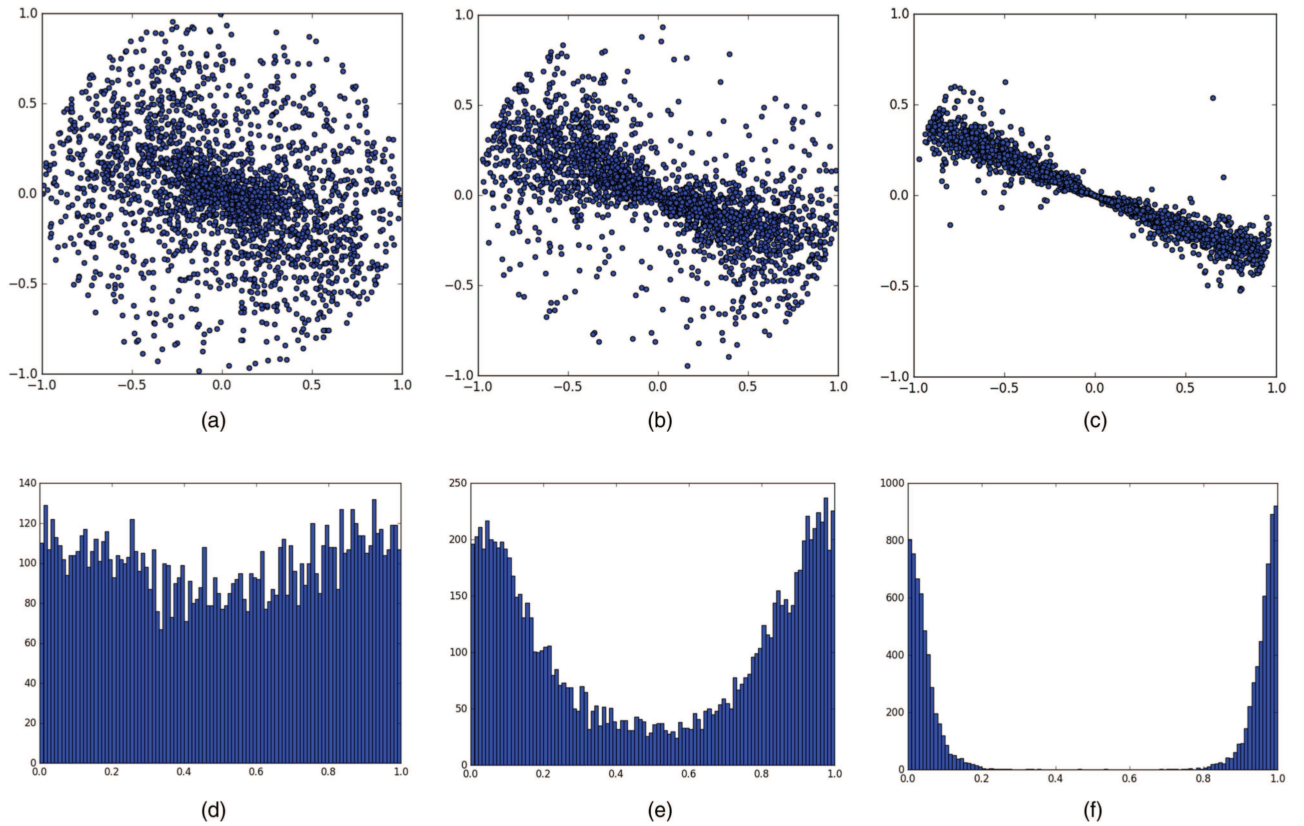

FIG. 5. (Top row) Opinions of $N = 2500$ agents in a two-dimensional opinion space at different time steps: (a) $t = 50\,000$, (b) $t = 70\,000$, and (c) $t = 100\,000$. (Bottom row) Distribution of the corresponding pairwise directional similarity $P[D^{ij}(t)]$, Eq. (9).

If we assume as before $\omega = 0.5$, we find in an explicit form,

$$\theta^i(t) = \xi(t)\phi^i(t) + \begin{cases} \dfrac{\left[\phi^j(t) + \phi^i(t)\right]}{2} & \text{if } \left|\phi^j(t) - \phi^i(t)\right| \le \pi, \\ \pi + \dfrac{\left[\phi^j(t) + \phi^i(t)\right]}{2} & \text{if } \left|\phi^j(t) - \phi^i(t)\right| > \pi. \end{cases} \tag{13}$$

This update rule is illustrated in Fig. 4, to be compared with Fig. 2 based on the Euclidean distance.

2. Model outcomes

We illustrate the outcome of the alignment model by means of agent-based simulations illustrated in Fig. 5. To make it comparable to Fig. 3, we first restrict ourselves to a two-dimensional opinion space. The initial state is the same as shown in Fig. 3(a) and follows from the setup described in Sec. III B. While the first row in Fig. 5 shows the positions of the agents in the two-dimensional opinion space at different times, the second row shows the distribution of the corresponding pairwise similarity measure, $D^{ij}(t)$, Eq. (9). The initial distribution of $D^{ij}(0)$, which matches the initial opinion distribution, Fig. 2(a), is shown in Fig. 7(a).

In the simulations shown in Fig. 5, the affective involvement e^i, which together with D^{ij} enters the function $\mathcal{F}[\cdot]$, Eq. (9), is set to a constant value $e^i \equiv e = 0.5$, equal for all agents. Random influences are set to zero, $\mathcal{Z}[\cdot] = 0$, Eq. (11); i.e., the dynamics are completely deterministic.

As we can see, in an early phase, the opinion vectors are broadly distributed, and the corresponding distribution $P[D^{ij}(t)]$, which reflects the angle $\Delta\phi^{ij}(t)$ between any two vectors, is almost uniform. This changes over time into a clear bimodal distribution. Its meaning becomes clear from the opinion positions in the two-dimensional space: agents tend to align their opinions such that a dominant direction emerges. Almost all agents align to this dominant direction but still position themselves on opposite sides. Hence, we do not observe consensus (which would also imply an alignment of opinions), but the coexistence of opinions from the left/right spectrum, i.e., polarization.

Note that, because of the assumed directional voting, agents do not adjust the magnitude of their opinion vectors, but just the angle. We add that a recent model built on a novel *weighted balance*

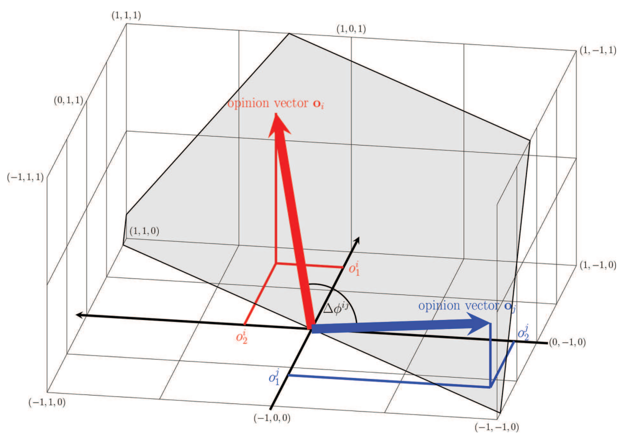

FIG. 6. Opinion vectors of two agents i and j in a three-dimensional opinion space. Note that the two vectors define a plane (in gray), on which the angle $\Delta\phi^{ij}$ is measured.

theory[57] is also able to reflect changes in the magnitude of the opinion vectors. Without that, we do not see a pronounced polarization, in which extreme opinions (with large magnitude) dominate. However, the emergence of a global alignment is clearly observed, which was the goal of this opinion dynamics model.

While our opinion dynamics model performs very promising in two dimensions, this raises the question how robust the outcome is if we change (i) interaction parameters or (ii) the dimensionality of the opinion space. Unfortunately, this robustness is not given, and in the following, we shortly explain the reasons for this, as a motivation for the model extension in Sec. III E.

The underlying dynamics assumes a critical threshold, in this case expressed by the affective involvement e. Very similar to the proximity voting model, this threshold decides about consensus or coexistence of opinions in the long run. Decreasing e for all agents increases the range of possible interactions because agents with a lower affective involvement have a larger tendency to change their opinions. This in turn destroys the coexistence of different opinions and fosters consensus. The same happens if we, instead of using a fixed value for all agents, increase the width of the distribution, $P[e]$, this way allowing more agents to have a lower value of e^i. Of course, there will be also more agents with large e^i, but what matters to reach consensus is the fraction of those agents that can still interact with others. If the agent i is no longer willing to approach the agent j in the opinion space, the agent j still can if its e^j value is low enough. Increasing the randomness in the dynamics by setting $\sigma_\xi > 0$ also favors the emergence of consensus.

