The British Psychological Society

Special section paper

The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers

Daniel Geschke[1]*, Jan Lorenz[2,3] (iD) and Peter Holtz[4] (iD)

[1]Institut für Demokratie und Zivilgesellschaft (Institute for Democracy and Civil Society, IDZ), Jena, Germany
[2]BIGSSS Bremen International Graduate School of Social Sciences, Jacobs University, Bremen, Germany
[3]Department of Computational Social Science, GESIS Leibniz Institute for the Social Sciences, Cologne, Germany
[4]Leibniz-Institut für Wissensmedien IWM (Knowledge Media Research Center), Tübingen, Germany

Filter bubbles and echo chambers have both been linked recently by commentators to rapid societal changes such as Brexit and the polarization of the US American society in the course of Donald Trump's election campaign. We hypothesize that information filtering processes take place on the individual, the social, and the technological levels (triple-filter-bubble framework). We constructed an agent-based modelling (ABM) and analysed twelve different information filtering scenarios to answer the question under which circumstances social media and recommender algorithms contribute to fragmentation of modern society into distinct echo chambers. Simulations show that, even without any social or technological filters, echo chambers emerge as a consequence of cognitive mechanisms, such as confirmation bias, under conditions of central information propagation through channels reaching a large part of the population. When social and technological filtering mechanisms are added to the model, polarization of society into even more distinct and less interconnected echo chambers is observed. Merits and limits of the theoretical framework, and more generally of studying complex social phenomena using ABM, are discussed. Directions for future research such as ways of comparing our simulations with actual empirical data and possible measures against societal fragmentation on the three different levels are suggested.

The ubiquitous availability of information in the age of social media and the personalization of information flows have had substantial effects on our daily lives and on our socio-political culture (Castells, 2010; Happer & Philo, 2013; Hermida, Fletcher, Korell, & Logan, 2012). In consequence, we are currently observing a polarization and

*Correspondence should be addressed to Dr. Daniel Geschke, IDZ Institut für Demokratie und Zivilgesellschaft, Talstr. 84, 07743 Jena, Germany (email: daniel.geschke@idz-jena.de).

fragmentation of the political sphere in many countries (e.g., Sunstein, 2018). Notable examples include the rise in populist politics and mass protests against immigration in Europe, the polarizing election campaign of Donald Trump, or the British people's vote for Brexit. These events developed relatively fast and were surprising for many observers, who wondered about the reasons for these rapid social changes. The possible influence of modern communication technologies on these events is discussed not only in academia, but also in newspapers and political debates (e.g., Ott, 2017).

Still, there is disagreement about if and under which circumstances these technological changes have positive or negative consequences for individual users and society as a whole. Whereas some optimists are confident that the Internet and social media eventually will expand everyone's chances to find unbiased information (e.g., Michal Kosinski in Noor, 2017), critics have warned of the emergence of *filter bubbles*, minimizing exposure to information that challenges individual attitudes (Pariser, 2011). Filter bubbles are defined here as an individual outcome of different processes of information search, perception, selection, and remembering the sum of which causes individual users to receive from the universe of available information only a tailored selection that fits their pre-existing attitudes. On the societal level, individuals tend to *share* a common social media bubble with like-minded friends (Boutyline & Willer, 2017; McPherson, Smith-Lovin, & Cook, 2001); over time, such communities in which Internet content that confirms certain ideologies is echoed from all sides are particularly prone to processes of group radicalization and polarization (Vinokur & Burnstein, 1978); this phenomenon has come to be known as the *echo chamber effect* (Garrett, 2009; Sunstein, 2001, 2009). Thus, echo chambers are a social phenomenon where the filter bubbles of interacting individuals strongly overlap. The dangers of a society falling apart into distinct echo chambers can be described as a lack of society-wide consensus and a lack of at least some shared beliefs among otherwise disagreeing people that are needed for processes of democratic decision-making (Sunstein, 2001, 2009). Furthermore, increasingly radicalized ideological online groups may at some point resort to real-life violence and terrorism to achieve their goals (e.g., Holtz, Wagner, & Sartawi, 2015; Weiman, 2006).

