Amsterdam University Press

COMPUTATIONAL COMMUNICATION RESEARCH 6.1 (2024) 1–40
HTTPS://DOI.ORG/10.5117/CCR2024.1.3.POGS

Agent-based modeling of diversity, new information and minority groups in opinion formation

Mark Pogson
Department of Communication and Media, University of Liverpool

Tom Nicholls
Department of Communication and Media, University of Liverpool

Abstract
Understanding the decentralized formation of public opinion is increasingly important to communication research. Although many key determinants cannot be observed empirically, we argue they can be explored through theoretical modeling. Building on an existing agent-based model of opinion dynamics, our study introduces more complex, but theoretically interesting and realistic, agent behavior. We model distinct opinion tendencies which represent individuals' diversity of belief, as well as external influences such as new information. Diversity increases the extremity of opinion in simulated consensus, radicalization and polarization. Simulation of new information demonstrates the ability of a minority group to shift majority opinion significantly in the long term, even with transient changes in behavior. Opposing minority groups do not counteract each other when their actions are delayed and may in fact amplify the original effect. We argue that modeling can help researchers and other stakeholders understand how these outcomes could arise in the real world, and thereby explore potential mitigations or exploitations.

Keywords: opinion dynamics, agent-based model, public opinion, polarization, radicalization, echo chambers, social network, social media, social influence, exogenous effects, opinion tendency, minority group, attitude, belief, identity

Introduction

Social media enables public opinion to develop and spread at scale through largely self-organizing networks (Neubaum and Krämer 2017). It therefore provides opportunities to strengthen and democratize public discourse, but also has the potential to spread harmful information with minimal safeguards or scrutiny (Gayo-Avello 2015; Wang et al. 2020; Lorenz-Spreen et al. 2020). Given the increasing significance of social media to society, and the apparent shortcomings in its regulation, it is important to understand how it can shape public opinion (Napoli 2015; Hitkul et al. 2021; Del Vicario et al. 2016). An important aspect of this is how external events can both affect and be affected by social media activity, as demonstrated by examples such as the Covid-19 infodemic, the January 6th American Capitol attack, and the collapse of Silicon Valley Bank (Cinelli et al. 2020; Suresh et al. 2023; Yerushalmy 2023).

Agent-based models (ABMs) provide a way to bridge the gap between data and theory in communication research, and have particular potential for computational communication researchers to understand the dynamics of public opinion (Waldherr and Wettstein 2019; Waldherr et al. 2021). ABMs explicitly represent individual social actors and their communications, thereby offering a tractable and realistic representation of how public opinion forms (Banisch, Lima, and Araújo 2012). Temporal, spatial and stochastic effects can all be incorporated in ABMs, and emergent phenomena result from the collection of individual behaviors, thus providing a natural transition from the micro- to macro-scale (Bonabeau 2002). ABMs can offer a valuable complement to surveys and text analysis to understand and predict opinion dynamics, taking a system-level view of input-output relationships based around causality, rather than simply a correlational one offered by statistics and machine learning (Baker et al. 2018; Noorazar 2020; Acemoglu and Ozdaglar 2011; Das, Gollapudi, and Munagala 2014).

Opinion dynamics models originate from the field of social physics, where ontological parsimony is prioritized (Jusup et al. 2022). Such models do not seek to represent all aspects of the system, meaning that important real-world details may be simplified or excluded (Anderson and Ye 2019). The aim is to enable more detailed forms of analysis and obtain insights into the system by isolating individual factors in a way which is hard to achieve empirically. For these reasons, opinion dynamics models have been adopted by researchers outside social physics, including in communication and behavioral sciences, to complement both traditional and big-data approaches in the study of public opinion (Song and Boomgaarden 2017).

The term 'public opinion' is used in the present study to mean an aggregate of separately held opinions, as commonly obtained using survey methods, although we are interested in finer-grained details than are generally possible empirically, especially at the temporal level, but also in terms of individual social influence. A range of alternative interpretations of its meaning exist, include polling outcomes, and discursive consensus in public spaces (Herbst 1993). However, we interpret public opinion as a continual and diverse process, rather than a single fixed outcome, hence we do not restrict our interpretation to be the majority opinion expressed in polls (Crespi 1997). Furthermore, while our focus is on how public discourse gives rise to aggregated opinion, we recognize there are other factors involved, including belief and communication beyond a particular public sphere. We also recognize there is potential for multiple meaningful aggregates of opinion, hence we do not restrict our interpretation to that of consensus.

Echo chambers are popularly associated with social media, and have been argued to arise where users interact primarily with others who share similar opinions to their own (Jasny, Waggle, and Fisher 2015; Cinelli et al. 2021). Some authors have claimed echo chambers are central to opinion polarization, but their existence has also been found to be overstated in many contexts and their connection to polarization less straightforward than often assumed (Dubois and Blank 2018; Fletcher, Kalogeropoulos, and Nielsen 2021; Bail et al. 2018). Radicalization is related to opinion polarization, but referring to a single pole (McCauley and Moskalenko 2017; Della Porta 2018). Although both polarization and radicalization long predate social media, the scale and algorithmic influence of social media platforms make them particularly salient today (Törnberg 2022; Kubin and von Sikorski 2021; Geiß et al. 2021).

Due to the multiple sources of complexity in opinion dynamics, opinions tend to be modeled in a 1-dimensional space, where each opinion can be represented either as a discrete choice, or a scalar value on a continuum, such as for the left-right political spectrum. The dissemination of 1-dimensional opinions can be inferred from social media data by observing links to different information sources in users' posts and analyzing the network of interactions based around these (Cinelli et al. 2021).

Many contentious and divisive political issues of recent times have been accompanied by extreme online opinions, including Brexit in the UK, and support for President Trump in the US (Del Vicario et al. 2017; Guo et al. 2020; Swol, Lee, and Hutchins 2022). When a single issue dominates current affairs, the assumption of a 1-dimensional opinion space becomes increasingly suitable, hence opinion dynamics models have an effective role to play in the study of extreme opinions, despite their simplified nature (Podobnik et al. 2017). This is perhaps most significant in cases of military conflict, where social media is increasingly influential, including the use of bots as influencing agents (Prier 2017; Ali and Fahmy 2013; Sacco and Bossio 2015).

In decentralized networks, as are common with social media, opinion dynamics tend to be driven either by a likeminded critical mass, or a highly confident minority (Baran 1964; Moussaïd et al. 2013). There are several factors which affect how opinions spread, including cultural, psychological and linguistic (Mckeown and Sheehy 2006; Centola and Baronchelli 2015; Wolf, Doorn, and Weissing 2008; Neubaum and Krämer 2017). There is extensive evidence from public opinion survey research that personal characteristics such as gender are important in opinion formation, driving tendencies and thereby affecting public opinion (Herek 2002; Kite, Whitley, and Wagner 2023).

ABMs typically focus on homophily and consensus-seeking based on current opinions, termed bounded-confidence models, but consideration has also been given to the role of memory, disagreement, and multidimensional opinions (Becchetti et al. 2023; Gargiulo and Gandica 2017; Minh Pham et al. 2020; Kozitsin 2023; Flache and Macy 2011; Parsegov et al. 2016).

Baumann et al. (2020) published an ABM of opinion dynamics where the emergence of consensus, radicalization and polarization was found to be governed by two main parameters, namely the controversy of the subject being discussed, and the level of homophily in the network. These parameters have recognizable real-world meanings. Controversy is treated abstractly and would manifest in culturally dependent ways, but it typically relates to the strength of feeling people have on a given topic, and the scope for disagreement. Homophily reflects both the desire and ability for individuals to engage with others they agree with, combining social, physical, technological and personal factors.

The model of Baumann et al. (2020) differed from previous bounded-confidence models of continuous opinion dynamics by allowing multiple social influences to occur simultaneously, unlike models derived from the influential model of Deffuant et al. (2000). It focused on rates of change in opinion rather than the opinions themselves, unlike models derived from another influential bounded-confidence model of Hegselmann and Krause (2002). Homophily was modeled probabilistically based on opinion differences relative to others in the network, rather than using a fixed threshold common to standard bounded-confidence models. Most usefully for the present study, the model formulation distinguished individual and social influences clearly, thus making it suitable for further development to explore these competing effects, and so reflect the theoretical importance of selective exposure (Song and Boomgaarden 2017; Stroud 2010). An unstructured network was assumed, where node connections were updated in line with current opinions. This approach reflects the fluid nature of many online networks, where platform algorithms are particularly instrumental, and connections may not even be directly between individuals but, rather, ephemeral content they produce and consume. This is unlike offline social networks, which are typically based on more meaningful and less changeable interpersonal relationships. Weak social ties have been observed to be particularly important to extreme opinions, so are a suitable focus for study (Fan, Xu, and Zhao 2020). The number of connections for nodes in the fluid network defined by Baumann et al. (2020) has statistical properties closer to real-world data than is represented in many opinion dynamics models based on static networks (Moinet, Starnini, and Pastor-Satorras 2015).

Baumann et al. (2020) validated the model against Twitter data and found it to be in good agreement. Their study considered long-running topics, such as gun control and abortion, in the absence of exogenous effects, i.e. influences from outside the social network. The only modeled sources of diversity were the initial opinions of agents, their message rates, and stochasticity. In the absence of social interaction, all agents were assumed to tend towards the same central opinion. Links to news sources were effectively able to be shared between agents by the messages they sent, but the effect of developing news was not considered. By ignoring exogenous effects, the model does not reflect real-time external influences, which are known to be important to real-world opinion dynamics (Page, Shapiro, and Dempsey 1987).

Exogenous influences on opinion dynamics have previously been investigated with binary issues using ABMs, including media environments in a Swiss referendum, and belief in climate change based on levels of internet access in the US (Wettstein 2020; Sikder et al. 2020). For continuous opinions, exogenous influences have been explored by extending the model of Hegselmann and Krause (2002) to treat external events similarly to other agents' opinions, finding that opinion polarization relied on these external events (Condie and Condie 2021). However, more recent work also built on Hegselmann and Krause (2002) to find that external events were not required to generate polarization, and could instead be produced by differences between privately and publicly held opinions (Lim and Bentley 2022).

To model exogenous effects, we use the work of Baumann et al. (2020) as a starting point due to its clear distinction between individual and social influences, and its ability to simulate observed real-world behaviors, such as polarization, independently of any external influences. Our updated model represents real-time exogenous influences in a way which is more closely coupled to continuous opinion dynamics than for the snapshot binary topics considered by Wettstein (2020) and Sikder et al. (2020), and it avoids the contested assumptions required to generate polarization in models derived from Hegselmann and Krause (2002). Our model therefore provides a plausible representation of social and exogenous influences on opinion dynamics, and enables meaningful examination of how micro-scale behaviors could produce macro-scale effects on public opinion under realistic scenarios. We have tested the wider generalizability of our model by using a sensitivity analysis, included in the Online Appendix[1].