The dimensionality M of the opinion space impacts the results in a less obvious, but interpretable manner. We recall that agents get assigned initial opinions on each dimension in a random manner. The angle between any two opinion vectors is, in a multi-dimensional space, calculated on the plane defined by the two vectors (see Fig. 6). This means, even in higher dimensions, there is only one angle $\Delta\phi^{ij}$. For $M = 2$, this can have initially any value between $(0, \pi)$ [shown in Fig. 7(a)]. However, the expectation value is $\langle \Delta\phi^{ij}(0) \rangle = \pi/2$. With each additional dimension, the probability to still find extreme values for $\Delta\phi^{ij}$ decreases and the distribution $P[D^{ij}(0)]$ narrows down toward the expectation value, $\langle D^{ij}(0) \rangle = 1 - \langle \Delta\phi^{ij}(0) \rangle / \pi = 0.5$. This can be clearly seen in Fig. 7, which shows the distribution of initial pairwise distances $P[D^{ij}(0)]$ for different dimensions. While this distribution is very broad and almost uniform for $M = 2$, it quickly approaches a uni-modal distribution centered around 0.5, if M is increased. In other words, it becomes unlikely for an agent to meet another agent with very similar or very dissimilar opinions. Most pairs of agents will have a mixture of congruent and opposing opinions.

For higher dimensions, the initial distribution $P[D^{ij}(0)]$ already ensures that all agents have (almost) the same alignment/disalignment of their opinion vectors, i.e., $\pi/2$. Then, the threshold value e, or the respective distribution $P[e]$, determines the outcome of the opinion dynamics. If the affective involvement is low, for example, $e \ll 0.5$, the majority of agents are able to interact, this way aligning their opinions even more, which eventually leads

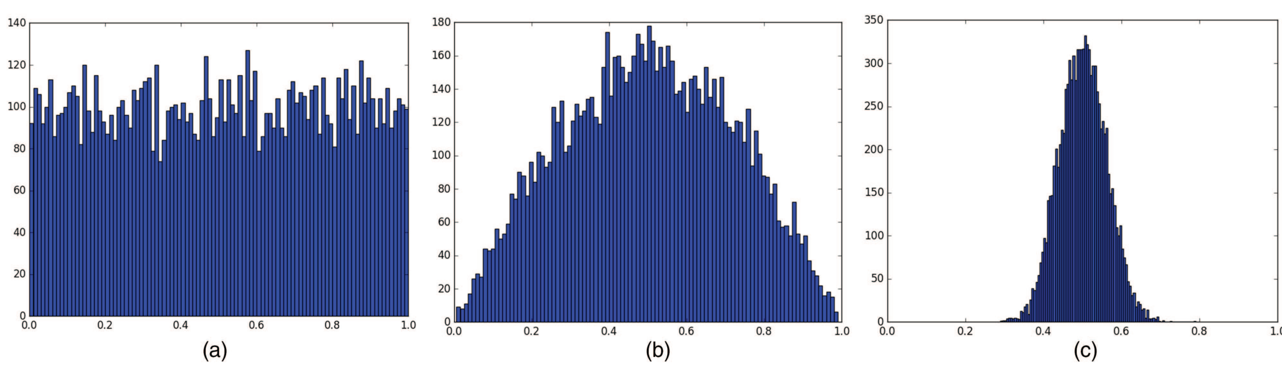

FIG. 7. Initial distributions of the pairwise directional similarity $P[D^{ij}(0)]$ for different dimensions of the opinion space: (a) $M = 2$, (b) $M = 3$, and (c) $M = 28$.

to a large alignment together with (almost perfect) consensus. This is shown in Fig. 16 of Appendix B. If the threshold is high, $e \gg 0.5$, the majority of agents is not able to interact. Then, their alignment distribution stays as a unimodal distribution centered around 0.5, very close to the initial distribution. In both cases, it is not possible to obtain the desired scenario of a bimodal alignment distribution, where agents align their opinions along the emerging *ideological dimension*, as shown in Fig. 5 for $M = 2$. The lack of diametrically opposed opinion vectors in the initial state makes the emergence of a polarized state very unlikely.

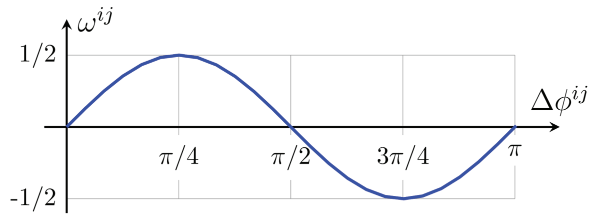

FIG. 8. Update parameter $\omega^{ij} = \omega[\Delta\phi^{ij}(t)]$, Eq. (15), dependent on the angle $\Delta\phi^{ij}$ between the opinion vectors of agents *i* and *j*.

E. Directional voting model with repulsion

1. Model description

So far, we have used a critical threshold to determine whether two agents still interact. This threshold has already become an individual parameter and was coupled to the affective involvement of the agents. Therefore, now, we go back to the argumentation in Sec. II A, where we discussed the options of two individuals that disagree on most issues and only agree on a few. Instead of not interacting, which we simulated in Secs. III B–III D, we now assume that they still interact but solve their cognitive dissonances by disagreeing even on the few issues they previously agreed on.

Because now all agent pairs can potentially interact, we have to change the respective function as follows:

$$\mathcal{F}\left[\mathbf{o}^i(t), \mathbf{o}^j(t)\right] = 1. \tag{14}$$

However, based on their interactions, two agents will not always align their opinion vectors. Instead, if $\Delta\phi^{ij}(t)$ is larger than a threshold δ, we assume that as a result of their interaction, they deviate even more in an opinion space. Precisely, the absolute distance between their updated angles, $\left|\theta^j(t) - \theta^i(t)\right|$ increases compared to $\Delta\phi^{ij}(t)$ if $\Delta\phi^{ij}(t) > \delta$, whereas it decreases if $\Delta\phi^{ij}(t) < \delta$. We set $\delta = \pi/2$, which means that agents with orthogonal opinion vectors do not influence each other's opinion. If their current alignment is less then $\pi/2$, they tend to align more, and if it is more than $\pi/2$, they tend to deviate even more.

This assumption is in line with the arguments in Secs. II A and II C because $\Delta\phi^{ij}(t)$ measures precisely whether agents agree or disagree on most issues. If $\Delta\phi^{ij}(t) < \pi/2$, they agree on most issues (but may still disagree on a few), and if $\Delta\phi^{ij}(t) > \pi/2$, it is the other way round. This means, $\Delta\phi^{ij}(t)$ effectively determines whether the relation between the two agents is positive, $r^{ij} = +1$, or negative, $r^{ij} = -1$. This way, we have implemented the theoretical arguments based on the combination of the cognitive dissonance theory and the structural balance theory in our agent-based model of opinion dynamics.

We can formally express this argument in our update function, $\mathcal{G}\left[\mathbf{o}^i(t), \mathbf{o}^j(t)\right]$, Eq. (10), if the parameter ω to update the angle, Eq. (12), becomes a function that depends on $\Delta\phi^{ij}(t)$ in a non-monotonous manner, for example,

$$\omega^{ij}(t) = \omega\left[\Delta\phi^{ij}(t)\right] = \frac{1}{2}\sin\left[2\Delta\phi^{ij}(t)\right]. \tag{15}$$

This is shown in Fig. 8. As we see, the previous dynamics, i.e., $\omega^{ij} = 1/2$, is regained if $\Delta\phi^{ij}(t) = \pi/4$. In this case, the two agents completely align their opinion; i.e., each one rotates its opinion vector by $\pi/8$ toward the other. Conversely, if $\Delta\phi^{ij}(t) = 3\pi/4$, we obtain $\omega^{ij} = -1/2$ and each agent rotates its opinion vector by $\pi/8$ away from the other.

We emphasize that the transition from alignment to disalignment is smooth and not abrupt. The largest changes in opinions occur when both the motivation to change opinions and the number of opinions that can be changed are high. This motivation is high if agents already have a sufficient agreement or disagreement on a number of issues. In cases of perfect alignment or disalignment, agents will not change their opinions based on the interaction with others because they already agree or disagree completely.

Finally, we need to specify the function $\mathcal{Z}\left[\mathbf{o}^i(t)\right]$ for the random influences. This still affects the angle ϕ^i, but instead of just considering random shocks $\xi(t)$, we now also consider the affective involvement e^i of an agent. Assuming that a higher emotional involvement in policy issues makes the opinion of an agent more resistant to a random change, we choose

$$\mathcal{Z}\left[\mathbf{o}^i(t)\right] = \phi^i(t)\,\xi(t)\left\{1 - e^i\right\}. \tag{16}$$

This means, agents with higher emotional involvement are less susceptible to noise. As before, ξ is sampled from a distribution with mean $\mu_\xi = 0$ and standard deviation σ_ξ, which regulates the overall level of randomness in the system.