In our study, we analyse the interplay between processes on the levels of individual minds, social groups, and technology, using agent-based modelling (ABM; Smith & Conrey, 2007). We employ an interdisciplinary approach in so far as our model takes into account theories from cognitive psychology, social psychology, and micro-sociology. The goal of our study is first to summarize relevant existing theories on filter bubbles and echo chamber effects in a formal ABM and then to explore the dynamics of this model. The resulting 'triple-filter-bubble model' is a step towards understanding how technological changes and their interplay with cognitive and social factors can contribute to rapid societal change.

Agent-based modelling

Rapid social change follows non-linear dynamics on interacting micro-, meso-, and macro-levels. It is therefore often hard to predict, when using traditional social scientific research methods that (mostly) rely on linear connections of a limited number of variables, rather than simulating multi-level and non-linear interactions. In ABM, interactions between individual agents are simulated as a consequence of rules that were set by the researcher. Effects of interest are usually systemic and appear on the macro-level (Flache *et al.*, 2017). That means, the dependent variables of interest are often properties of the society and not of individuals. ABM is particularly suited for studying interactions between individuals

instead of interactions between variables (Smith & Conrey, 2007) and how such interpersonal influence processes play out (Mason, Conrey, & Smith, 2007). This approach enables us to distinguish individual effects from effects which emerge in their interaction with macroscopic phenomena and, in turn, to tentatively gauge possible effects of changing technological environments on individual users *and* society. Thus, ABM can be used to test the dynamics of complex, non-linear, theoretical assumptions and as a tool for theory building. The method has been applied in previous studies to simulate such different social phenomena as the occurrence of traffic jams (Bazzan & Klügl, 2014) or segregation in housing (Huang, Parker, Filatova, & Sun, 2014; Schelling, 1971). In the next paragraph, we will briefly summarize the available empirical research on filter bubble and echo chamber effects and afterwards present our triple-filter-bubble framework.

Empirical findings on filter bubbles and echo chambers

Already in 2009, Garrett had found evidence that within the political domain, Internet users preferred to consume information that confirmed their ideologies. These findings build upon earlier research on the effects of 'traditional' mass media such as television (e.g., Vallone, Ross, & Lepper, 1985). Del Vicario *et al.* (2016) were recently able to show in a study using Facebook data how information is passed along ideological fault lines in scientific as well as conspiracy-theory communities. Bakshy, Messing, and Adamic (2015) analysed the data of 10.1 million US American Facebook users who identified themselves as being either politically liberal, moderate, or conservative. They found that most information filtering is the result of *homophily*, in the sense that Facebook users have significantly more friends with a political orientation similar to their own. The Facebook newsfeed then relies on information that was shared by at least one person in the friend network, and this already leads to information selection with a severe bias in favour of information confirming a certain ideology. However, an earlier study of the same research group (Bakshy, Rosenn, Marlow, & Adamic, 2012) found that the majority of the information that was displayed in the Facebook newsfeed was not shared by close friends with whom the Facebook users exchanged chats and comments and updates on a regular basis (*strong ties*), but by acquaintances with whom users only communicated on a casual basis (*weak ties*; cf. Granovetter, 1973). As long as there is at least some heterogeneity within a user's friend network, the user will at least have some exposure to differing points of view. Such information would be totally out of the user's reach if information were only accessible via offline communication with close friends who normally share a person's ideologies and beliefs.