We hope that the present study is useful both in advancing the state of the art of opinion dynamic modeling, and also in highlighting the opportunities that mathematical modeling can offer computational communication researchers to examine and triangulate the assumptions underpinning their empirical models. We further hope the model provides useful insights into the effects of decentralized communication on public opinion by extending existing knowledge and complementing other research approaches, which may benefit stakeholders beyond communication research.

Background

Modeling overview

Models of multi-agent systems which explicitly represent individual components and their interactions are termed ABMs. These typically involve discrete time steps to simulate agent behaviors, and can include a range of mathematical and computational techniques. ABMs are often based around discrete state transitions of agents, such as with Markov chains, but we use ordinary differential equations (ODEs) to describe changes in the continuous opinions of agents. This is a form of dynamical system modeling which takes a bottom-up rather than top-down approach. ODEs are differential equations with a single independent variable, in this case time. The opinion of each agent is represented by a single ODE, and the ODEs are coupled, i.e. interdependent. The approach has similarities to vector autoregressive modeling, but without relying on statistical fitting.

Baumann et al. (2020) model

To provide context, the model of Baumann et al. (2020) is summarized in Equation 1 - Equation 3. A social actor in the model is described as an agent, identified by index i. The opinion of agent i is represented as x_i, which is a function of time t. The model is simulated using discrete time steps in which agents can send and receive messages representing their current opinions, and update their opinions accordingly. The social network in the model is defined by the exchange of messages, rather than agents holding fixed connections. The model imposes a limit of m messages that can be received by each agent during a time step. Most individuals in a typical social network have relatively few connections, while a small number are highly connected, as determined by preferential attachment (Dorogovtsev 2010). This property holds at all scales, and it is thus termed a scale-free network. These are well approximated by a power law distribution for agent node connections, represented by the message rate in the fluid network of Baumann et al. (2020), i.e. the probability of an agent being able to send a message at each time step (Moinet, Starnini, and Pastor-Satorras 2015). The power law probability density function F for agent message rate a is:

$$F(a) = \frac{1-\gamma}{1-\varepsilon^{1-\gamma}} a^{1-\gamma} \tag{1}$$

where γ is a parameter which governs the shape of the power-law distribution, and ε is the minimum permitted message rate, i.e. the lowest possible message rate for an agent to be considered part of the network, preventing an abundance of uncommunicative agents. Prior to simulation, each agent is assigned a fixed message rate according to Equation 1. This is a convenient way to generate scale-free network behavior, while allowing agents to choose their own connections based on homophily.

Messages are sent probabilistically in the model according to agents' message rates. At each time step in the simulation, the opinion of agent i changes at a rate $\dot{x}_i$, balancing the agent's tendency towards 0 opinion, and the influences it receives from other agents j:

$$\dot{x}_i = -x_i + K \sum_{j \neq i}^{N} A_{ij} \tanh(\alpha x_j) \qquad (2)$$

where N is the number of agents in the network, K is a sociality parameter representing the level of attention agents give to other agents' opinions, A_{ij} is the adjacency matrix value for agents i and j (which is time-varying, described further below), and α is the level of controversy for the topic being communicated.

The tanh function in Equation 2 is the hyperbolic tangent, often termed a sigmoid or logistic curve. Its role is to moderate the opinion of other agents, as observed experimentally (Jayles et al. 2017). The nature of tanh means all influencing opinions are constrained to be between ±1, before being scaled by K. The value of α gives the gradient of the function at the origin, so $\alpha > 1$ amplifies opinions close to 0, thus making transitions sharper between opposing opinions.

The first term in Equation 2 moves the opinion of agent i towards 0, i.e. a common central tendency for all agents. The social summation term aggregates the moderated opinion of influencing agents at the current time step.

Parameter K effectively includes a factor of $\frac{1}{m}$, hence the summation provides an average of other agents' opinions, similar to DeGroot learning (DeGroot 1974). However, it differs in that the summation contributes to the rate of change in opinion of agent i, rather than giving the new opinion itself. Furthermore, the adjacency matrix value A_{ij} is either 1 or 0, depending on whether agent i receives a message from agent j, unlike the matrix used by DeGroot (1974) which weights opinions according to levels of trust.

The moderating effect of the tanh function assumes there is a common central reference point of 0 for all agents, which in this case coincides with the common central tendency. It is therefore only social influence which moves agents away from 0 opinion, meaning the model allows an agent to be radicalized by any opinion of the same sign as the one it already holds. This is considered further in the Method. The adjacency matrix describes connections between agents and is determined according to homophily. Agents are more likely to communicate if they share similar opinions, and ignore agents they have less in common with, in accordance with selective exposure. The probability of agent i sending a message to agent j is calculated by:

$$p_{ij} = \frac{|x_i - x_j|^{-\beta}}{\sum_{j \neq i}^{N} |x_i - x_j|^{-\beta}} \tag{3}$$

where β is a parameter which represents the strength of homophily in the network. If $|x_i - x_j|$ is small, it will tend to dominate the summation in the denominator, hence p_{ij} will approach 1, and if β is large, the effect will be increased. Equation 3 is calculated for each agent at each time step in the simulation (noting i and j are reversed for use in Equation 2), and the m most probable connections are chosen for each agent, from which the adjacency matrix is obtained. There is a probability r of agents sending reciprocal messages, which is unconstrained by the limit m.

Baumann et al. (2020) parameterized the original model to explore the dependence of public opinion on social network behavior, using a sensitivity analysis to explain parameter value choices. They found that more active agents had more extreme opinions, and that agent opinions correlated with those of agents they communicated with. Both these findings matched well with social media data, as well as empirical evidence for the effect of selective news exposure in online social networks (Bakshy, Messing, and Adamic 2015).

Research aims

The Baumann et al. (2020) model successfully produces different qualitative outcomes in opinion dynamics, specifically for consensus, radicalization and polarization. It also matches well with social media data, particularly for homophilous opinions.

However, the model assumes a common 0 tendency for all agents, i.e. the opinion agents would reach in the absence of social interaction. As social scientists we are interested in understanding the consequences of a model which better reflects heterogenous opinion tendencies, i.e. the observation that humans' opinions are intrinsically diverse rather than tending towards a single common norm. Enabling these tendencies to change over time would also allow new information from outside the network to be represented in the model.

Using the new model functionality for agent tendencies, we first aim to investigate the effects of diversity on opinion dynamics:

RQ1 Can heterogeneous opinion tendencies change global opinion in consensus, radicalization and polarization?

We then aim to investigate the effect of real-time news on opinion dynamics, focusing on minority groups, which may have different information sources, beliefs or agendas to the majority. Using an abrupt temporary shift in tendencies to represent sudden events, and a gradual shift in tendencies to represent slow-building stories, we will investigate how minority groups can influence majority opinion:

RQ2 Can different changes in the tendencies of minority groups have different effects on majority opinion?

Method

Opinion tendencies

An explicit opinion tendency $\hat{x}_i$ is introduced for each agent, which replaces the implied 0 tendency in Equation 2. The tendency can either be constant or a function of time, giving an updated model form of:

$$\dot{x}_i = \hat{x}_i - x_i + K \sum_{j \neq i}^{N} A_{ij} \tanh(\alpha x_j) \tag{4}$$

Tendency brings interesting and realistic possibilities to the model, including scope to represent prejudice, bias and belief, as well as the ability for agents to change these in light of new information, experience and other external influences. This represents the assumed distinction of public opinion and public discourse, but also their interdependence.

Although the introduction of $\hat{x}_i$ removes the assumption of a common 0 central tendency, there is still an implicit assumption of a common 0 central reference point in how the tanh function is used, but there is no special meaning to this value. If an arbitrary central reference point of μ were assumed, the tanh function would be applied to $x_i - \mu$ instead of x_i. However, since it has no bearing on the opinion dynamics being considered, we omit this without loss of generality.

As with Baumann et al. (2020), the updated model assumes agent connections are determined by current agent opinions, rather than long-term relationships. The extent to which agents prioritize communication with like-minded individuals is controlled by homophily β, the level of attention they pay to the messages they receive is controlled by sociality K, and the relative effect of non-zero messages is controlled by controversy α. It would be possible to estimate parameter values from real-world data, but this is beyond the scope of the present study. We outline a framework for harnessing empirical data in the Discussion.

Computation

We first reproduced the model of Baumann et al. (2020) using original Python code, before obtaining new results from the updated model[2]. Baumann et al. (2020) used an explicit fourth-order Runge-Kutta method to integrate the system of equations represented by Equation 2 numerically, using a time step of 0.01 time units. We instead used a first-order Euler method to make opinion changes more clearly linked to incoming messages, which produced very similar results. The number of agents used by Baumann et al. (2020) was 1000, but the present study used $N = 500$ to give shorter run times. In practice, any sufficiently large N produced similar results. Simulations were run for 1000 time steps, equivalent to 10 time units in Baumann et al. (2020).

To match Baumann et al. (2020), initial opinions of agents were randomly sampled from a uniform distribution between -1 and 1. Message rate distribution parameter values were $\gamma = 2.1$ and $\varepsilon = 0.01$, with message limit $m = 10$, sociality parameter $K = 3$, and reciprocal message probability $r = 0.5$. Messaging at each time step was evaluated in randomized order. The adjacency matrix was calculated probabilistically at each time step based on Equation 3.

Baumann et al. (2020) presented results for (a) consensus towards the initial central opinion, generated by low controversy and moderate homophily ($\alpha = 0.05$, $\beta = 2$); (b) radicalization (i.e. collective movement away from the initial central opinion in a single direction), generated by high controversy and no homophily ($\alpha = 3$, $\beta = 0$); and (c) polarization (i.e. an equal split away from the initial central opinion in opposing directions), generated by high controversy and high homophily ($\alpha = \beta = 3$).

All opinion tendencies were first set as 0 to enable direct comparison of results with Baumann et al. (2020). The direction of radicalization is random in the model, but we present results with the radicalized opinion as positive to make interpretation simpler. This was achieved by reversing the sign of opinion dynamics if the mean final opinion was negative in the radicalization simulations. Note that the term 'radicalization' is used for convenience, but may be misplaced in terms of social outcome, since it simply represents a shift in the average opinion of the social network, as occurs with all kinds of sociopolitical change. No position on the scale is modelled as inherently better or worse than any other.

Model simulation and analysis were performed using NumPy, SciPy, random, pandas and Joblib libraries in Python (Harris et al. 2020; Virtanen et al. 2020; van Rossum 2020; McKinney 2010; Joblib 2023). Plotting was performed using matplotlib and seaborn libraries in Python (Hunter 2007; Waskom 2021).

Model experiments

After comparison with Baumann et al. (2020), the updated model was used to address the RQs. Unless otherwise specified, parameter values were the same as stated above. The term 'group' is used as shorthand for agents which shared particular properties, but communication with all other agents was unaffected, i.e. there was no concept of insider or outsider status. This enabled anonymous opinion-based communication to be directly simulated – something hard to observe empirically, but which social media quite uniquely enables – where, for example, experts, conspiracy theorists and bots can appear equally credible (Luceri et al. 2019).