Combining all these ingredients, the updated angle $\theta^i(t)$ is now slightly different from Eq. (12),

$$\theta^i(t) = \phi^i(t)\left[1 + \xi(t)\left\{1 - e^i\right\}\right] + \omega^{ij}(t)\Delta\phi^{ij}(t), \tag{17}$$

where $\omega(t)$ follows from Eq. (15) and $\mathcal{G}\left[\mathbf{o}^i(t), \mathbf{o}^j(t)\right]$ is still given by Eq. (10).

2. Model outcomes

In Fig. 9, we present the results for the multi-dimensional opinion space with $M = 28$, using the same initial setup as before. We first highlight that our opinion dynamics model is indeed able to produce an outcome with a bimodal pairwise directional similarity distribution. This is achieved despite the fact that the initial distribution, because of the high dimensionality, is unimodal and quite narrow, as shown in Fig. 9(a). Second, we note that this outcome is robust because it is achieved even in those cases where a broader distribution $P[e]$ and an increased noise level are considered.Different

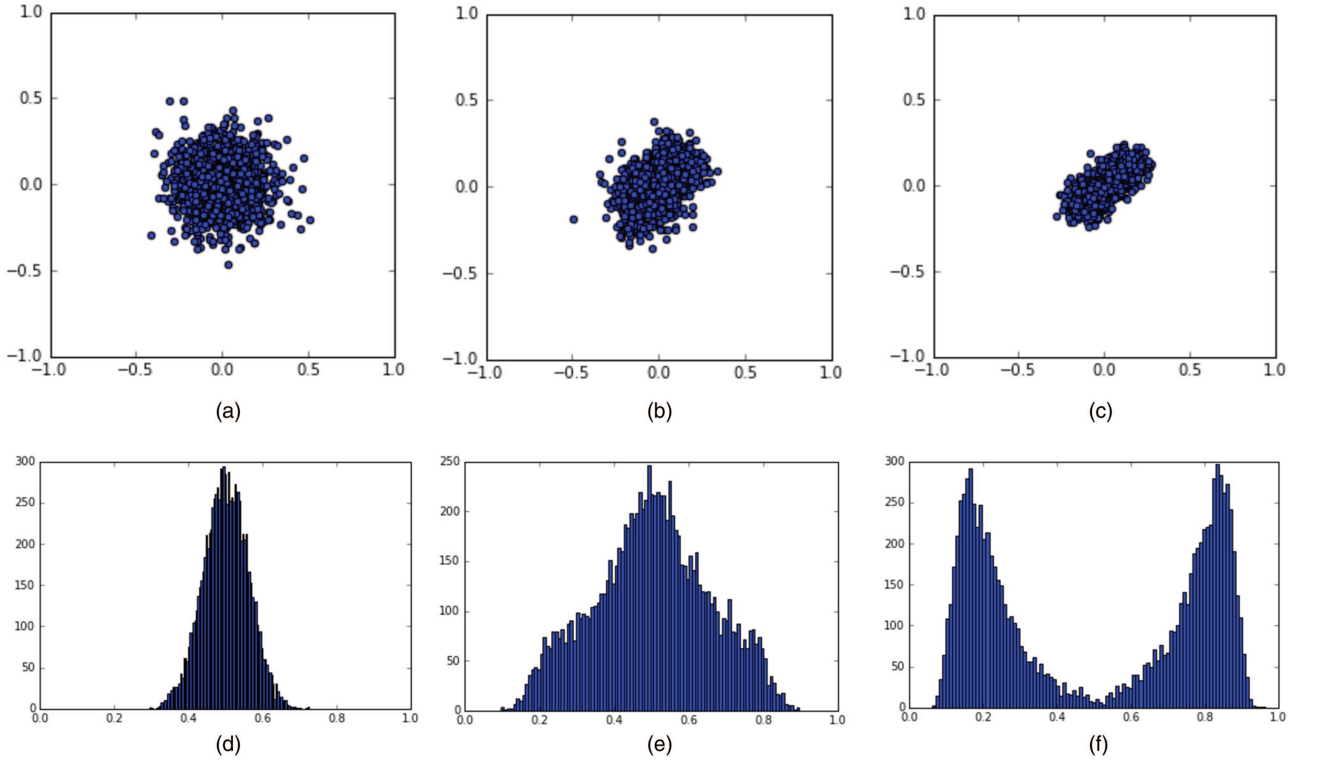

FIG. 9. Opinions of $N = 2500$ agents in a multi-dimensional opinion space ($M = 28$) at different time steps: (a) $t = 0$, (b) $t = 380\,000$, and (c) $t = 500\,000$. (Top row) Agents' positions on the first two opinion dimensions $\mathbf{s}_1$ and $\mathbf{s}_2$. (Bottom row) Distribution of the corresponding pairwise directional similarity $P[D^{ij}(t)]$, Eq. (9). Further parameters: $\mu_e = 0.6$, $\sigma_e = 0.5$ for the affective involvement and $\sigma_\xi = 0.2$ for the noise.

from the case discussed above, this does not lead to consensus in the end. Alternatively, we clearly observe the emergence of polarized opinions along a dominant ideological dimension. The scatterplots in Fig. 9 show how the opinion vector endpoints slowly align themselves to one dimension, pointing in both directions from the origin. Furthermore, similarity histograms show that the model produces a polarized state: In the last stage, there is about an equal number of pairs of agents with very high and with very low similarity values—exactly what is expected from a polarized outcome.

IV. ISSUE ALIGNMENT AND AFFECTIVE INVOLVEMENT

A. Global alignment

In this section, we further explore the emergence of a dominant opinion dimension in an M-dimensional opinion space in our last model (Sec. III E) and, in particular, how this emergence is dependent on the level of affective involvement e. One way to extract this dominant dimension from the simulated data is the *principal component analysis* (PCA). By means of an orthogonal transformation of the original opinions, $\mathbf{o}^i(t) = o^i_1(t)\mathbf{s}_1 + o^i_2(t)\mathbf{s}_2 + \cdots + o^i_M(t)\mathbf{s}_M$, in the opinion space $\mathbf{s}_1, \ldots, \mathbf{s}_M$, PCA identifies the principal components $\mathbf{c}_1, \ldots, \mathbf{c}_M$, i.e., the axes of a transformed opinion space, such that $\mathbf{c}_1$ shows ("explains") the largest variance in the data, $\mathbf{c}_2$ the second largest, etc. In most cases, already the first few principal components $m = 1, 2, 3$ are sufficient to explain the variance observed. Compared to the M dimensions, this means a dimensionality reduction, which necessarily implies a loss of information. However, PCA is a way to minimize this loss of information, provided that certain assumptions, for instance, about linear correlations between the opinions hold.

In the following, we only use the first component, $\mathbf{c}_1$, based on the insight that the main ideological dimension was found to be the most important one explaining real opinion distributions on policy issues. In order to calculate how much of the variance is explained by $\mathbf{c}_1$, we have to follow the standard procedure of PCA, just summarized here: (a) center the data, i.e., $o^i_1 - \langle o \rangle_1$ etc., (b) compute the covariance matrix, (c) calculate the eigenvalues and corresponding eigenvectors, (d) normalize the eigenvectors to unit vectors, and (e) transform the covariance matrix into a diagonal matrix. The diagonal elements of this matrix then give us the variance explained by the corresponding axes. This means, the largest eigenvalue λ_1 refers to the variance explained by the first principal component

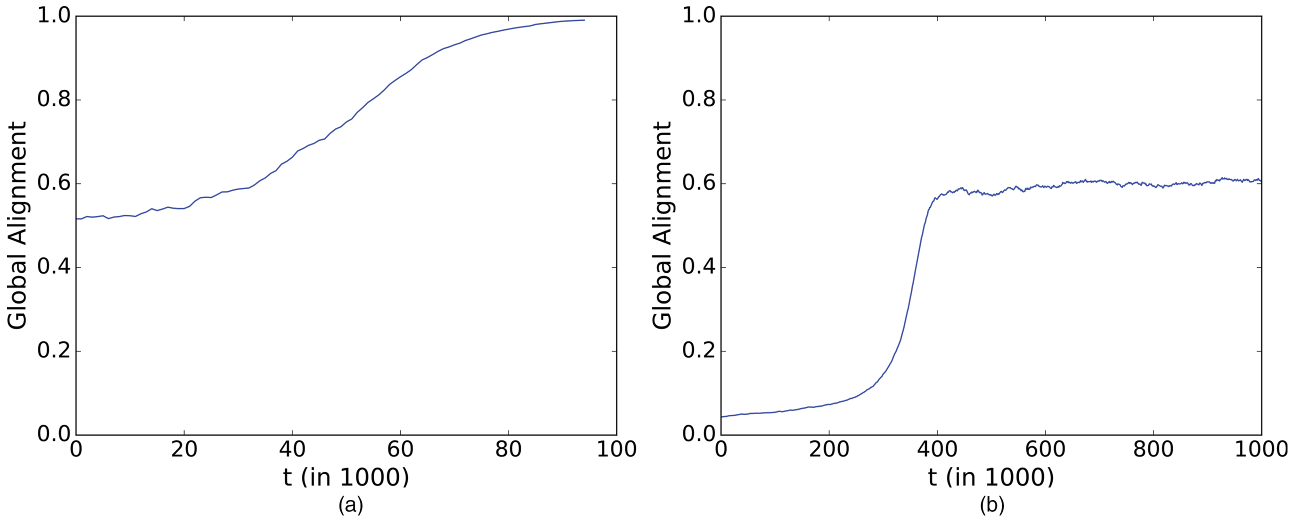

FIG. 10. Global alignment $A(t)$ over time measured in simulation steps. (a) Model with opinion alignment, Eqs. (9), (10), and (13) and no noise, for $M = 2$, also shown in Fig. 5. (b) Model with opinion alignment and repulsion, Eqs. (10), (14), and (17), for $M = 28$, also shown in Fig. 9.