A somewhat related result was found using Twitter data (Vaccari *et al.*, 2016): On the one hand, Twitter users more frequently interact (comment or retweet) with authors with a similar political ideology; still, Twitter is used frequently as well to interact with representatives of networks that display an oppositional ideology. The authors draw the conclusion that apparently there are not only echo chambers on Twitter, but also *contrarian clubs*. However, another group of researchers found a definite longitudinal political polarization of primarily US American Twitter users: Over time (between 2009 and 2016), the number of politicians and media with similar ideologies that the users followed increased continuously. However, this finding could also be a result of Donald Trump's polarizing election campaign (Garimella & Weber, 2017).

Recently, the specific role of recommender systems in the emergence of filter bubbles and echo chambers has been investigated in a number of empirical studies. Technically speaking, recommender systems provide recommendations in three basic ways: They

recommend content that was previously selected by a user or other (to some extent similar) users (*collaborative filtering*). They recommend content based on similarities of properties and characteristics between previously chosen and available content (*content-based filtering*). Or they combine both approaches (*hybrid recommender systems*; Burke, 2002). Nguyen, Hui, Harper, Terveen, and Konstan (2014) found mixed results when they analysed the effects of a movie rating page's collaborative filtering-based recommender system on its users' range of interests. The users' average movie diversity decreased over time, but the effect was stronger for those users that did not usually follow the recommendations than for those who did frequently click the recommended links. In a study on the effects of a music platform's recommender system, Hosanagar and colleagues found little empirical evidence for fragmentation over time (Hosanagar, Fleder, Lee, & Buja, 2013). The development of recommendation strategies that counter possible filter bubble or echo chamber effects has become a topic of interest for software engineers in the last several years (e.g., Resnick, Garrett, Kriplean, Munson, & Stroud, 2013). Overall, there seems to be no common interpretation of the available evidence in the research community as to whether technological features such as recommender systems or many-to-many communication patterns in social media facilitate or attenuate the emergence of filter bubbles and echo chambers. Therefore, the present study aims to shed light on their emergence by empirically testing the joint effects of three levels of filters in an ABM.[1]

The triple-filter-bubble model

We refer to *filters* in a very general way as processes that lead to a limitation of information that is available to individuals. In our models, we will take into account filtering processes on three different levels: The individual, the social, and the technological level (Geschke, 2017).

Individual filters

The first group of filters – cognitive motivational processes – has been studied extensively in cognitive and social psychology. As a means of confirming pre-existing attitudes (Nickerson, 1998), verifying their self-views (Swann, Pelham, & Krull, 1989), avoiding cognitive dissonance (Festinger, 1957), and boosting social identity (Brewer, 1991), individuals are to different extents cognitively motivated to search for and add fitting bits of information and to ignore or deny conflicting ones. Similar effects have also been studied under the term *confirmation bias* (Jonas, Schulz-Hardt, Frey, & Thelen, 2001; Knobloch-Westerwick, Mothes, & Polavin, 2017).

In all these cases, filtering refers to selective exposure due to an individual's information search, processing, and memory. Curiosity may, however, motivate individuals to have a preference for consuming information that is at least to some degree novel and surprising (Loewenstein, 1994).

Social filters

Human beings display a tendency to form friendships and other social network structures preferably with people with whom they share 'sociodemographic, behavioral, and intrapersonal characteristics' (McPherson *et al.*, 2001, p. 415). In social media

---

[1] *The authors would like to kindly thank Alex Haslam for proposing to empirically test the triple-filter-bubble model by using the method of ABM.*

communities as well, processes of self-categorization (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987) contribute to the formation of communities with a shared social identity (Ridings & Gefen, 2004). Furthermore, social media users frequently *unfriend* acquaintances holding different views on conflictual topics (John & Dvir-Gvirsman, 2015).

In the age of social media, information is often passed along such online networks (Bakshy *et al.*, 2012). Hence, homogeneous network structures can potentially limit the width of information to which a social media user is exposed. The tendency for homophily is even stronger among social media users holding conservative or more extreme views (Boutyline & Willer, 2017). Hence, in particular in certain milieus social homophily can be a strong contributing factor to the emergence of filter bubbles and echo chambers and consequently group polarization effects (Sunstein, 2018; Vinokur & Burnstein, 1978).