For RQ1, random constant tendencies of ±0.5 were used. These were chosen to be inside the range of initial opinions, and to average as 0 (see Online Appendix for a sensitivity analysis). They could represent differences in intrinsic personal beliefs, or the time-invariant effect of influences outside the network, such as an individual's habitual newspaper choice. Agents were randomly assigned −0.5 or +0.5 constant tendency prior to simulation, with equal probability. When making radicalization results positive (as described above), the tendency groups were correspondingly reversed.

For RQ2, time-dependent tendencies were applied to a minority group, while all other agents had 0 constant tendency. Parameter values for high controversy and high homophily were used ($\alpha = \beta = 3$), which would cause symmetrical polarization in the absence of the minority group. Polarizing conditions were chosen to explore the effects of minority groups due to their well-documented prevalence and importance to extreme opinion (Heltzel and Laurin 2020). A square pulse function was used to provide an abrupt temporary shift in minority group tendencies, for example news triggering a bank run, while a linear ramp function was used to provide a gradual change in minority group tendencies, such as mounting evidence for a smoking ban. All other agents were unaffected by these time-dependent tendencies, other than through influences they received directly or indirectly from the minority group. This was performed first with a single minority group, and then with an opposing minority group of the same size, but with a delay in the tendency function.

We constructed a minority group to hold tendencies significantly outside the initial opinion range, using a tendency amplitude of 2. We wanted the group to be small enough to plausibly represent a fringe opinion, but large enough to have an influence on the broader population, so we used a minority group size of 5% of all agents. The pulse had an onset time of 0, and a duration of 500 time steps, chosen to affect the initial stages of opinion dynamics and finish well before the end of the simulation. The ramp reached a tendency amplitude of 2 after 1000 time steps, chosen to provide the same area under the curve as the transient pulse, and last for the full simulation. The opposing pulse (i.e. −2 amplitude) had a time delay of 100 time steps, chosen to provide rapid but not immediate opposition to the original pulse. In practice, the results do not substantively depend on these choices (see the Online Appendix for a sensitivity analysis).

Each model configuration is termed a trial. Results for each trial are presented from a single simulation, but to account for stochasticity, repeated simulations (termed repetitions) were run, and the aggregated results are also shown. Trials and parameter values are summarized in Table 1 and Table 2. The effect of aggregating across repetitions is considered in the Discussion.

For each trial, the outcome of interest was agent opinions, i.e. modeled public opinion. The dynamics of these were plotted with related properties such as the agent group. For statistical comparison of different trials and groups, the final opinion of agents was used, i.e. their opinions at the end of the simulation.

To answer the RQs, distributions of final opinions were compared between trials by using a one-way analysis of variance (ANOVA) F-test, and a Kruskal-Wallis (K-W) H-test. The ANOVA p value gives the probability of means being the same between distributions, assuming normality, and the K-W p value gives the probability of medians being the same, without assuming normality (Johnson 2022). To provide more meaningful statistical comparisons for RQ1 with consensus and polarization, opinion magnitudes were

Trial	α	β	T	Description
A	0.05	2	0	Consensus
B	3	0	0	Radicalization
C	3	3	0	Polarization
D	0.05	2	±0.5	Trial A variant: opposing constant groups
E	3	0	±0.5	Trial B variant: opposing constant groups
F	3	3	±0.5	Trial C variant: opposing constant groups
G	3	3	0	Trial C variant: single pulse minority group
H	3	3	0	Trial C variant: single ramp minority group
I	3	3	0	Trial C variant: opposing pulse minority groups

Table 1: Model trials and baseline parameter values. For all trials, $K = 3$, $m = 10$, $r = 0.5$, $\gamma = 2.1$ and $\varepsilon = 0.01$. Parameter T is the constant tendency magnitude for the majority group, i.e. $|\hat{x}_i|$ in Equation 4. Baseline minority group parameter values are given in Table 2.

Parameter description	Parameter value
Minority group size (fraction)	0.05
Pulse amplitude	2
Pulse time duration	500
Pulse onset time	0
Opposing pulse time delay (Trials G–H)	N/A
Opposing pulse time delay (Trial I)	100

Table 2: Baseline parameter values for minority groups. Minority group size is expressed as a fraction of all agents, where 0.002 is 1 agent in 500.

used for reasons of symmetry.

Results were aggregated over 32 repetitions of each trial for use in statistical comparisons. They were also aggregated in plots of opinion dynamics with a bootstrapped estimate of the mean and 95% confidence bands for each tendency group.

Trials and analyses are summarized as follows:

- Figure 1: Opinion dynamics for consensus, radicalization and polarization, with 0 tendency for all agents (Trials A–C).
- Figure 2: Opinion dynamics with heterogeneous constant opinion

tendencies of ±0.5, using conditions for consensus, radicalization and polarization (Trials D–F).

- Figure 3: Comparison of final global opinion between Trials A–C and Trials D–F (RQ1).
- Figure 4: Opinion dynamics with time-dependent minority group tendencies using square pulse and linear ramp functions (Trials G–H).
- Figure 5: Opinion dynamics with minority group tendencies as for Trial G, but with an equally sized delayed opposing group (Trial I).
- Figure 6: Comparison of final majority opinion between Trial C and Trials G–I, and between Trial G and Trial I (RQ2).
- Figure 7: Opinion dynamics aggregated by tendency and initial opinion group for Trials G–I.

Results

Comparison with Baumann et al. (2020)

Results are shown in Figure 1 for consensus, radicalization and polarization (Trials A–C), which closely match results presented by Baumann et al. (2020). In Figures 1a - 1c, results are colored by initial opinion, showing that initial opinion strongly determines the polarization of agents (Figure 1c), i.e. agents with a positive initial opinion will likely be polarized in the positive direction, and vice versa for agents with a negative initial opinion. Conversely, agents in both consensus and radicalization are relatively unaffected by their initial opinion (Figures 1a - 1b).

In Figure 1d - 1f, results are colored by the number of messages received by each agent. This number depends on the number of homophilous agents present throughout the simulated time period, as well as the message rate of those agents, the message limit m, and stochastic effects. It is evident that agents which received the most messages generally had the most extreme opinions. As all agents had an opinion tendency of 0 in Trials A–C, it was only social influence which moved them away from 0 opinion, so it makes sense that the most influenced agents tended to hold the most extreme opinions. There is empirical evidence to support this, but it is also highly dependent on other variables such as overall engagement, political context, and trust (Amit, Mannan, and Islam 2020; Lorenz-Spreen et al. 2023).

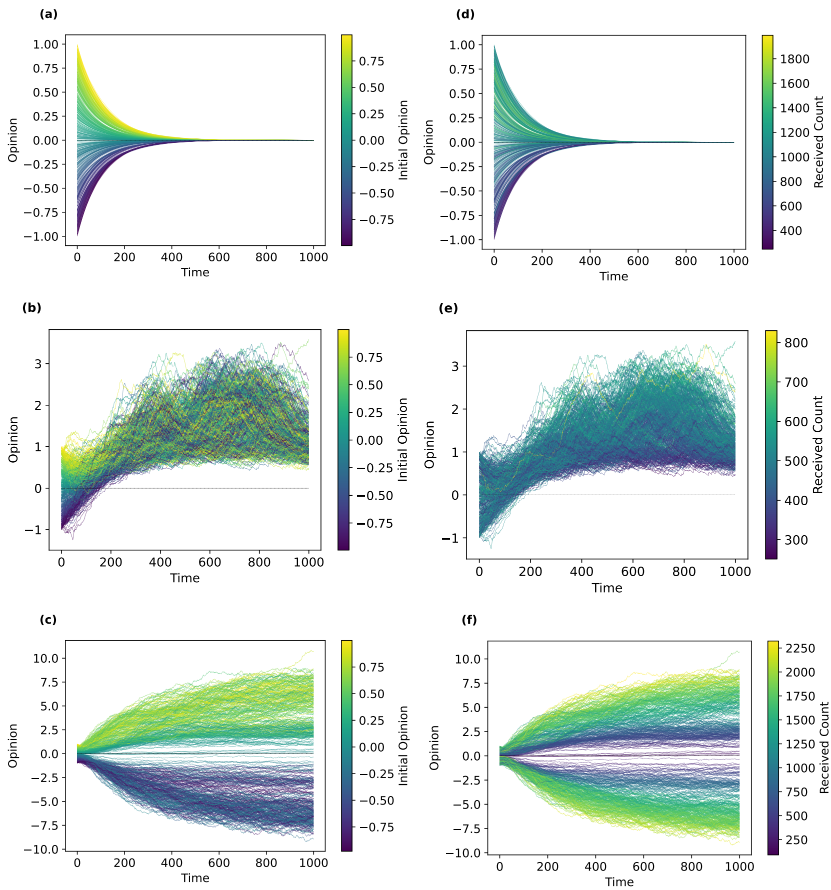

Figure 1: Opinion dynamics for (a) consensus, (b) radicalization and (c) polarization (Trials A–C) from a reimplementation of the original model of Baumann et al. (2020). In plots (a-c), lines are colored according to the initial opinion of agents, where polarization is strongly determined by initial opinion. Plots (d-f) show equivalent results colored by the number of messages received by agents, where the most extreme agents tend to have received the most messages. Results in each plot are from a single simulation, matched between plots (a-c) and (d-f).

Heterogeneous constant tendencies

Results for random constant tendencies of ±0.5 are shown in Figure 2, including aggregated results from multiple simulations to account for stochastic effects. The aggregation distinguishes the tendency groups of agents.

In Figure 2a, agents reached separate group consensuses ('split consensus') which corresponded to their respective tendencies, producing fragmentation of opinion instead of a single central consensus, i.e. distinct aggregates of public opinion. There was a greater spread of final opinions within each consensus group than before, due to the smaller subpopulations having a slower process of consensus-reaching and a correspondingly weaker overall cohesion. Figure 2d shows that consensuses were slightly more extreme than the tendencies themselves, due to agents which had initial opinions beyond their tendencies pulling others towards them. Otherwise, behavior was similar to Figure 1a for a single central consensus. Although the effective polarization in Figure 2a appears in some ways to be greater than in Figure 1c, this arises from the tendencies alone; note also the difference both in opinion magnitudes and underlying mechanisms. This is important when considering polarization more generally, as behaviors which reflect existing opinion variations are not in themselves evidence for a growth in polarization through social effects.

For radicalization in Figure 2b, it is evident that agents were able to be radicalized against their own tendency ('divergent radicalization'), given a sufficiently large sociality value . The two tendency groups were distinct in their radicalization, reflecting the respective tendencies of agents, i.e. agents with +0.5 tendency had a more positive final opinion than those with −0.5 tendency, and there was little mixing between the two groups. Figure 2e shows that the two tendency groups initially began to diverge towards their respective tendencies, before following a single radicalization direction, while maintaining the opinion gap from divergence. Interestingly, the initial divergence was not caused by the echo-chamber effect, since without homophily, all agents were equally likely to communicate with each other, so divergence was driven only by tendency. As noted above for Figure 2a, this mechanism differs to social polarization, although the outcome looks similar.