$\mathbf{c}_1$. In the following, we define this as our measures of *global alignment*, A,

$$A = \mathrm{Var}(\mathbf{c}_1) = \lambda_1. \tag{18}$$

If $A = 1$, all individual opinion vectors lie on the dominant ideological dimension. The lowest value of global alignment that can be attained is $A = 1/M$, meaning that there is no global alignment whatsoever. In the following, we use the term "global alignment" to distinguish it from individual alignment, which we are going to define further below. Here, just note that A is not defined as an average over individual alignments. Figure 10 illustrates how the global alignment A evolves over time for the two simulation runs shown in Figs. 5 and 9. Figure 10(a) shows the almost perfect global alignment along the dominating dimension, for $M = 2$. In Fig. 10(b), we clearly see that due to the noise, the model never comes completely to rest and never fully aligns to the first component of the PCA because of the high number of dimensions, $M = 28$. However, after 400 000 iterations, the global alignment stabilizes around a relatively high value of 0.6.

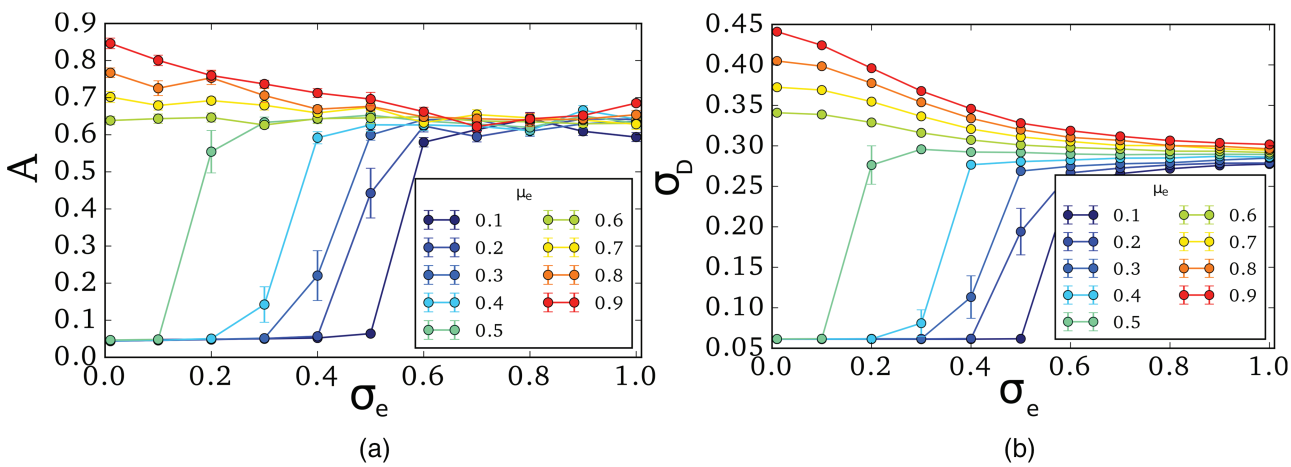

FIG. 11. (a) Global alignment A for varying distributions $\mathcal{N}(\mu_e, \sigma_e)$ to describe the affective involvement of agents. Different lines refer to different values of μ_e, the x-axis to different values of σ_e. (b) Standard deviation σ_D of the distribution $P[D^{ij}]$ of pairwise directional similarities. Error bars indicate the standard errors with respect to the mean.

B. Impact of affective involvement

We now investigate how the global alignment A depends on the second variable that characterizes each agent, the affective involvement $e^i \in [0, 1]$. We focus on the opinion dynamics with alignment and repulsion and concentrate on the high-dimensional opinion space, $M = 28$. A discussion of how the affective involvement impacts the model with only alignment was already given above.

We recall that the value of e^i is constant over time but in general drawn from a truncated normal distribution $\mathcal{N}(\mu_e, \sigma_e)$. This means, varying μ_e between 0 and 1 allows us to increase the expected value for the affective involvement, which decreases the ability to randomly change the opinion as it directly impacts $\mathcal{Z}[\cdot]$, Eq. (16). Varying σ_e between 0 and 1, on the other hand, allows making agents more heterogeneous regarding this ability. Thus, high values of μ_e combined with low values of σ_e would refer to a deterministic limit, while low values of μ_e combined with low values of σ_e would refer to a random limit, where all agents are impacted by the noise in the same (large) manner.

In our agent-based computer simulations, we vary both μ_e and σ_e in steps of 0.1. For each combination (μ_e, σ_e), we run 10 simulations for 1 000 000 time steps to ensure that a quasi-stationary global alignment is reached [see also Fig. 10(b)]. Due to the noise, the simulations never reach a completely stationary state. To determine the values of ξ^i, we sample from the truncated normal distribution $\mathcal{N}(\mu_\xi, \sigma_\xi)$ with $\mu_\xi = 0$ and $\sigma_\xi = 0.2$.

The results are shown in Fig. 11(a). We clearly see that for large μ_e, i.e., a low overall level of randomness in the dynamics, the global alignment A is always high. In the deterministic limit, it reaches a level of 85%, and even for a large heterogeneity in the agents' affective involvement, it is still above 60%. This contrasts with the random limit of small μ_e, where the global alignment drops to zero if σ_e is below 0.5. If it is above 0.5, i.e., if by chance there are still sufficiently many agents with a larger affective involvement, this again allows a global alignment of opinions. Hence, we confirm again that our opinion dynamics with alignment and repulsion is very robust against variations in the agent's parameters. On the other hand, we find that the transition between aligned and non-aligned global states is rather steep. This means, there is a critical level of randomness that can destroy the global alignment of opinions, as it should be rightly expected.

The second variable to characterize the alignment of agents' opinions to the dominant dimension is the pairwise directional alignment $D^{ij}(t) = 1 - \Delta\phi^{ij}(t)/\pi$. We have shown in Figs. 5 and 9 that the histogram of these values over time approaches a bimodal distribution. This indicates that the system develops a state of polarization where agents form clusters of opinions along opposite directions of the dominant dimension. This means that there is a coexistence of opposite opinions in the long term, whereas a unimodal distribution would refer to consensus. These two outcomes can be characterized by the standard deviation σ_D of the distribution $P[D]$. While the mean of this distribution is in both cases $\mu_D \approx 0.5$, consensus would refer to small values of σ_D, while polarization refers to large values of σ_D. We note that because the values for D^{ij} are bound between 0 and 1, a large value of σ_D means 0.5.

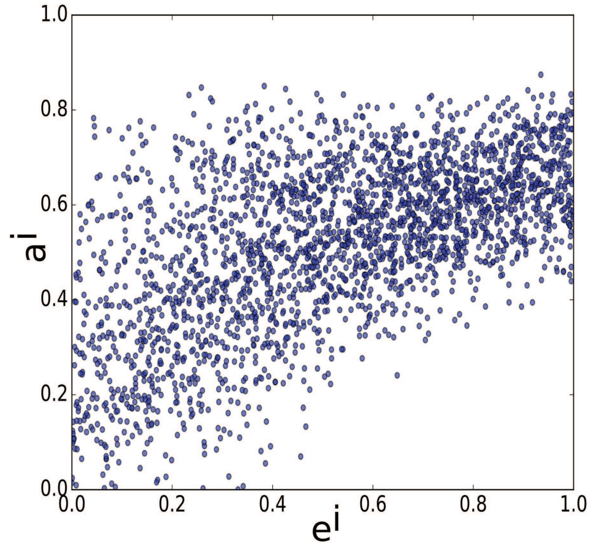

FIG. 12. Scatterplot of individual alignment a_i vs emotional involvement a_i for the results shown in Fig. 10(b) with $M = 28$.