Technological filters

The third group of filters – *algorithms* – operates on the technological level: Online media providers, such as Google or Facebook, compete for user attention. Therefore, they filter the provided information according to individual users' assumed wants and needs, leading to individually selected media offers (Pariser, 2011). The goal of this filtering is to maximize the time users spend on their respective sites, in order to maximize profits generated through advertising. To accomplish this, companies use proprietary, non-transparent automatic algorithms. In effect, this leads to different information offers tailored to the individual. For instance, none of us gets the same output on any given Google search; instead, each user gets an individualized selection of information. We assume that stronger automatic filtering leads to a decreased variety of information that is offered to individuals. This eventually leads to a decreased spectrum of attitudes that are cognitively available and salient in individuals and, thus, to smaller filter bubbles.

However, these recommender systems also constantly confront the user with novel not yet consumed information to maximize click-through rates, thereby potentially increasing the exposure to different points of view (Herlocker, Konstan, Terveen, & Riedl, 2004). Therefore, an alternative assumption is that, in spite of the filtering processes mentioned above, online media increase the spectrum of attitudes that are cognitively available and salient in individuals.

In sum, the filters on these three levels are expected to influence how much of the abundant information is cognitively available to individuals. More importantly, this influence is not random, but systematic: Information is more likely sought, delivered, or perceived when it fits the individual's needs as determined by individual characteristics, and this is partly gauged through automatic recommender systems. Additionally, the outcome depends as well on attitudinal characteristics of the peer groups that individuals interact with. For the present study, we created a parsimonious ABM to simulate these different processes.

Research questions

We ran several simulations using the aforementioned parsimonious ABM to gauge the effects of different ways of information propagation (see the paragraph 'Simulation scenarios' below) on the formation of echo chambers and filter bubbles. We also implemented rudimentary social structures in the form of friendship networks analogue to online social media platforms such as Facebook. We wanted to find out which combinations of these social and technological 'filters' facilitate and attenuate the

emergence of echo chambers and filter bubbles. Details of our model, the different scenarios, and the outcome variables (see the paragraph 'Possible outcomes' below) are explained in the section below.

Methods

Model synopsis

We designed a dynamic ABM where several individuals (together representing a society) position themselves in a two-dimensional attitude space based on attitudinal bits of information they hold in memory. In the model, individuals repeatedly receive new information with differing attitudinal messages from different sources. The sources of new information represent the technological filters and can be (1) individual discovery, (2) central announcement, for example through mass media, or personalized recommendations, for example through online media providers, that either (3) fit or (4) challenge the attitudes of the individual. Further on, individuals may also receive posted information from their friends through their social network, when a social media channel is provided. The friendship network thus provides a social filter for the individual. Individuals integrate the information they receive through cognitive filters: They integrate a particular bit of information more likely when the distance of its attitudinal message to their own attitude is below a latitude of acceptance. This means that it is unlikely that they integrate information that does not fit their pre-existing average attitudes. The concept of the *latitude of acceptance* goes back to Sherif and Hovland (1961). In the social simulation literature, the concept is known as *bounded confidence* (Hegselmann & Krause, 2002). Individuals have a limited memory and can only integrate a certain amount of information. When their memory is full, they have to forget bits of information to integrate new ones. These processes lead to repositioning of individuals in the attitudinal space according to the average information they consequently hold in their memory. We have implemented the model in *NetLogo* (Wilensky, 1999). In the following, we lay out the model following the procedure proposed by Jackson, Rand, Lewis, Norton, and Gray (2017) adapted to our model. In the *Info Tab* of the model (Lorenz, Holtz, & Geschke, 2018), we provide a description which follows the implementation code closely. The model can be downloaded here: https://doi.org/10.5281/zenodo.1407733.