Many agents in Figure 2c were polarized against their own tendency, similar to the effect found for radicalization in Figure 2b. There was more intermixing between the two tendency groups than for the radicalization in Figure 2b,

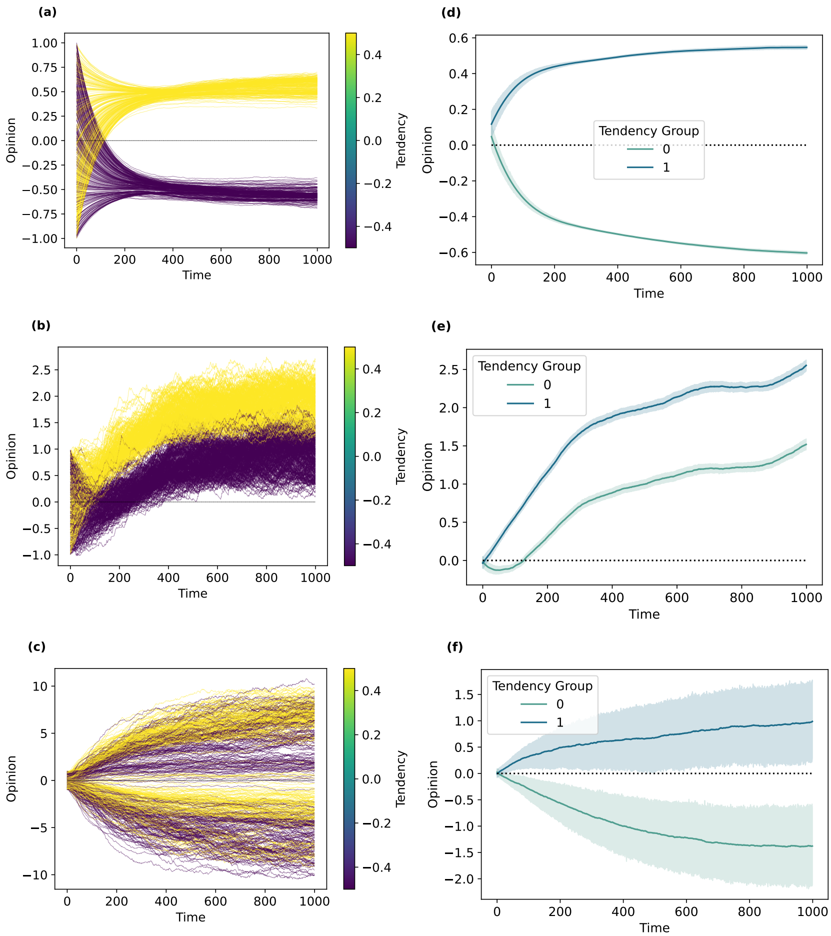

Figure 2: Opinion dynamics with random constant tendencies of ±0.5 for (a) consensus, (b) radicalization and (c) polarization (Trials D–F). Results in each plot (a-c) are from a single simulation, with lines colored according to the opinion tendency of agents. Plots (d-f) show corresponding aggregated results. Lines are colored according to the tendency groups (group 0 has -0.5 tendency, and group 1 has $+0.5$ tendency).

but agents with +0.5 tendency were more likely to be polarized positively, and also to be at the higher end of whichever polarization group they became part of, and vice versa for agents with −0.5 tendency ('biased polarization'). There were conflicts between tendencies and social influences for many agents, which did not occur in the same way when all agents started from a neutral position, where Figure 1c showed that initial opinion strongly determined polarization. However, Figure 2f shows that tendency tended to override initial opinion in determining the eventual polarization of agents.

RQ1

To answer RQ1, the final global opinion of agents was compared between Trials A–C and Trials D–F, as shown in Figure 3. Note the difference between tendency group, as shown in Figure 2f, and polarization group, as effectively shown in Figure 3c from the magnitude of opinions.

Figure 3a shows that opinions in consensus diverged in line with agent tendencies. This is unsurprising, but the influence of more extreme agents is apparent in the average consensus exceeding the 0.5 tendency magnitude, as suggested by Figure 2a and 2d. Despite the tendencies averaging as 0 and not exceeding the initial opinion range, and moreover all agents initially being intermixed, global average final opinion was significantly more extreme than with 0 tendency ($p < 0.005$). Given that real-world opinion tendencies are more broadly distributed than modeled here, this should reduce expectations of how much consensus in public opinion to expect in reality, even assuming low controversy and reasonable homophily.

Figure 3b shows that heterogeneous tendencies significantly increased average radicalization ($p < 0.005$). The separation of the two tendency groups seen in Figure 2e is evident in the two peaks in Figure 3b. Even agents with −0.5 tendency became significantly more extreme than when all agents had 0 tendency ($p < 0.005$, but group-specific comparison not shown for brevity). This suggests that opposing tendencies are overridden by radicalization, rather than serving to moderate it.

Figure 3c shows that the effect of diversity on polarization was to increase it significantly ($p < 0.005$). All agents were intermixed in their initial opinions, leading to many agents being polarized against their own tendency, with conflicts between exogenous and social influences. Were agents initially segregated, these conflicts would be reduced, suggesting a further increase in polarization would occur. Intermixing of individuals could, therefore,

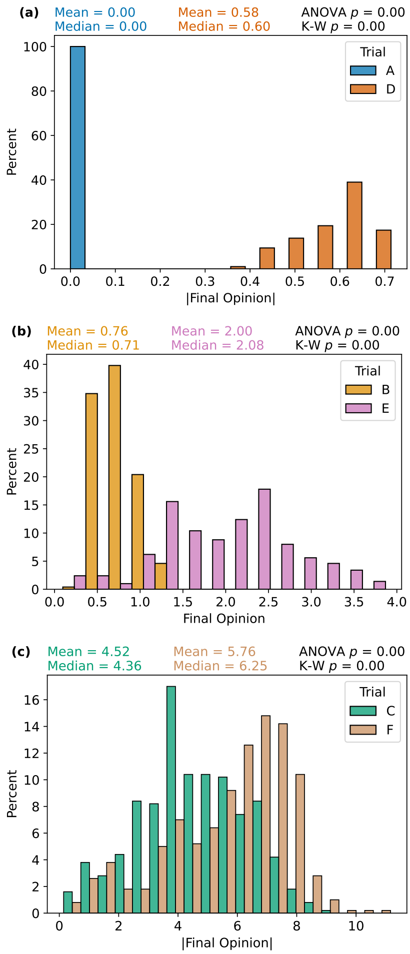

Figure 3: Comparison of final agent opinions with 0 tendency and random ±0.5 tendency for (a) consensus (Trials A and D), (b) radicalization (Trials B and E), and (c) polarization (Trials C and F). Opinion magnitude was used in plots (a) and (c) for reasons of symmetry. Bars are offset in each plot to avoid overlap, and colored by trial. Heterogeneous tendencies caused a significant increase in extreme opinion in all cases ($p < 0.005$).

help to mitigate the increase in polarization caused by diversity.

Please see Online Appendix, Figures A1-4, for a sensitivity analysis of RQ1, which supports the findings presented above.

Time-dependent tendencies

Results for time-dependent tendencies in a minority group of agents are shown in Figure 4 (Trials G–H). In the absence of the minority group, results in Figure 4 would be the same as Figure 1c for polarization (Trial C). Two time-dependent functions were used, as described in the Method: a transient square pulse (Figure 4a and 4c), and a gradual linear ramp (Figure 4b and 4d).

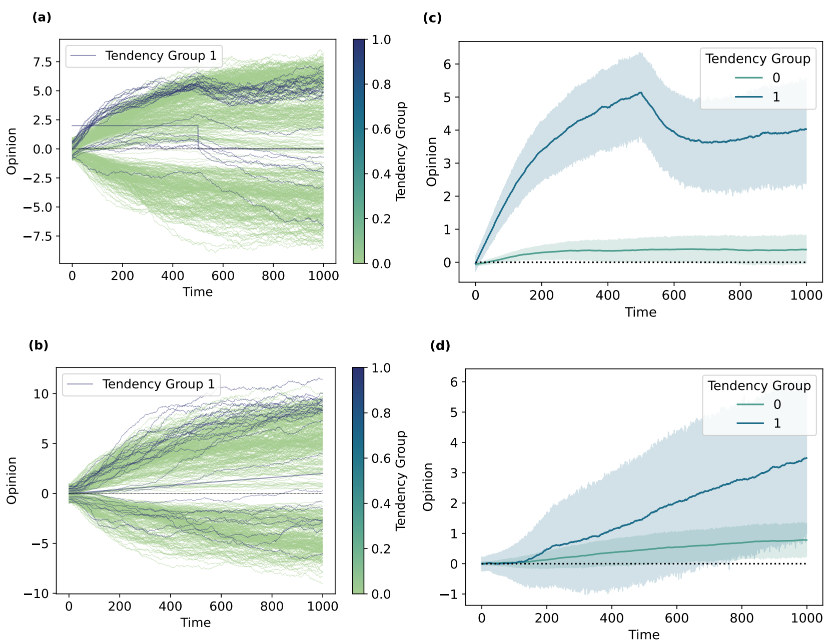

Figure 4: Opinion dynamics in polarizing conditions, with time-dependent tendencies for a minority group of agents, using (a) a square pulse function (Trial G), and (b) a linear ramp function (Trial H). Results in each plot (a-b) are from a single simulation, and lines are colored according to the tendency group of agents (group 0 is the majority group, and group 1 the minority group). Straight lines represent the time-dependent tendency function for the minority group. Plots (c-d) show corresponding aggregated results.

From Figure 4a and 4c, it is evident that a temporary change in tendency in the minority group caused a lasting shift in majority opinion. The minority group returned to the same 0 tendency as the majority group halfway through the simulation, by which time the majority group had already moved to a stable positive overall opinion, despite its 0 tendency ('shifted polarization'). It is notable that the minority group quickly exceeded the tendency function, despite receiving social influence from all agents.

Figure 4b and 4d show that a gradual shift in minority tendencies exhibited far greater variability both in the minority group and the majority, but that on average, majority opinion appeared similarly affected to the pulse in Figure 4a and 4c. The increased variability from the linear ramp may be due to the minority agents being more susceptible to polarization in both directions at the start of the simulation, unlike with the square pulse where the initial tendency was distinctly positive.

The linear ramp was omitted from further consideration for simplicity, but results with an equally sized delayed opposing square pulse minority group are shown in Figure 5 (Trial I). It is evident the opposing group was unable to counteract the effect of the first minority group, which is explored further below.