We have also investigated how the distribution $P[D]$ depends on the distribution of the affective involvement $\mathcal{N}(\mu_e, \sigma_e)$. The results are shown in Fig. 11(b). For the deterministic limit of large μ_e and small σ_e, we find high values for σ_D, i.e., a clear polarization. This even holds if the heterogeneity of the agents' affective involvement e^i is increased. For the random limit of small μ_e and small σ_e, on the other hand, we see that the global polarization is destroyed by the noise; alternatively, consensus is obtained. This is in line also with our previous discussions that an increased noise level fosters consensus. Again, we note that the transition between consensus and polarization is rather steep; i.e., there exists a critical level of randomness.

C. Individual alignment

We also investigate how the affective involvement e^i of individual agents impacts their individual ability to align to the dominating opinion dimension. For this alignment, we can define an angle $\psi^i\left[\mathbf{o}^i, \mathbf{c}_1\right]$ between the individual opinion vector $\mathbf{o}^i$ and the main PCA component $\mathbf{c}_1$. The agent i is perfectly aligned to the main ideological dimension if $\psi^i = 0$; i.e., the opinion vector points into the direction of the PCA component $\mathbf{c}_1$ or if $\psi^i = \pi$, i.e., the opinion vector points into the direction opposite to $\mathbf{c}_1$. The latter case indicates individual opposition and forms the basis for polarization. However, still, this opposition can be expressed in terms of the ideological dimension. Thus, we define the *individual alignment* $a^i(t)$ of the agent i to the main ideological dimension as

$$a^i(t) = \left| \frac{2\psi^i(t)}{\pi} - 1 \right|; \qquad (19)$$

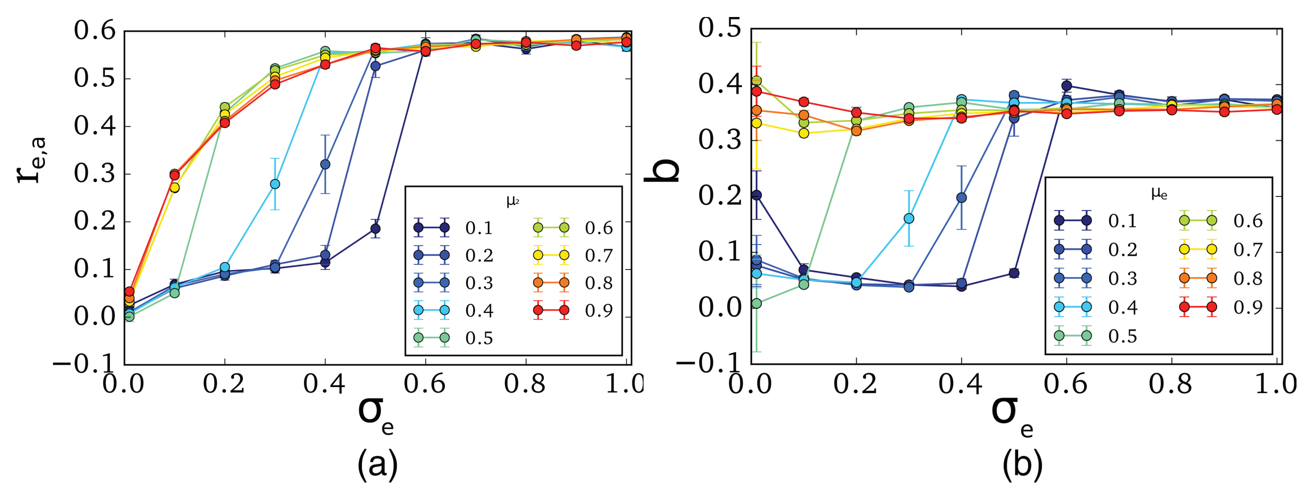

FIG. 13. (a) Correlation $r_{e,a}$, Eq. (20), between individual affective involvement e^i and individual alignment a^i and (b) the regression parameter b, Eq. (21), of e^i on a^i for varying distributions $\mathcal{N}(\mu_e, \sigma_e)$ to describe the affective involvement of agents. Different lines refer to different values of μ_e, the x-axis to different values of σ_e. Error bars indicate the standard errors with respect to the mean.

$a^i = 0$ if the opinion vector $\mathbf{o}^i$ is orthogonal to the dominant ideological dimension $\mathbf{c}_1$ and $a^i = 1$ if it is either 0 or π, pointing in either direction from the origin.

The scatterplot shown in Fig. 12 gives us a first indication of how the two agent variables e^i and a^i relate. We see that a higher affective involvement, i.e., a lower level of random opinion changes, indeed correlates with a higher level of individual alignment. This reflects the dissolving role of noise on alignment, already discussed for the global alignment A. We note again that A is not defined as an average over individual alignments.

To study this relation in a more systematic manner, we repeat the simulation procedure used in Sec. IV B. We define the Pearson correlation coefficient between e^i and a^i as

$$r_{e,a} = \frac{N\sum_i e^i a^i - \sum_i e^i \sum_i a^i}{\sqrt{N\sum_i (e^i)^2 - \left(\sum_i e^i\right)^2}\sqrt{N\sum_i (a^i)^2 - \left(\sum_i a^i\right)^2}}. \tag{20}$$

We then vary the distribution of affective involvement $\mathcal{N}(\mu_e, \sigma_e)$ from which the e^i are sampled to see how this impacts $r_{e,a}$. The results are shown in Fig. 13(a). We find again that in the random limit, this correlation breaks down. For the deterministic limit of large μ_e, we see that the correlations also decrease with σ_e. This is quite obvious because $\sigma_e \to 0$ means that all agents have the same affective involvement $e^i \to e$. Their individual alignment a^i may still vary, but its correlation with a constant e is zero.

Eventually, we can also analyze the scatterplot of Fig. 12 by means of a linear regression model,

$$a^i = c + b\,e^i + \varepsilon^i, \tag{21}$$

where ε^i is the error term (residual) and c is the intercept. The regression parameter b varies with the parameters of the distribution $\mathcal{N}(\mu_e, \sigma_e)$ as shown in Fig. 13(b). Again, we see that in the random limit of small μ_e, b is close to zero; i.e., correlations are low, while in the deterministic limit, its value is reasonably large. The phase transition at a critical σ_e is also clearly visible.

V. CONCLUSION

A. Agent-based modeling

In this paper, we apply agent-based modeling as one possible methodology to understand an empirically observed macro-phenomenon, in our case the alignment of individual opinions. Agent-based modeling requires us to provide reasonable micro-mechanisms of how agents influence another in their opinions. We base the proposed mechanisms in established psychological theories, notably (structural) balance theory and the theory of directional voting.

Our model then allows testing how different assumptions about such mechanisms impact the macroscopic dynamics. This means, we do not follow a data-driven modeling approach, which tries to reproduce a specific real-world outcome by estimating interaction parameters from observations.[67] Alternatively, we aim at a *generative explanation*,[21] a thought experiment to find out which mechanisms are necessary and sufficient to generate a stylized version of empirical reality—and, perhaps even more important, which mechanisms are not sufficient. By adding or removing mechanisms and tuning model parameters, we can improve our model step by step until we finally attain a model that is able to reproduce the desired macro-phenomenon. This of course does not prove the validity of the model assumptions, but it is a clear indicator which modeling hypotheses are compatible with a given macroscopic outcome.

B. Obtaining global issue alignment and polarization

Our model shall be able to reproduce two features of opinion dynamics observed in the political domain: (i) The emergence of global issue alignment: Individual opinions on different policy issues are correlated such that a dominant "left–right" ideological dimension can explain most of them. (ii) A polarization of opinions on this ideological dimension: While individual opinion vectors are aligned, they point to opposite directions from the origin.

Global issue alignment already assumes an underlying *multi-dimensional opinion space*, which is neglected in many opinion dynamics models. It means that instead of a scalar value, opinions are characterized by vectors in this M-dimensional space. The existence of global alignment was empirically demonstrated, and it was also theoretically discussed in political science with respect to voting and coalition formation.[2,37,49] However, there is still a lack of models that are able to generate this phenomenon.

We fill this research gap by investigating the conditions under which global issue alignment is obtained. Specifically, in our model, we vary (a) what information about the opinions of others agents take into account and (b) how they respond to this information. We have shown that the so called *proximity voting*, which is equivalent to using the Euclidean distance to evaluate the similarity of opinions, fails to generate a global issue alignment. *Directional voting*, however, in which agents measure similarities of opinions dependent on the "right" and "wrong" side, has the potential to generate global alignment, at least in low-dimensional opinion spaces. If we combine directional voting with a repulsive force between far-distant opinions, we find that global issue alignment also emerges in high-dimensional opinion spaces and is very robust against parameter changes in the model.

The repulsive force is not just postulated to improve the model, but it is motivated by the mentioned psychological mechanisms of cognitive balance. The tendency to increase cognitive balance could either lead to an attractive force, i.e., opinions of agents become more similar to minimize the dissonance, or to a repulsive force, i.e., opinions of agents become more different. Hence, we propose a reasonable micro mechanism. Even more, we also demonstrate how these assumptions can be formalized in an opinion dynamics model by providing a formal model in polar coordinates.