Individuals and news items in attitude space

The model has two types of agents: individuals and bits of information. Both live in a two-dimensional world that represents an attitude space. Thus, the position of an information bit represents its attitudinal content on two dimensions, and the position of an individual represents the individual's attitudes on the same dimensions. Attitudes range from −1 to +1 on both dimensions. A dimension can, for example, represent political attitudes on the economic left-right and the societal progressive-traditional axes, or valence attitudes on two issues. Individuals are connected among each other through bidirectional friendship links. These links represent a friendship graph as it exists in social media platforms. Individuals also connect to bits of information. Such info-links represent that the individual integrated this bit of information in memory. Consequently, individuals can also be connected as sharers of information when they hold at least one common bit of information in their memories. The attitude space and all types of agents and links are outlined in Figure 1.

Agents' activities

In every time step, new bits of information appear in attitude space at random positions. They come to the attention of individuals, who try to integrate them. Further on, individuals might also post a randomly selected bit of information they hold in memory to the attention of their friends. For example, in Figure 1, the red individual might additionally receive one bit of the information from its friend (linked in yellow).

Whenever a new bit of information comes to the attention of an individual, it tries to integrate it by creating an info-link. We model the integration of information as a probabilistic event. The integration probability is a function of the attitude distance between the individual and the information. We use the following functional formula:

$$P(d; D, \delta) = \frac{D^{\delta}}{d^{\delta} + D^{\delta}} \tag{1}$$

where d is the attitude distance, D the latitude of acceptance, and δ a sharpness parameter that specifies how steep the integration probability drops from one to zero around the latitude of acceptance. The probability of integration decreases with d (cf. Abelson, 1964; Fishbein & Ajzen, 1975; Fisher & Lubin, 1958). This means that information fitting the individual's average pre-existing attitudes is more likely to be integrated. While the concept of the *latitude of acceptance* goes back to Sherif and Hovland (1961), the functional form is taken from the formalization of the *Social Judgment Theory* by Hunter, Danes, and Cohen (1984). However, they only dealt with the case of $\delta = 2$ and did not take into consideration a sharper decline around the latitude of acceptance. The limit in the case of very large δ coincides with the bounded confidence model (Deffuant, Neau, Amblard, & Weisbuch, 2000; Krause, 2000). The integration probability is 0.5 when the distance is equal to the latitude of acceptance ($d = D$). At the limit of very large sharpness parameters δ, integration becomes deterministic. In this case, info-bits are integrated with certainty if the distance is less than D, and are rejected otherwise. The functional form is shown in Figure 2.
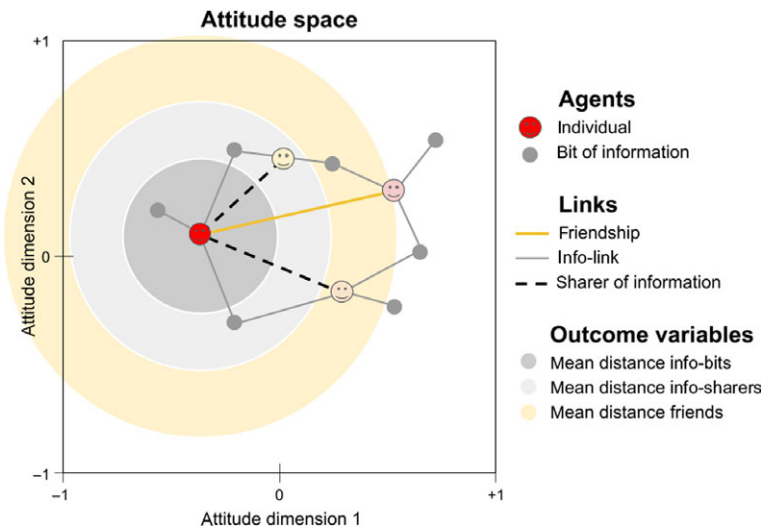