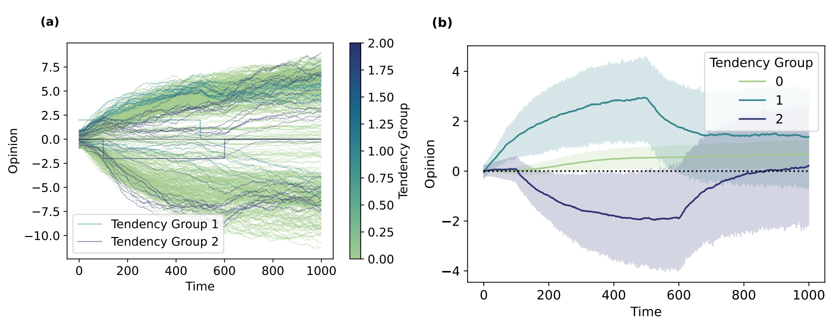

Figure 5: Opinion dynamics in polarizing conditions, with time-dependent tendencies for two opposing minority groups of agents with a temporal offset in behaviors, using a square pulse function (Trial I). Lines are colored according to the tendency group of agents (group 0 is the majority group, group 1 the first minority group, and group 2 the delayed opposing minority group). Plot (a) shows results from a single simulation, with lines colored according to the tendency group of agents. Straight lines represent the time-dependent tendency functions for the minority groups. Plot (b) shows corresponding aggregated results.

RQ2

To answer RQ2, the final majority opinion of agents was compared between Trial C (i.e. polarizing conditions, with all agents having 0 tendency), Trials

G–H (i.e. with pulse and ramp minority tendencies), and Trial I (i.e. with delayed opposing pulse minority tendencies). Comparisons are shown in Figure 6a - 6c for agents specifically in the majority group, i.e. agents with constant 0 tendency in each trial. To explore further the effect of the opposing minority groups, Trials G and I are compared directly in Figure 6d. Unlike the polarization comparison in Figure 3c, results in Figure 6 do not use opinion magnitudes, as there was no expectation of symmetry, due to the presence of asymmetric minority groups.

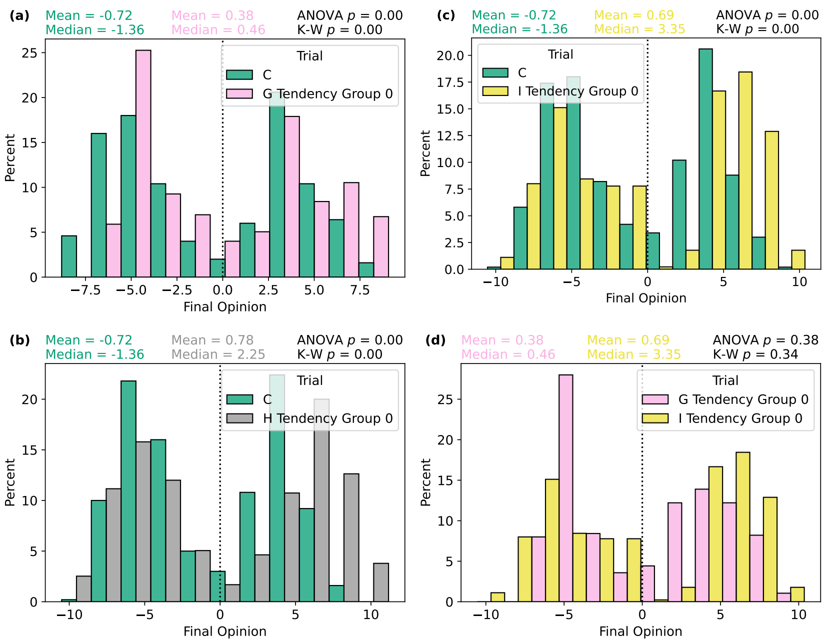

Figure 6: Comparison of final agent opinions for unbiased polarization (Trial C) with the unbiased majority group under influence from (a) pulse function minority group (Trial G), (b) ramp function minority group (Trial H), and (c) pulse function minority group with a delayed opposing group of equal size (Trial I). Plot (d) shows an equivalent comparison between Trials G and I. Bars are offset in each plot to avoid overlap and colored by trial. Plots show that the effect of the minority groups was significant in all cases ($p < 0.005$). The delayed opposing group did not significantly alter the effect of the first minority group (minimum $p = 0.34$), though both the mean and median were larger in the presence of the opposing minority group.

Figure 6a shows that the pulse minority group (Trial G) had a significant effect on majority opinion ($p < 0.005$), demonstrating that minority groups can significantly shift long-term majority opinion with only a transient behavioral change. Figure 6b shows that gradual changes in the minority group (Trial H) are also able to have a significant effect on majority opinion, seemingly at least as large as for the abrupt pulse. Figure 6c - 6d show that the shift in majority opinion was not cancelled out by the delayed opposing minority group (Trial I), and may even have increased the original effect, although this was not found to be significant (minimum $p = 0.34$).

Please see the Online Appendix, Figures A5-10, for a sensitivity analysis of RQ2, which supports the findings presented above.

To illustrate further the effects of the minority groups, aggregated opinion dynamics for Trials G–I are decomposed in Figure 7 by tendency group and initial opinion group. It is notable that initial opinion strongly affects resultant dynamics in each of Trials G–I. Minority group agents were initialized with the same random spread of opinions as all other agents, but if they had been aligned with their future tendencies, it is possible the effect of their tendency would be even greater. This is useful to consider with real-world examples, where changes in tendency may well correlate with initial opinion.

Conclusion

With the aim of improving our understanding of online communication and public opinion by using ABMs, we have extended the model of Baumann et al. (2020) to account for individual opinion tendencies of agents, which may be constant or time-dependent in response to external events. Results for consensus, radicalization and polarization matched well with Baumann et al. (2020) when all agents were given 0 tendency (Figure 1), which in turned matched well with real-world social media data.

In answer to RQ1 (Figures 2 - 3), it was found that heterogeneous opinion tendencies increased extreme opinions in consensus and polarization, in both cases with average opinions greater than the tendency magnitudes themselves. However, heterogeneous tendencies particularly increased the extremity of opinion in radicalization, despite the average tendency being 0. As shown in the sensitivity analysis presented in the Online Appendix, results were to some extent dependent on the choice of sociality parameter value and tendency magnitude, but this does not affect overall findings for RQ1, or understanding of the system in general terms.

In the real-world, these findings suggest that diverse tendencies encourage greater disconnect between opposing groups, either due to the tendencies themselves (split consensus), or the effects of homophily being amplified by

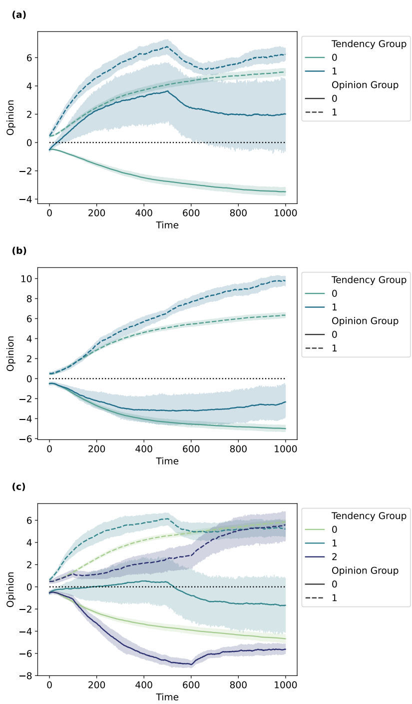

Figure 7: Opinion dynamics for Trials G–I with aggregation of multiple repetitions decomposed by tendency group and initial opinion group. Plot (a) shows results for Trial G, plot (b) for Trial H, and plot (c) for Trial I. Lines are colored according to the tendency group of agents (group 0 is the majority group, group 1 the first minority group, and group 2 the delayed opposing minority group), and styled according to the initial opinion group (group 0 denotes initial opinion ≤ 0, and group 1 denotes initial opinion > 0). In each case, majority opinion is shifted by the minority, in accordance with the form of minority tendency function. The opinion dynamics of agents are highly affected by the combination of initial opinion and tendency group.

opposing tendencies (biased polarization). The most severe effect of diverse tendencies appears to be in (divergent) radicalization, where tendencies amplify the most extreme opinions, rather than moderating them. The result is a greater likelihood of extremes in global opinion. There is also a greater likelihood of entrenched and exaggerated opposing views arising with diversity, either though split consensus or biased polarization. This has the potential to be exploited for financial or political gain, which is potentially concerning, especially given the effects algorithms can have on homophily (Kaiser and Rauchfleisch 2020).

In answer to RQ2 (Figures 4 - 7), a minority group of 5% of agents with a common time-dependent tendency was observed to have a significant effect on majority opinion (shifted polarization). This was found with both an abrupt transient shift in tendencies and a gradual long-term one (Figure 6a - 6b). Exploring time-dependent tendencies further, a second minority group of the same size but with a delayed opposing tendency was found not to counteract the first minority group (Figure 6c - 6d). This highlights the power a small, coordinated group can have in decentralized social networks, and how hard they can be to counteract (Vysotsky and McCarthy 2017).

The timing and magnitude of the minority tendency shift was important to these results, but the findings still generalize (see Online Appendix). Modeling suggests that long-term shifts in mainstream opinion can be caused by transient changes in a minority of individuals if they act in a coordinated way, with opinions that are in a suitable range for the majority to be receptive. These results are consistent with Moussaïd et al. (2013) for a single static social actor, and they also have relevance to the popular concept of the Overton window, i.e. the movable range of opinions to which the majority may be receptive (Szałek 2013).

Gradual changes in the minority group were found to affect majority opinion in a similar way to the transient pulse, though with greater variability, which is useful to note for slow-building issues such as evidence for long-term health risks, and human-caused climate change. Although majority opinion can shift with gradual changes in the minority group, an abrupt intervention has the potential to shift opinion more quickly and consistently, with similar effectiveness, even when the minority group behavior is not sustained. These findings could help inform social and behavior change communication, and also how to handle viral misinformation on social media (Briscoe and Aboud 2012; Jahanbakhsh et al. 2021).

It is noteworthy that agents in the minority groups were no more active than other agents, their initial opinions no different, their tendencies no more strongly held, and there was no difference in how they exerted or received social influence. Although results were obtained only for a finite time period, the effect of the temporary pulse was found to produce stable behavior in the majority group for as long after the pulse as it was active for (Figure 4a and 4c, but also see Online Appendix). How long this would continue has not been established, but the scope for long-term effects resulting from transient events has been credibly demonstrated in the model.

As with all models, the findings reflect the underlying assumptions. The biggest simplification is treating opinion as a single numerical value, where agents change their opinion to reduce disagreement with others. Another important simplification is the global judgement made for homophily, although this is treated as an effect of online platforms. Modeling multi-dimensional opinions, willful disagreement, and different modalities of communication would likely change at least some of the presented findings, but this does not necessarily limit the insights gained here. The simplicity of the model means that cultural, social and political contexts were not represented explicitly, and so generalizability is not bound by these. Similarly, there are all kinds of real-world influences which the model did not seek to represent, such as framing and psychology. A more complex model could attempt to account for these, but would require greater integration with empirical work and theory from other disciplines, and would lose some of the advantages of parsimony.