C. Affective involvement

As an asset, our model includes an emotional component in the opinion dynamics. This extension is rooted in arguments from political science theory that global issue alignment is driven by "*passion*," i.e., affective involvement in politics.[54] We implement the emotional component in our agent-based model by means of a heterogeneous parameter, e^i, drawn from a distribution $\mathcal{N}(\mu_e, \sigma_e)$ with given mean and variance. The higher the level of affective involvement, the more resistant agents are to change their opinions. We discuss two different ways of implementing this relation (a) by defining a threshold for directional similarity and (b) by impacting the level of random opinion changes.

In our paper, we systematically study the influence of affective involvement on global issue alignment and on individual alignment. We find that two different types of outcome can be observed: If the level of affective involvement, expressed by μ_e, is low and the heterogeneity across agents, expressed by σ_e, is also low, we end up in regime with little global alignment, low opinion polarization, and little individual alignment. In this disorganized state, no dominant ideological dimension emerges to which agents align their opinions.

If on the other hand, the heterogeneity of agents' affective involvement is high, which means that (for both low and high μ_e) there is a sufficiently a large number of agents with a high level of affective involvement, we always find outcomes with high global issue alignment, high opinion polarization, and high individual alignment. This is a highly organized state in the opinion space, and we have pointed out that there is a rather sharp transition between the disorganized and the organized states dependent on the parameters of the affective involvement. Therefore, we can conclude that affective involvement, the way it is considered in our model, fosters the global issue alignment, as argued also by political scientists. Even more, global alignment can only be observed beyond a critical level of affective involvement.

APPENDIX A: SIMULATION OF THE TWO-DIMENSIONAL BOUNDED CONFIDENCE MODEL

Here, we present snapshots of the dynamics for which the initial state is shown in Fig. 3(a) and the final states in Figs. 3(b) and 3(c). The model parameters are chosen as follows: $\mu_o = 0$ and $\sigma_o = 0.8$ for the initial opinions and $\mu_e = 0.5$ and $\sigma_e = 0$ for the emotional involvement; i.e., all agents have the same $e^i \equiv e = 0.5$. For the confidence interval, two different values are chosen: $\varepsilon = 0.5$ in Figs. 14(a) and 14(b) and $\varepsilon = 0.25$ in Figs. 14(c) and 14(d).

To further quantify the dynamics of the agent-based model, we analyze the evolution of the pairwise Euclidean similarity,

$$S^{ij}(t) = 1 - \frac{d^{ij}(t)}{\sqrt{8}}. \tag{A1}$$

$S^{ij}(t)$ is a linear transformation of the pairwise Euclidean distance and is shown in Fig. 15. We see that initially, the distribution $P[S]$ is rather broad, but becomes more narrow over time, to converge almost to a delta peak at $S = 1$. This means that the pairwise similarity is maximized for all agents. In the case of consensus, this happens because all agents have reached the same opinion vector, also shown in Fig. 2(b). In the case of coexistence, the outcome is almost identical because the opinion clusters in the periphery, also shown in Fig. 2(c), contain only very few agents. Hence, the very small contribution at about $S = 0.7$ in Fig. 15(c) is barely noticeable. After all, this configuration is stable because those agents that still interact have reached the same opinion vector and belong to the same cluster in the opinion space.

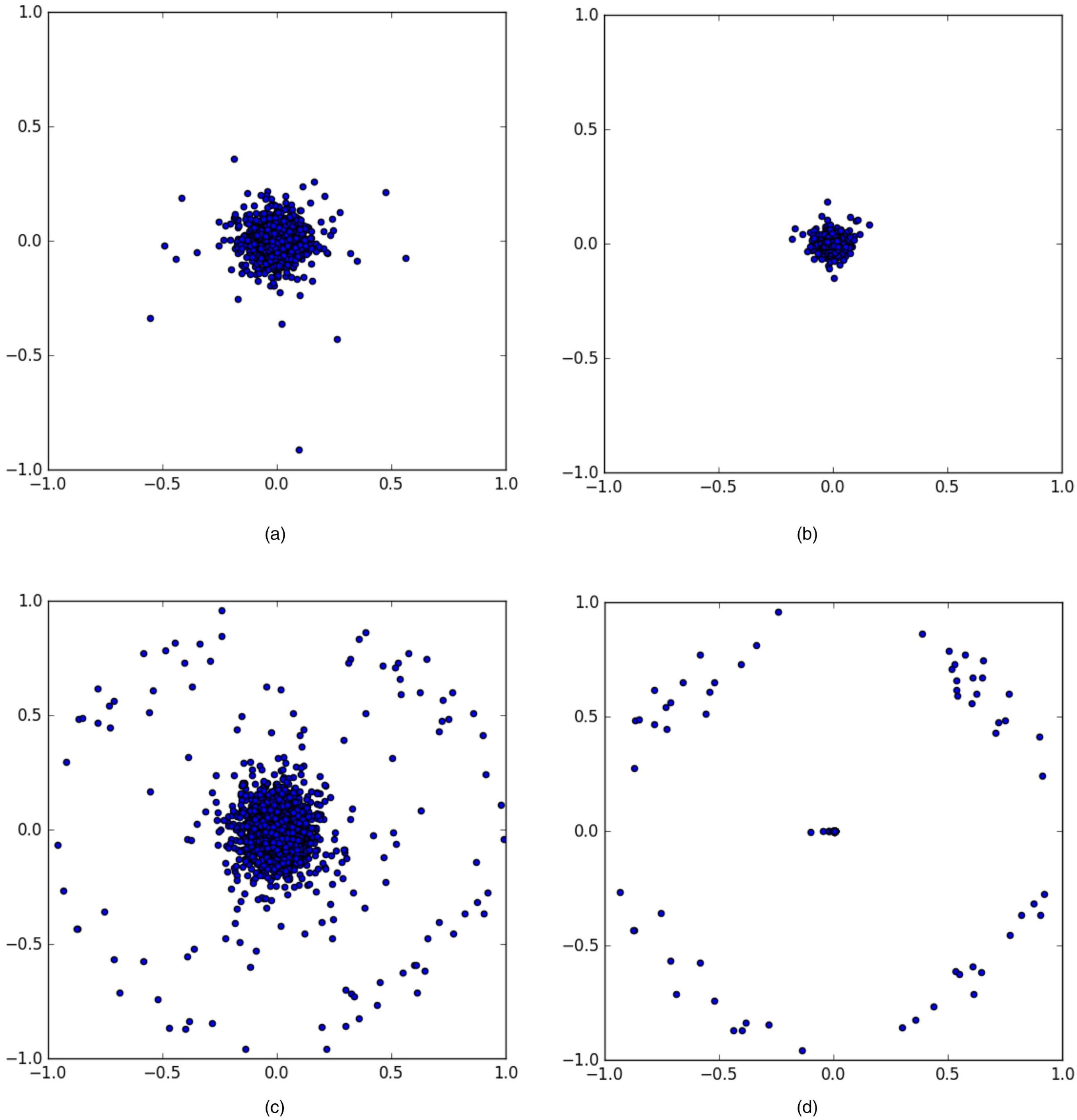

FIG. 14. Snapshots of the positions of agents in the two-dimensional opinion space: (a) and (b) $\varepsilon = 0.5$ and (c) and (d) $\varepsilon = 0.25$. (a) $t = 30\,000$, (b) $t = 40\,000$, (c) $t = 30\,000$, and (d) $t = 90\,000$.

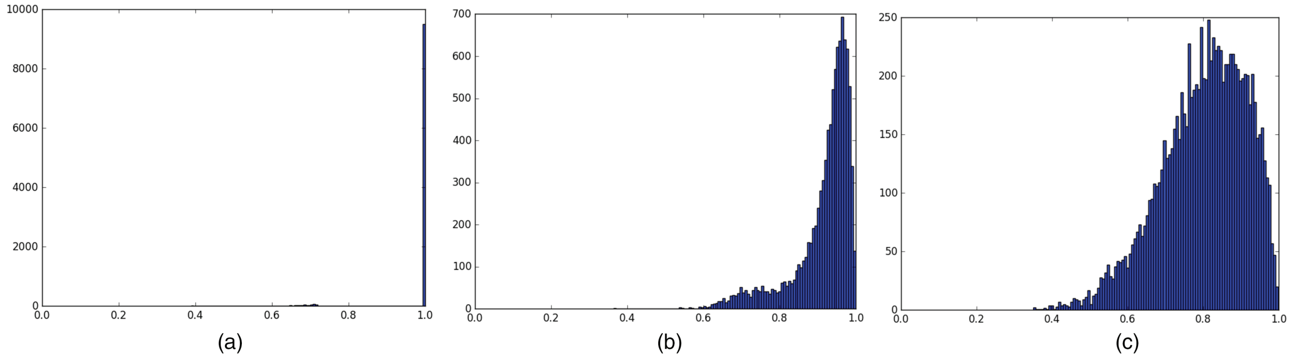

FIG. 15. Distribution of the pairwise Euclidean similarity $S^{ij}(t)$, Eq. (A1), at different time steps: (a) $t = 0$; for the distribution in the opinion space, see Fig. 2(a). (b) $t = 30\,000$. (c) $t = 210\,000$. For parameters, see Fig. 14, $\varepsilon = 0.25$.