Figure 1. Conceptual figure showing the attitude space and all types of agents and links of the model, as well as the three central outcome measures. [Colour figure can be viewed at wileyonlinelibrary.com]

Individuals can only maintain a limited number of info-links due to their limited memory. Therefore, they drop a random info-link when necessary to integrate a new bit of information. Whenever an individual integrates a new bit of information, it adjusts its own attitudes. On both attitude dimensions, the individual sets the attitude to the average attitude of all the bits of information it holds in memory, following Anderson's (1971) *integration theory*. Thus, the integration of a new bit of information pulls the individual a bit towards attitudinal values communicated in it. Conceptual Figure 1 shows the individuals at the attitudinal barycentre of the information they hold in memory.

Individual may also change their friends, typically at a much lower rate. If a friendship is up to be potentially dropped, the probability to keep the friendship depends on the attitude distance to the friend analogue to the integration probability as defined in Equation (1). When a friendship is dropped, the individual selects a new friend at random from the friends of its friends.

Initial conditions and simulation time

In all simulations discussed in the following, we initialized 500 individuals with random positions in attitude space (uniformly distributed in both dimensions) and no bits of information initially.

As initial friendship networks, we created networks where each individual has on average twenty friends. For the network generation, we assigned each individual to one of four pre-defined equally sized groups and made a random network such that for each individual on average 80% of its friends come from the same group and 20% from the other three groups. We used this group structure network to check whether friendship groups have an impact on evolving echo chambers. As robustness check, we also checked random networks without group structure and small world networks of the Watts–
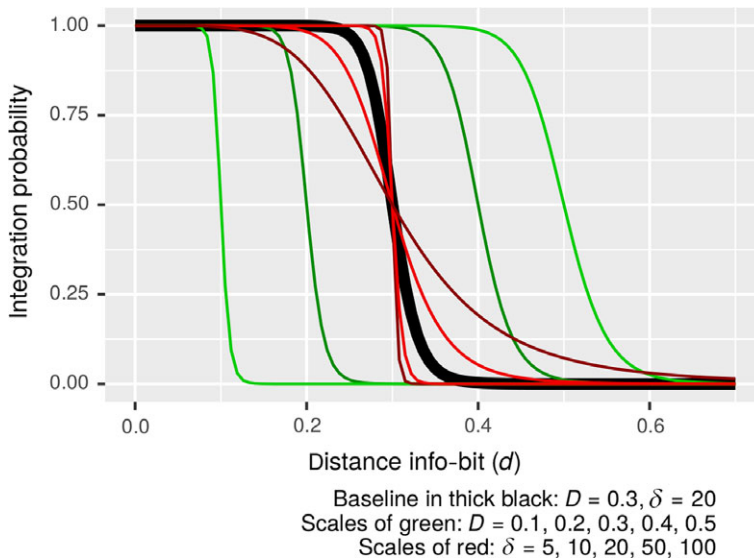

Figure 2. The functional form of the integration probability $P(d) = D^{\delta}/(d^{\delta} + D^{\delta})$ from Equation (1). The thick black line marks the baseline case ($D = 0.3$, $\delta = 20$) used for the simulation results in most of the following figures. Green lines mark other latitudes of acceptance D, and red lines other sharpness parameters δ. [Colour figure can be viewed at wileyonlinelibrary.com]

Strogatz type, with a 20% fraction of long-distance links (Watts & Strogatz, 1998). The effects presented in the following appear essentially identical for all these types of networks.

Each simulation lasted for 10,000 time steps; this ensures that a metastable configuration is reached in all configurations we simulated. In each time step, each agent is independently of others exposed to one new bit of information. Additionally, it might be exposed to more bits of information – one posted from each of its friends. The 10,000 time steps would represent 3 years where each individual is exposed to about nine new bits of information per day and – with twenty posting friends – about 180 posted bits of information per day. Note that these are the bits of information the individual is exposed to, not necessarily the number of bits it integrates each day.