The potential for a minority to shift opinions in the majority has previously been found in experiments and modeling of closed social networks with endogenous discrete choices (Centola et al. 2018). Previous work has also investigated the role of exogenous network influences on opinion dynamics with binary choices, as well as in a continuum using a model which relied on exogenous events to generate polarization (Wettstein 2020; Sikder et al. 2020; Condie and Condie 2021). Huang and Wen (2014) previously modeled the influence of minority groups by combining group norms with concepts of privately and publicly held opinions, as similarly used by Lim and Bentley (2022) to model polarization. However, as far as we are aware, the present study is the first to examine the role of exogenous influences on opinion dynamics in a decentralized scale-free social network for opinions in a continuum where polarization did not rely on exogenous factors, and also to consider the influence of minority groups with access to new or changing information, without assumptions of fixed group attitudes or identities. The relevance to real-world events such as those described in the Introduction is hopefully clear.

In particular, we believe the present study has helped understand the effects of new information during the Covid pandemic, where misinformation was spread by a minority on social media in response to evolving news, causing potentially life-threatening shifts in wider opinion (Cinelli et al. 2020). It also has relevance to the American Capitol attack, where a minority of individuals with pre-existing beliefs were mobilized on social media in sufficient numbers to cause insurrection (Suresh et al. 2023). It is additionally relevant to the Silicon Valley Bank run, where rumors from a minority spread rapidly on social media to shift public opinion enough to cause a terminal loss of investor confidence (Yerushalmy 2023). In each of these examples, there appears to be a crossover of private and public communication, as well as pre-existing beliefs, which the present study has represented with the tendency term of the model. This has helped identify communication and social mechanisms which can cause crucial shifts in public opinion through minority influence. However, we acknowledge the generic and simplified nature of the presented results, and do not claim to have modelled any of these examples specifically.

Differences in personal characteristics have previously been empirically observed as being important to public opinion (Herek 2002; Kite, Whitley, and Wagner 2023). We have investigated how such differences may affect consensus, radicalization and polarization at a theoretical level (Figures 1 - 3). Of these, polarization has previously received most attention (Kubin and von Sikorski 2021). Building on this, the present study has helped distinguish homophily-driven polarization (Figure 1c) from tendency-driven split consensus (Figure 2a). Both appear similar in outcome, but have different mechanistic causes in the model, which is pertinent to interpreting polarization in the real world.

We believe these findings are relevant to stakeholders beyond communication researchers. For example, they could help journalists and politicians gain a better understanding of observed polarization in society. The present study suggests polarization may be produced by distinct mechanisms, i.e. underlying beliefs and response to social influence, the effects of which are likely to be amplified by their combination. Public discourse and policy making could benefit from identifying these distinctions more clearly.

Regarding the process of opinion change, previous research has tended to focus on how individuals engage with wider social narratives, where civil protest and media coverage were found to be particularly important (Collingwood, Lajevardi, and Oskooii 2018). The effect of minorities on public opinion has been well studied, mainly concentrating on how an individual's sense of identity affects both publicly and privately held opinion (Gardikiotis 2011). The present study adds a new perspective to these empirical observations by reducing the modeled system to a small number of tangible parameters which can replicate many observed qualitative behaviors and their hypothesized causes. Although the study does not imply that more complex aspects of opinion dynamics are unimportant, including the psychology of group behavior, it also shows how realistic outcomes can be generated from relatively simple assumptions, which can be used as the basis for further hypothesis testing.

In particular, the concept of group identity was not present in the model, only the common tendencies and coordination of agents in the same group. This nuance to the effect of identity on decentralized communication is helpful for making sense of real-world observations. Specifically, individuals need not knowingly have anything in common to become an influential minority group, and instead may simply share a news source to which the majority is unaware, as may have occurred in the Silicon Valley Bank run, or a prior belief or experience which causes them to respond differently to news in the public domain, as may have occurred with Covid misinformation and the American Capitol attack. As with the different mechanisms of polarization, we believe these examples have relevance beyond communication research, and could be helpful for politicians, journalists and social media platforms to interpret and anticipate minority influence on public opinion in response to new information. Importantly, minority groups need not recognize themselves as such. Where minority influence could cause harm, interventions may benefit from focusing less on identity, and more on how information is assimilated into the public sphere.

The present study did not make direct use of empirical data, but it is based closely on the model of Baumann et al. (2020) which was validated against social media data. Future work could focus on using empirical data both to inform modeling at the micro-scale and to test it at the macro-scale. There are some suggestions in the literature for how this could be achieved, and we would advocate building on the news link-based approaches used by Baumann et al. (2020) and Cinelli et al. (2021). Other possibilities include experiments using human participants, and linguistic analysis of social media data, considering both the timing and sentiment of interactions (Bazarova, Walther, and McLeod 2012; Liu, Liu, and Chen 2020). This combination of approaches could provide a framework both for parameter estimation and model validation.

A key finding of the present study is that minority groups can shift the modeled majority opinion in dynamic news contexts if they act in a coordinated way. Furthermore, a delayed opposition group may not counteract the first minority group, and could in fact serve to amplify the first group's effect. Beyond empirical validation of this, future work could explore further the distinction between agent-level homophily decisions (e.g. personality and relationship effects) and global-scale mediation (e.g. algorithmic and cultural effects).

Some random effects were evident in the presented results, (e.g. the non-zero average opinion for Trial C in Figure 6), but running multiple simulations helped reduce the issue. However, there is more work to be done in understanding the role of stochasticity, as although aggregating results helped to identify general trends, this also has the risk of ironing out interesting behavior, and conflating variability within and across repetitions. The use of statistical tests was affected by the number of repetitions used, and although keeping this constant made the comparisons fair, future work could give more attention to effect size (Sawilowsky 2009). We recommend further use of sensitivity analysis, including a global variance-based method to improve understanding of the model parameters and their interactions (Saltelli and Annoni 2010). It would also be useful to explore the effects of different network topologies, distributions and restructuring processes (Sikder et al. 2020; Kan, Feng, and Porter 2023).

The model could be extended to include other forms of agent-specific and time-dependent behavior, similar to the treatment of opinion tendencies in the present study. This could involve any model parameters, but perhaps most salient would be to apply a similar approach to the message rate of agents in response to receiving messages they particularly agree with, as observed empirically, and encouraged through common platform affordances such as likes and shares (Myers and Leskovec 2014). To build on this, the role of message virality would be another suitable consideration for future work (Kim 2018).

Future work could also investigate more structured and stable networks, reflecting, for example, friend and follower functionalities. This could include investigation of heterophilous and parasocial interactions, where social tension may be particularly relevant (Lozares et al. 2014; Dibble, Hartmann, and Rosaen 2016). Tactics of opposing minority groups could also be investigated with the model, as well as the effect of multiple groups (Gallagher et al. 2018). The effect of network size could also be explored, including behaviors within small private groups, which is an increasingly common and under-analyzed usage of social media, and one where agent-based modeling could be particularly informative (Dargahi Nobari et al. 2021).

Acknowledgements

We thank Fabian Baumann (Center for Humans & Machines, MPI Berlin) for helpful replies to questions about the original model. We also thank the editor and anonymous reviewers for insightful comments. This work was undertaken on Barkla, part of the High Performance Computing facilities at the University of Liverpool.

Notes

[1]See https://doi.org/10.5281/zenodo.10718505 for the Online Appendix.

[2]See https://github.com/markpogson/odmodel for the full code used in this paper.

References

Acemoglu, D., and A. Ozdaglar. 2011. "Opinion Dynamics and Learning in Social Networks." *Dynamic Games and Applications* 1: 3–49. https://doi.org/10.1007/s13235-010-0004-1.

Ali, S. R., and S. Fahmy. 2013. "Gatekeeping and Citizen Journalism: The Use of Social Media During the Recent Uprisings in Iran, Egypt, and Libya." *Media, War & Conflict* 6 (1): 55–69. https://doi.org/10.1177/1750635212469906.

Amit, S., S. Mannan, and M. A. Islam. 2020. "Bangladesh: Time Spent Online, Conflict and Radicalization." *Conflict Studies Quarterly* 31. https://doi.org/10.24193/csq.31.1.

Anderson, B. D., and M. Ye. 2019. "Recent Advances in the Modelling and Analysis of Opinion Dynamics on Influence Networks." *International Journal of Automation and Computing* 16 (2): 129–49. https://doi.org/10.1007/s11633-019-1169-8.

Bail, C. A., L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. F. Hunzaker, and A. Volfovsky. 2018. "Exposure to Opposing Views on Social Media Can Increase Political Polarization." *Proceedings of the National Academy of Sciences* 115 (37): 9216–21. https://doi.org/10.1073/pnas.1804840115.

Baker, R. E., J. M. Pena, J. Jayamohan, and A. Jérusalem. 2018. "Mechanistic Models Versus Machine Learning, a Fight Worth Fighting for the Biological Community?" *Biology Letters* 14 (5): 20170660. https://doi.org/10.1098/rsbl.2017.0660.

Bakshy, E., S. Messing, and L. A. Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science (New York, N.Y.)* 348 (6239): 1130–32. https://doi.org/10.1126/science.aaa1160.

Banisch, S., R. Lima, and T. Araújo. 2012. "Agent Based Models and Opinion Dynamics as Markov Chains." *Social Networks* 34 (4): 549–61. https://doi.org/10.1016/j.socnet.2012.06.001.

Baran, P. 1964. "On Distributed Communications Networks." *IEEE Transactions on Communications Systems* 12 (1): 1–9. https://doi.org/10.1109/TCOM.1964.1088883.

Baumann, F., P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini. 2020. "Modeling Echo Chambers and Polarization Dynamics in Social Networks." *Physical Review Letters* 124 (4). https://doi.org/10.1103/PhysRevLett.124.048301.

Bazarova, N. N., J. B. Walther, and P. L. McLeod. 2012. "Minority Influence in Virtual Groups: A Comparison of Four Theories of Minority Influence." *Communication Research* 39 (3): 295–316. https://doi.org/10.1177/0093650211399752.

Becchetti, Luca, Andrea Clementi, Amos Korman, Francesco Pasquale, Luca Trevisan, and Robin Vacus. 2023. "On the Role of Memory in Robust Opinion Dynamics." February 16, 2023. http://arxiv.org/abs/2302.08600.

Bonabeau, E. 2002. "Agent-Based Modeling: Methods and Techniques for Simulating Human Systems." *Proceedings of the National Academy of Sciences* 99: 7280–87. https://doi.org/10.1073/pnas.082080899.

Briscoe, Ciara, and Frances Aboud. 2012. "Behaviour Change Communication Targeting Four Health Behaviours in Developing Countries: A Review of Change Techniques." *Social Science & Medicine*, Part Special Issue: Challenges to changing health behaviours in developing countries, 75 (4): 612–21. https://doi.org/10.1016/j.socscimed.2012.03.016.