APPENDIX B: SIMULATION OF THE OPINION ALIGNMENT IN $M = 28$ DIMENSIONS

In Fig. 16, we present results of the multi-dimensional opinion alignment without repulsion. The results are discussed in Sec. III D.

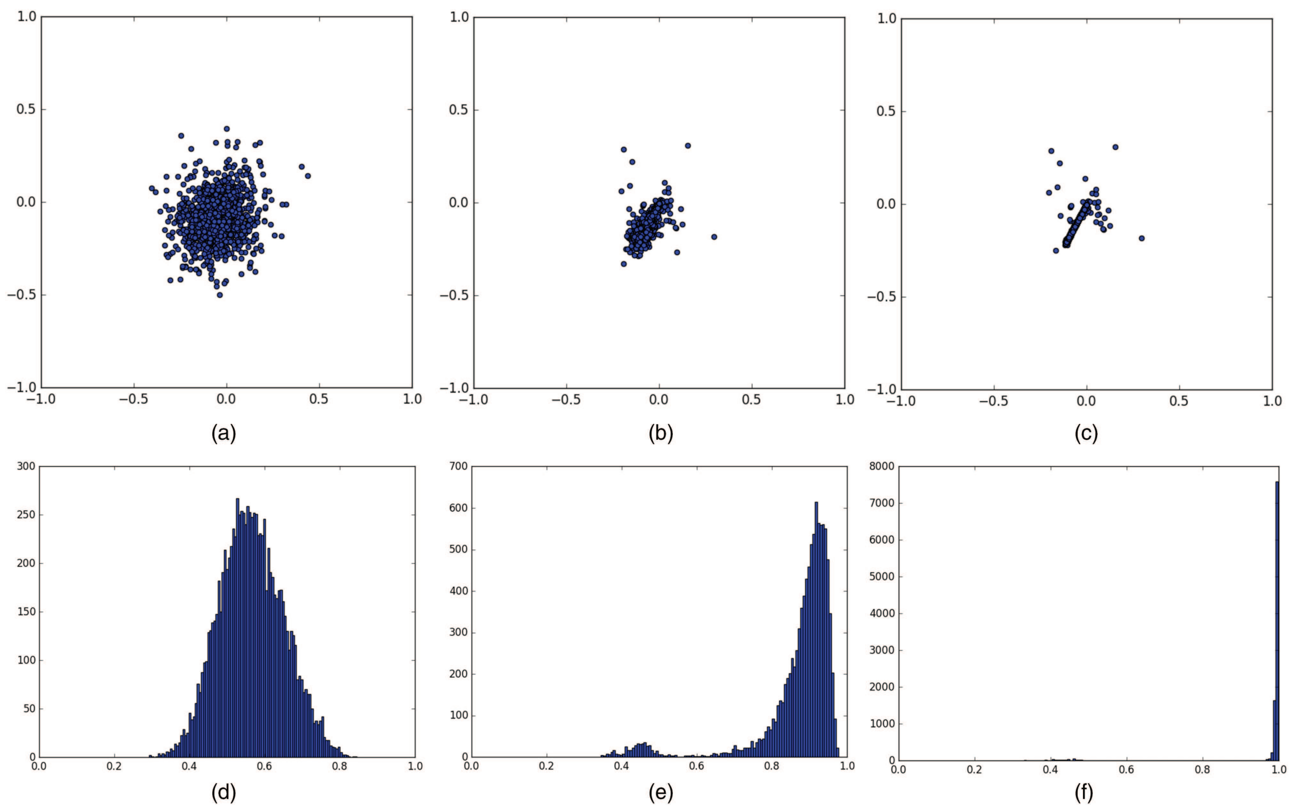

FIG. 16. Opinions of $N = 2500$ agents in a multi-dimensional opinion space ($M = 28$) at different time steps: (a) $t = 50\,000$ and (b) $t = 70\,000$ and $t = 100\,000$. (top row) The projection of the opinions on the space of the two principal components $\mathbf{c}_1$, $\mathbf{c}_2$ is shown. (bottom row) Distribution of the pairwise directional similarity, $P[D^{ij}(t)]$, Eq. (9). (c) $P[D^{ij}(0)]$ is shown in Fig. 7(c). Further parameters: $e^i \equiv e = 0.5$, no noise.