Centola, D., and A. Baronchelli. 2015. "The Spontaneous Emergence of Conventions: An Experimental Study of Cultural Evolution." *Proceedings of the National Academy of Sciences* 112 (7): 1989–94. https://doi.org/10.1

073/pnas.1418838112.
Centola, D., J. Becker, D. Brackbill, and A. Baronchelli. 2018. "Experimental evidence for tipping points in social."
Cinelli, M., G. D. Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini. 2021. "The Echo Chamber Effect on Social Media." *Proceedings of the National Academy of Sciences* 118 (9). https://doi.org/10.1073/pnas.2023301118.
Cinelli, M., W. Quattrociocchi, A. Galeazzi, C. M. Valensise, E. Brugnoli, A. L. Schmidt, P. Zola, F. Zollo, and A. Scala. 2020. "The COVID-19 social media infodemic." *Scientific Reports* 10 (1): 1–10. https://doi.org/10.1038/s41598-020-73510-5.
Collingwood, L., N. Lajevardi, and K. A. Oskooii. 2018. "A Change of Heart? Why Individual-Level Public Opinion Shifted Against Trump's 'Muslim Ban'." *Political Behavior* 40: 1035–72. https://doi.org/10.1007/s11109-017-9439-z.
Condie, S. A., and C. M. Condie. 2021. "Stochastic Events Can Explain Sustained Clustering and Polarisation of Opinions in Social Networks." *Scientific Reports* 11 (1): 1355. https://doi.org/10.1038/s41598-020-80353-7.
Crespi, Irving. 1997. *The Public Opinion Process: How the People Speak*. 1st edition. Mahwah, NJ: Routledge.
Dargahi Nobari, A., M. H. K. M. Sarraf, M. Neshati, and F. E. Daneshvar. 2021. "Characteristics of Viral Messages on Telegram; The World's Largest Hybrid Public and Private Messenger." *Expert Systems with Applications* 168: 114303. https://doi.org/10.1016/j.eswa.2020.114303.
Das, A., S. Gollapudi, and K. Munagala. 2014. "February. Modeling Opinion Dynamics in Social Networks." In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, 403–12. https://doi.org/10.1145/2556195.2559896.
Deffuant, Guillaume, David Neau, Frederic Amblard, and Gérard Weisbuch. 2000. "Mixing Beliefs Among Interacting Agents." *Advances in Complex Systems* 03 (January): 87–98. https://doi.org/10.1142/S0219525900000078.
DeGroot, M. H. 1974. "Reaching a Consensus." *Journal of the American Statistical Association* 69 (345): 118–21. https://doi.org/10.1080/01621459.1974.10480137.
Del Vicario, M., A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi. 2016. "The Spreading of Misinformation Online." *Proceedings of the National Academy of Sciences* 113 (3): 554–59. https://doi.org/10.1073/pnas.1517441113.
Del Vicario, M., F. Zollo, G. Caldarelli, A. Scala, and W. Quattrociocchi. 2017.

"Mapping Social Dynamics on Facebook: The Brexit Debate." *Social Networks* 50: 6–16. https://doi.org/10.1016/j.socnet.2017.02.002.

Della Porta, D. 2018. "Radicalization: A Relational Perspective." *Annual Review of Political Science* 21: 461–74. https://doi.org/10.1146/annurev-polisci-042716-102314.

Dibble, J. L., T. Hartmann, and S. F. Rosaen. 2016. "Parasocial Interaction and Parasocial Relationship: Conceptual Clarification and a Critical Assessment of Measures." *Human Communication Research* 42 (1): 21–44. https://doi.org/10.1111/hcre.12063.

Dorogovtsev, Sergey. 2010. *Lectures on Complex Networks: 20.* Illustrated edition. Oxford ; New York: OUP Oxford.

Dubois, Elizabeth, and Grant Blank. 2018. "The Echo Chamber Is Overstated: The Moderating Effect of Political Interest and Diverse Media." *Information, Communication & Society* 21 (5): 729–45. https://doi.org/10.1080/1369118X.2018.1428656.

Fan, Rui, Ke Xu, and Jichang Zhao. 2020. "Weak Ties Strengthen Anger Contagion in Social Media." May 4, 2020. https://doi.org/10.48550/arXiv.2005.01924.

Flache, Andreas, and Michael W. Macy. 2011. "Small Worlds and Cultural Polarization." *The Journal of Mathematical Sociology* 35 (1-3): 146–76. https://doi.org/10.1080/0022250X.2010.532261.

Fletcher, Richard, Antonis Kalogeropoulos, and Rasmus Kleis Nielsen. 2021. "More Diverse, More Politically Varied: How Social Media, Search Engines and Aggregators Shape News Repertoires in the United Kingdom." *New Media & Society*, July, 14614448211027393. https://doi.org/10.1177/14614448211027393.

Gallagher, R. J., A. J. Reagan, C. M. Danforth, and P. S. Dodds. 2018. "Divergent Discourse Between Protests and Counter-Protests: #BlackLivesMatter and #AllLivesMatter." *PLoS One* 13 (4): 0195644. https://doi.org/10.1371/journal.pone.0195644.

Gardikiotis, A. 2011. "Minority Influence." *Social and Personality Psychology Compass* 5 (9): 679–93. https://doi.org/10.1111/j.1751-9004.2011.00377.x.

Gargiulo, F., and Y. Gandica. 2017. "The Role of Homophily in the Emergence of Opinion Controversies." *Journal of Artificial Societies and Social Simulation* 20 (3): 8. https://doi.org/10.18564/jasss.3448.

Gayo-Avello, D. 2015. "Social media, democracy, and democratization." *IEEE Multimedia* 22 (2): 10–16. https://doi.org/10.1109/MMUL.2015.47.

Geiß, Stefan, Melanie Magin, Pascal Jürgens, and Birgit Stark. 2021. "Loopholes in the Echo Chambers: How the Echo Chamber Metaphor Oversimplifies the Effects of Information Gateways on Opinion Expression." *Digital Journalism* 9 (5): 660–86. https://doi.org/10.1080/21670811.2021.1873811.

Guo, L., A. Rohde, J., and H. D. Wu. 2020. "Who Is Responsible for Twitter's Echo Chamber Problem? Evidence from 2016 US Election Networks." *Information, Communication & Society* 23 (2): 234–51. https://doi.org/10.1080/1369118X.2018.1499793.

Harris, C. R., K. J. Millman, S. J. Walt, R. Gommers, P. Virtanen, D. Cournapeau, and T. E. Oliphant. 2020. "Array Programming with NumPy." *Nature* 585 (7825): 357–62. https://doi.org/10.1038/s41586-020-2649-2.

Hegselmann, R., and U. Krause. 2002. "Opinion Dynamics and Bounded Confidence: Models, Analysis and Simulation." *Journal of Artifical Societies and Social Simulation* 5 (3). https://jasss.org/5/3/2/2.pdf.

Heltzel, Gordon, and Kristin Laurin. 2020. "Polarization in America: Two Possible Futures." *Current Opinion in Behavioral Sciences*, Political Ideologies, 34 (August): 179–84. https://doi.org/10.1016/j.cobeha.2020.03.008.

Herbst, Susan. 1993. "The Meaning of Public Opinion: Citizens' Constructions of Political Reality." *Media, Culture & Society* 15 (3): 437–54. https://doi.org/10.1177/016344393015003007.

Herek, G. M. 2002. "Gender Gaps in Public Opinion about Lesbians and Gay Men." *Public Opinion Quarterly* 66 (1): 40–66. https://doi.org/10.1086/338409.

Hitkul, Avinash Prabhu, Dipanwita Guhathakurta, Jivitesh Jain, Mallika Subramanian, Manvith Reddy, Shradha Sehgal, et al. 2021. "Capitol (Pat)riots: A Comparative Study of Twitter and Parler." January 18, 2021. https://doi.org/10.48550/arXiv.2101.06914.

Huang, Chung-Yuan, and Tzai-Hung Wen. 2014. "A Novel Private Attitude and Public Opinion Dynamics Model for Simulating Pluralistic Ignorance and Minority Influence." *Journal of Artificial Societies and Social Simulation* 17 (3): 8. https://doi.org/10.18564/jasss.2517.

Hunter, J. D. 2007. "Matplotlib: A 2D Graphics Environment." *Computing in Science & Engineering* 9 (3): 90–95. https://doi.org/10.1109/MCSE.2007.55.

Jahanbakhsh, Farnaz, Amy X. Zhang, Adam J. Berinsky, Gordon Pennycook, David G. Rand, and David R. Karger. 2021. "Exploring Lightweight Interventions at Posting Time to Reduce the Sharing of Misinformation on Social Media." *Proceedings of the ACM on Human-Computer Interaction* 5 (April): 18:1–42. https://doi.org/10.1145/3449092.

Jasny, L., J. Waggle, and D. R. Fisher. 2015. "An Empirical Examination of

Echo Chambers in US Climate Policy Networks." *Nature Climate Change* 5 (8): 782–86. https://doi.org/10.1038/nclimate2666.

Jayles, B., H. R. Kim, R. Escobedo, S. Cezera, A. Blanchet, T. Kameda, and G. Theraulaz. 2017. "How Social Information Can Improve Estimation Accuracy in Human Groups." *Proceedings of the National Academy of Sciences* 114 (47): 12620–25. https://doi.org/10.1073/pnas.1703695114.

Joblib. 2023. "Joblib: Running Python Functions as Pipeline Jobs." https://joblib.readthedocs.io/.

Johnson, Roger W. 2022. "Alternate Forms of the One-Way ANOVA F and Kruskal–Wallis Test Statistics." *Journal of Statistics and Data Science Education* 30 (1): 82–85. https://doi.org/10.1080/26939169.2021.2025177.

Jusup, M., P. Holme, K. Kanazawa, M. Takayasu, I. Romić, Z. Wang, and M. Perc. 2022. "Social physics." *Physics Reports* 948: 1–148. https://doi.org/10.1016/j.physrep.2021.10.005.

Kaiser, Jonas, and Adrian Rauchfleisch. 2020. "Birds of a Feather Get Recommended Together: Algorithmic Homophily in YouTube's Channel Recommendations in the United States and Germany." *Social Media + Society* 6 (4): 2056305120969914. https://doi.org/10.1177/2056305120969914.

Kan, Unchitta, Michelle Feng, and Mason A Porter. 2023. "An Adaptive Bounded-Confidence Model of Opinion Dynamics on Networks." *Journal of Complex Networks* 11 (1): 415–44. https://doi.org/10.1093/comnet/cnac055.

Kim, Ji Won. 2018. "They Liked and Shared: Effects of Social Media Virality Metrics on Perceptions of Message Influence and Behavioral Intentions." *Computers in Human Behavior* 84 (July): 153–61. https://doi.org/10.1016/j.chb.2018.01.030.