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study. The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]O. Artime Vila, A. Carro, A. F. Peralta, J. J. Ramasco, M. San Miguel, and R. Toral, "Herding and idiosyncratic choices: Nonlinearity and aging-induced transitions in the noisy voter model," C. R. Phys. **20**(4), 262–274 (2019).
[2]D. Baldassarri and A. Gelman, "Partisans without constraint: Political polarization and trends in American public opinion," Am. J. Sociol. **114**(2), 408 (2008).
[3]K. Benoit and M. Laver, "The dimensionality of political space: Epistemological and methodological considerations," Eur. Union Politics **13**(2), 194–218 (2012).
[4]K. Benoit and M. Laver, *Party Policy in Modern Democracies* (Routledge, 2006).
[5]S. Bornschier, "The new cultural conflict, polarization, and representation in the Swiss party system, 1975–2011," Swiss Political Sci. Rev. **21**(4), 680–701 (2015).
[6]A. Carro, R. Toral, and M. San Miguel, "The role of noise and initial conditions in the asymptotic solution of a bounded confidence, continuous-opinion model," J. Stat. Phys. **151**(1–2), 131–149 (2013).
[7]A. Carro, R. Toral, and M. San Miguel, "The noisy voter model on complex networks," Sci. Rep. **6**(1), 24775 (2016).
[8]D. Cartwright and F. Harary, "Structural balance: A generalization of Heider's theory," Psychol. Rev. **63**(5), 277 (1956).
[9]C. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics," Rev. Mod. Phys. **81**(2), 591 (2009).
[10]M. D. Conover, B. Gonçalves, A. Flammini, and F. Menczer, "Partisan asymmetries in online political activity," EPJ Data Sci. **1**(1), 6 (2012).
[11]P. E. Converse, "The nature of belief systems in mass public," in *Ideology and Discontent*, edited by D. E. Apter (Free Press, New York, 1964).
[12]G. Deffuant, F. Amblard, G. Weisbuch, and T. Faure, "How can extremism prevail? A study based on the relative agreement interaction model," J. Artif. Soc. Soc. Simul. **5**(4) (2002); available at https://ideas.repec.org/a/jas/jasssj/2002-25-2.html.
[13]G. Deffuant, I. Bertazzi, and S. Huet, "The dark side of gossips: Hints from a simple opinion dynamics model," Adv. Complex Syst. **21**(06n07), 1850021 (2018).
[14]G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, "Mixing beliefs among interacting agents," Adv. Complex Syst. **3**(01n04), 87–98 (2000).
[15]D. DellaPosta, Y. Shi, and M. Macy, "Why do liberals drink lattes?," Am. J. Sociol. **120**(5), 1473–1511 (2015).
[16]P. DiMaggio, J. Evans, and B. Bryson, "Have American's social attitudes become more polarized?," Am. J. Sociol. **102**(3), 690–755 (1996).
[17]M. Dimock, C. Doherty, J. Kiley, and R. Oates, "Political polarization in the American public," Pew Research Center, 2014.
[18]I. Dornic, H. Chaté, J. Chave, and H. Hinrichsen, "Critical coarsening without surface tension: The universality class of the voter model," Phys. Rev. Lett. **87**(4), 045701 (2001).
[19]J. N. Druckman, E. Peterson, and R. Slothuus, "How elite partisan polarization affects public opinion formation," Am. Polit. Sci. Rev. **107**(01), 57–79 (2013).
[20]B. Edmonds *et al.*, "Assessing the safety of (numerical) representation in social simulation," in *Agent-Based Computational Modelling* (Physica Verlag, 2005), pp. 195–214.
[21]J. M. Epstein, *Generative Social Science: Studies in Agent-Based Computational Modeling* (Princeton University Press, 2006).
[22]J. Fernández-Gracia, K. Suchecki, J. J. Ramasco, M. San Miguel, and V. M. Eguíluz, "Is the voter model a model for voters?," Phys. Rev. Lett. **112**(15), 158701 (2014).
[23]A. Flache, M. Mäs, T. Feliciani, E. Chattoe-Brown, G. Deffuant, S. Huet, and J. Lorenz, "Models of social influence: Towards the next frontiers," J. Artif. Soc. Soc. Simul. **20**(4) (2017).
[24]S. Galam, "Democratic voting in hierarchical structures or how to build a dictatorship," Adv. Complex Syst. **03**(01n04), 171–180 (2000).
[25]D. Garcia, A. Abisheva, S. Schweighofer, U. Serdült, and F. Schweitzer, "Ideological and temporal components of network polarization in online political participatory media," Policy Internet **7**(1), 46–79 (2015).
[26]D. Garcia, F. Mendez, U. Serdult, and F. Schweitzer, "Political polarization and popularity in online participatory media: An integrated approach," in *Proceedings of the 1st Workshop on Politics, Elections and Data—PLEAD '12* (ACM, 2012), pp. 3–10.
[27]D. Grechyna, "On the determinants of political polarization," Econ. Lett. **144**, 10–14 (2016).
[28]P. Groeber, J. Lorenz, and F. Schweitzer, "Dissonance minimization as a microfoundation of social influence in models of opinion formation," J. Math. Sociol. **38**(3), 147–174 (2014).
[29]P. Groeber, F. Schweitzer, and K. Press, "How groups can foster consensus: The case of local cultures," J. Aritif. Soc. Soc. Simul. **12**(2), 1–22 (2009); available at http://jasss.soc.surrey.ac.uk/12/2/4.html.
[30]P. Hedström, *Dissecting the Social: On the Principles of Analytical Sociology* (Cambridge University Press, Cambridge, 2005).
[31]R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis, and simulation," J. Artif. Soc. Soc. Simul. **5**(3) (2002); available at http://jasss.soc.surrey.ac.uk/5/3/2.html.
[32]R. Hegselmann and U. Krause, "Opinion dynamics under the influence of radical groups, charismatic leaders, and other constant signals: A simple unifying model," Networks Heterogeneous Media **10**(3), 477 (2015).
[33]F. Heider, "Attitudes and cognitive organization," J. Psychol. **21**(1), 107–112 (1946).
[34]M. J. Hetherington, "Review article: Putting polarization in perspective," Br. J. Polit. Sci. **39**(02), 413–448 (2009).
[35]J. A. Hołyst, K. Kacperski, and F. Schweitzer, "Social impact models of opinion dynamics," in *Annual Reviews of Computational Physics* (World Scientific, 2001), Vol. 9, pp. 253–273.
[36]C. I. Hovland, O. Harvey, and M. Sherif, "Assimilation and contrast effects in reactions to communication and attitude change," J. Abnorm. Soc. Psychol. **55**(2), 244 (1957).
[37]I. H. Indridason, "Coalition formation and polarisation," Eur. J. Polit. Res. **50**, 689–718 (2011).
[38]S. Iyengar and S. J. Westwood, "Fear and loathing across party lines: New evidence on group polarization," Am. J. Pol. Sci. **59**(3), 690–707 (2015).
[39]N. Kaldor, "Capital accumulation and economic growth," in *The Theory of Capital* (Springer, 1961), pp. 177–222.
[40]H. Kim, G. B. Powell, Jr., and R. C. Fording, "Electoral systems, party systems, and ideological representation: An analysis of distortion in Western democracies," Comp. Polit. **42**(2), 167–185 (2010).
[41]E. Kurmyshev, H. A. Juarez, and R. A. Gonzalez-Silva, "Dynamics of bounded confidence opinion in heterogeneous social networks: Concord against partial antagonism," Physica A **390**(16), 2945–2955 (2011).
[42]D. Lacy and P. Paolino, "Testing proximity versus directional voting using experiments," Elect. Stud. **29**(3), 460–471 (2010).
[43]P. Leimgruber, D. Hangartner, and L. Leemann, "Comparing candidates and citizens in the ideological space," Swiss Polit. Sci. Rev. **16**(3), 499–531 (2010).
[44]J. Lorenz, "Continuous opinion dynamics under bounded confidence: A survey," Int. J. Mod. Phys. C **18**(12), 1819–1838 (2007).
[45]A. Malka and Y. Lelkes, "More than ideology: Conservative–liberal identity and receptivity to political cues," Soc. Justice Res. **23**(2–3), 156–188 (2010).
[46]L. Mason, "The rise of uncivil agreement: Issue versus behavioral polarization in the American electorate," Am. Behav. Sci. **57**(1), 140–159 (2013).
[47]L. Mason, "Distinguishing the polarizing effects of ideology as identity, issue positions, and issue-based identity," in paper presented at the Center for the Study of Democratic Politics Conference on Political Polarization: Media and Communication Influences, Princeton University, Princeton, NJ, 2015.
[48]X. F. Meng, R. A. Van Gorder, and M. A. Porter, "Opinion formation and distribution in a bounded-confidence model on various networks," Phys. Rev. E **97**(2), 022312 (2018).
[49]S. Merrill and B. Grofman, *A Unified Theory of Voting: Directional and Proximity Spatial Models* (Cambridge University Press, 1999).

[50] B. Min and M. San Miguel, "Fragmentation transitions in a coevolving nonlinear voter model," Sci. Rep. **7**(1), 1 (2017).
[51] E. Pariser, *The Filter Bubble: What the Internet Is Hiding from You* (Penguin, UK, 2011).
[52] N. Perony, R. Pfitzner, I. Scholtes, C. J. Tessone, and F. Schweitzer, "Enhancing consensus under opinion bias by means of hierarchical decision making," Adv. Complex Syst. **16**(06), 1350020 (2013).
[53] M. Pineda, R. Toral, and E. Hernández-García, "Diffusing opinions in bounded confidence processes," Eur. Phys. J. D **62**(1), 109–117 (2011).
[54] K. T. Poole, *Spatial Models of Parliamentary Voting* (Cambridge University Press, 2005).
[55] G. Rabinowitz and S. E. Macdonald, "A directional theory of issue voting," Am. Polit. Sci. Rev. **83**(1), 93–121 (1989).
[56] R. C. Schank and E. Langer, *Beliefs, Reasoning, and Decision Making: Psycho-logic in Honor of Bob Abelson* (Psychology Press, 2013).
[57] S. Schweighofer, F. Schweitzer, and D. Garcia, "A weighted balance model of opinion hyperpolarization," J. Artif. Soc. Soc. Simul. (2020); SocArXiv.
[58] F. Schweitzer, *Brownian Agents and Active Particles: Collective Dynamics in the Natural and Social Sciences* (Springer Science & Business Media, 2007).
[59] F. Schweitzer, "Sociophysics," Phys. Today **71**(2), 40–46 (2018).
[60] F. Schweitzer, "The law of proportionate growth and its siblings: Applications in agent-based modeling of socio-economic systems," in *Complexity, Heterogeneity, and the Methods of Statistical Physics in Economics*, edited by H. Aoyama, Y. Aruka, and H. Yoshikawa (Springer, Tokyo, 2020), pp. 145–176.
[61] F. Schweitzer and L. Behera, "Nonlinear voter models: The transition from invasion to coexistence," Eur. Phys. J. B **67**(3), 301–318 (2009).
[62] F. Schweitzer, T. Krivachy, and D. Garcia, "An agent-based model of opinion polarization driven by emotions," Complexity **2020**, 5282035.
[63] A. Sirbu, V. Loreto, V. D. P. Servedio, and F. Tria, "Opinion dynamics with disagreement and modulated information," J. Stat. Phys. **151**(1–2), 218–237 (2013).
[64] H.-U. Stark, C. J. Tessone, and F. Schweitzer, "Decelerating microdynamics can accelerate macrodynamics in the voter model," Phys. Rev. Lett. **101**(1), 018701 (2008).
[65] E. Suhay, "Explaining group influence: The role of identity and emotion in political conformity and polarization," Polit. Behav. **37**(1), 221–251 (2015).
[66] K. Takács, A. Flache, and M. Mäs, "Discrepancy and disliking do not induce negative opinion shifts," PLoS One **11**(6), e0157948 (2016); available at https://doi.org/10.1371/journal.pone.0157948.
[67] G. Vaccario, M. V. Tomasello, C. J. Tessone, and F. Schweitzer, "Quantifying knowledge exchange in R&D networks: A data-driven model," J. Evol. Econ. **28**(3), 461–493 (2018).
[68] F. Vazquez, P. L. Krapivsky, and S. Redner, "Freezing and slow evolution in a constrained opinion dynamics model," J. Phys. A **36**, L61–L68 (2003), Ms vorh.
[69] G. Weisbuch, "Bounded confidence and social networks," Eur. Phys. J. B **38**(2), 339–343 (2004).
[70] A. M. Wilford, "Polarization, number of parties, and voter turnout: Explaining turnout in 26 OECD countries," Soc. Sci. Q. **98**(5), 1391–1405 (2017).