Kite, Mary E., Bernard E. Whitley, and Lisa Wagner. 2023. *Psychology of Prejudice and Discrimination.* Fourth edition. New York London: Routledge, Taylor & Francis Group. https://doi.org/10.4324/9780367809218.

Kozitsin, Ivan V. 2023. "Opinion Dynamics of Online Social Network Users: A Micro-Level Analysis." *The Journal of Mathematical Sociology* 47 (1): 1–41. https://doi.org/10.1080/0022250X.2021.1956917.

Kubin, Emily, and Christian von Sikorski. 2021. "The Role of (Social) Media in Political Polarization: A Systematic Review." *Annals of the International Communication Association* 45 (3): 188–206. https://doi.org/10.1080/23808985.2021.1976070.

Lim, Soo Ling, and Peter J. Bentley. 2022. "Opinion Amplification Causes Extreme Polarization in Social Networks." *Scientific Reports* 12 (1, 1): 18131. https://doi.org/10.1038/s41598-022-22856-z.

Liu, Yanni, Dongsheng Liu, and Yuwei Chen. 2020. "Research on Sentiment Tendency and Evolution of Public Opinions in Social Networks of Smart City." *Complexity* 2020 (June): e9789431. https://doi.org/10.1155/2020/9789431.

Lorenz-Spreen, Philipp, Stephan Lewandowsky, Cass R. Sunstein, and Ralph Hertwig. 2020. "How Behavioural Sciences Can Promote Truth, Autonomy and Democratic Discourse Online." *Nature Human Behaviour* 4 (11, 11): 1102–9. https://doi.org/10.1038/s41562-020-0889-7.

Lorenz-Spreen, Philipp, Lisa Oswald, Stephan Lewandowsky, and Ralph Hertwig. 2023. "A Systematic Review of Worldwide Causal and Correlational Evidence on Digital Media and Democracy." *Nature Human Behaviour* 7 (1, 1): 74–101. https://doi.org/10.1038/s41562-022-01460-1.

Lozares, C., J. M. Verd, I. Cruz, and O. Barranco. 2014. "Homophily and Heterophily in Personal Networks. From Mutual Acquaintance to Relationship Intensity." *Quality & Quantity* 48: 2657–70. https://doi.org/10.1007/s11135-013-9915-4.

Luceri, Luca, Ashok Deb, Adam Badawy, and Emilio Ferrara. 2019. "Red Bots Do It Better: Comparative Analysis of Social Bot Partisan Behavior." In *Companion Proceedings of The 2019 World Wide Web Conference*, 1007–12. WWW '19. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3308560.3316735.

McCauley, C., and S. Moskalenko. 2017. "Understanding Political Radicalization: The Two-Pyramids Model." *American Psychologist* 72 (3): 205. https://doi.org/10.1037/amp0000062.

Mckeown, G., and N. Sheehy. 2006. "Mass Media and Polarisation Processes in the Bounded Confidence Model of Opinion Dynamics." *Journal of Artificial Societies and Social Simulation* 9 (1). https://www.jasss.org/9/1/11.html.

McKinney, W. 2010. "Data Structures for Statistical Computing in Python." In *Proceedings of the 9th Python in Science Conference*, edited by S. Walt and J. Millman, 51–56. https://doi.org/10.25080/Majora-92bf1922-00a.

Minh Pham, T., I. Kondor, R. Hanel, and S. Thurner. 2020. "The Effect of Social Balance on Social Fragmentation." *Journal of the Royal Society Interface* 17 (172): 20200752. https://doi.org/10.1098/rsif.2020.0752.

Moinet, A., M. Starnini, and R. Pastor-Satorras. 2015. "Burstiness and Aging in Social Temporal Networks." *Physical Review Letters* 114 (10): 108701. https://doi.org/10.1103/PhysRevLett.114.108701.

Moussaïd, M., J. E. Kämmer, P. P. Analytis, and H. Neth. 2013. "Social Influence and the Collective Dynamics of Opinion Formation." *PLoS One* 8

(11): 78433. https://doi.org/10.1371/journal.pone.0078433.

Myers, Seth A., and Jure Leskovec. 2014. "The Bursty Dynamics of the Twitter Information Network." In *Proceedings of the 23rd International Conference on World Wide Web*, 913–24. WWW '14. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2566486.2568043.

Napoli, P. M. 2015. "Social Media and the Public Interest: Governance of News Platforms in the Realm of Individual and Algorithmic Gatekeepers." *Telecommunications Policy* 39 (9): 751–60. https://doi.org/10.1016/j.telpol.2014.12.003.

Neubaum, G., and N. C. Krämer. 2017. "Monitoring the Opinion of the Crowd: Psychological Mechanisms Underlying Public Opinion Perceptions on Social Media." *Media Psychology* 20 (3): 502–31. https://doi.org/10.1080/15213269.2016.1211539.

Noorazar, H. 2020. "Recent Advances in Opinion Propagation Dynamics: A 2020 Survey." *The European Physical Journal Plus* 135: 1–20. https://doi.org/10.1140/epjp/s13360-020-00541-2.

Page, Benjamin I., Robert Y. Shapiro, and Glenn R. Dempsey. 1987. "What Moves Public Opinion?" *The American Political Science Review* 81 (1): 23–43. https://doi.org/10.2307/1960777.

Parsegov, S. E., A. V. Proskurnikov, R. Tempo, and N. E. Friedkin. 2016. "Novel Multidimensional Models of Opinion Dynamics in Social Networks." *IEEE Transactions on Automatic Control* 62 (5): 2270–85. https://doi.org/10.1109/TAC.2016.2613905.

Podobnik, B., M. Jusup, D. Kovac, and H. E. Stanley. 2017. "Predicting the Rise of EU Right-Wing Populism in Response to Unbalanced Immigration." *Complexity*, no. 2017. https://doi.org/10.1155/2017/1580526.

Prier, J. 2017. "Commanding the Trend: Social Media as Information Warfare." *Strategic Studies Quarterly* 11 (4): 50–85.

Rossum, Guido van. 2020. *The Python Library Reference, Release 3.8.2*. Python Software Foundation.

Sacco, V., and D. Bossio. 2015. "Using Social Media in the News Reportage of War & Conflict: Opportunities and Challenges." *The Journal of Media Innovations* 2 (1): 59–76. https://doi.org/10.5617/jmi.v2i1.898.

Saltelli, A., and P. Annoni. 2010. "How to Avoid a Perfunctory Sensitivity Analysis." *Environmental Modelling & Software* 25 (12): 1508–17. https://doi.org/10.1016/j.envsoft.2010.04.012.

Sawilowsky, Shlomo. 2009. "New Effect Size Rules of Thumb." *Journal of Modern Applied Statistical Methods* 8 (2). https://doi.org/10.22237/jmasm/1257035100.

Sikder, O., R. E. Smith, P. Vivo, and G. Livan. 2020. "A minimalistic model of bias, polarization and misinformation in social networks." *Scientific Reports* 10 (1): 5493. https://doi.org/10.1038/s41598-020-62085-w.

Song, Hyunjin, and Hajo G. Boomgaarden. 2017. "Dynamic Spirals Put to Test: An Agent-Based Model of Reinforcing Spirals Between Selective Exposure, Interpersonal Networks, and Attitude Polarization." *Journal of Communication* 67 (2): 256–81. https://doi.org/10.1111/jcom.12288.

Stroud, Natalie Jomini. 2010. "Polarization and Partisan Selective Exposure." *Journal of Communication* 60 (3): 556–76. https://doi.org/10.1111/j.1460-2466.2010.01497.x.

Suresh, Vishnuprasad Padinjaredath, Gianluca Nogara, Felipe Cardoso, Stefano Cresci, Silvia Giordano, and Luca Luceri. 2023. "Tracking Fringe and Coordinated Activity on Twitter Leading up to the US Capitol Attack." July 17, 2023. http://arxiv.org/abs/2302.04450.

Swol, L., S. Lee, and R. Hutchins. 2022. *The Banality of Extremism: The Role of Group Dynamics and Communication of Norms in Polarization on January 6*. Group Dynamics: Theory, Research, and Practice.

Szałek, B. Z. 2013. "Some Praxiological Reflections on the so-Called 'Overton Window of Political Possibilities,' 'Framing' and Related Problems." *Reality of Politics. Estimates-Comments-Forecasts* 4: 237–57. https://doi.org/10.15804/rop201315.

Törnberg, P. 2022. "How Digital Media Drive Affective Polarization Through Partisan Sorting." *Proceedings of the National Academy of Sciences* 119 (42): 2207159119. https://doi.org/10.1073/pnas.2207159119.

Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, et al. 2020. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python." *Nature Methods* 17 (3, 3): 261–72. https://doi.org/10.1038/s41592-019-0686-2.

Vysotsky, Stanislav, and Adrienne L. McCarthy. 2017. "Normalizing Cyberracism: A Neutralization Theory Analysis." *Journal of Crime and Justice* 40 (4): 446–61. https://doi.org/10.1080/0735648X.2015.1133314.

Waldherr, A., S. Geise, M. Mahrt, C. Katzenbach, and C. Nuernbergk. 2021. "Toward a Stronger Theoretical Grounding of Computational Communication Science: How Macro Frameworks Shape Our Research Agendas." *Computational Communication Research* 3 (2): 1–28. https://doi.org/10.5117/CCR2021.02.002.WALD.

Waldherr, A., and M. Wettstein. 2019. "Bridging the Gaps: Using Agent-Based Modeling to Reconcile Data and Theory in Computational Communication Science." *International Journal of Communication* 13: 3976–99.

Wang, X., A. D. Sirianni, S. Tang, Z. Zheng, and F. Fu. 2020. "Public Discourse and Social Network Echo Chambers Driven by Socio-Cognitive Biases." *Physical Review X* 10 (4): 041042. https://doi.org/10.1103/PhysRevX.10.041042.

Waskom, M. L. 2021. "Seaborn: Statistical Data Visualization." *Journal of Open Source Software* 6 (60): 3021. https://doi.org/10.21105/joss.03021.

Wettstein, Martin. 2020. "Simulating Hidden Dynamics: Introducing Agent-Based Models as a Tool for Linkage Analysis." *Computational Communication Research* 2 (1, 1): 1–33. https://doi.org/10.5117/CCR2020.1.001.WETT.

Wolf, M., G. S. Doorn, and F. J. Weissing. 2008. "Evolutionary Emergence of Responsive and Unresponsive Personalities." *Proceedings of the National Academy of Sciences* 105 (41): 15825–30. https://doi.org/10.1073/pnas.0805473105.

Yerushalmy, Jonathan. 2023. "'The First Twitter-fuelled Bank Run': How Social Media Compounded SVB's Collapse." *The Observer: Business*, March 16, 2023. https://www.theguardian.com/business/2023/mar/16/the-first-twitter-fuelled-bank-run-how-social-media-compounded-svbs-collapse.