Possible outcomes of our simulations

For every individual, we computed the mean attitude distance to all bits of information in memory, the mean attitude distance to all sharers of information, and the mean attitude distance to all friends. All these distances are shown in Figure 1 for the red individual. The average of these three outcome variables over all individuals after 10,000 time steps characterizes how the society they form evolved. Let us consider three prototypical cases.

1. The distance to bits of information is smaller than the distance to sharers of information which is smaller than the distance to friends. In this society, individuals share information with others, some of whom hold very different attitudes, while some of their friends hold even more different attitudes. The posted information they received from these friends is, of course, usually not integrated, since it is beyond their latitude of acceptance.

2. The distance to sharers of information is smaller than the distance to bits of information which is smaller than the distance to friends. In this society, individuals do share information only with others who hold very similar attitudes. Thus, communities of information sharers usually have their bits of information exclusively within this society. Individuals have friends in other communities of info-sharers, but usually do not integrate their posted bits of information.

3. The distance to friends is smaller than the distance to sharers of information which is smaller than the distance to bits of information. In this society, all the friends as well as all information sharers of individuals have very close attitudes. As the bits of information they hold in individual memories are more diverse than in their social surroundings, they may even have the perception of living in a attitude-diverse information environment, while at the same time there may exist another community of the same type with drastically different attitudes, representing strongly disconnected echo chambers.

Simulation scenarios

In the simulation analysis, we were interested to see which of the three societies (1)–(3) evolves under different modes for the generation of new information, and different characteristics of cognitive, social, and technological filters. To that end, we set up twelve scenarios, whose configurations and outcomes are summarized in Table 1. The three output measures in the table are direct computations after one simulation run. We extensively tested that these reproduce with minor variation in other simulation runs, which can be tested (Lorenz *et al.*, 2018). In that sense, the numbers are generalizable

Table 1. Summary of the scenario configurations, quantitative output measures, and qualitative society characteristics of Scenarios 1 to 12. Output measures were computed based on the results of one simulation run

Fig.	#	Mode new information	Social posting	Latitude of acc.	Refriend prob.	Info-bits	Info-sharer	Friends	Type of society
						Scenario configuration			Output measures: mean distance to . . .
2	1	Individual	Off	0.3	0	0.182	0ᵃ	0.840	(2)
2	2	Central	Off	0.3	0	0.182	**0.285**	0.890	(1)
2	3	Individual	On	0.3	0	**0.150**	0.045	0.520	(2)
2	4	Central	On	0.3	0	**0.150**	0.044	0.730	(2)
3	5	Filter close	Off	0.3	0	**0.164**	0.075	0.593	(2)
3	6	Filter distant	Off	0.3	0	0.278	**0.295**	0.829	(1)
3	7	Filter close	On	0.3	0	**0.144**	0.042	0.707	(2)
3	8	Filter distant	On	0.3	0	**0.148**	0.042	0.692	(2)
4	9	Individual	On	0.3	0.01	**0.156**	0.041	0.038	(3)
4	10	Individual	On	0.3	1	**0.153**	0.031	0.029	(3)
5	11	Individual	On	0.5	0	**0.257**	0.075	0.070	(3)ᵇ
5	12	Central	Off	0.5	0	0.307	**0.456**	0.579	(1)

Notes. Output measures after 10^4 time steps.

Type of society: (1) info-bits < info-sharer < friends, (2) info-sharer < info-bits < friends, (3) friends < info-sharer < info-bits.

ᵃIn this scenario, shared bits of information do not exist.

ᵇA fundamental difference to Scenarios 9 and 10 is that 11 has is only one attitude community. The number of communities is not specified by the types of society.

Bold numbers mark the larger value of the Info-bits and Info-sharer output measure.

output measures for the corresponding input parameters, and the characteristics of the evolving society (1)–(3) can be inferred from the order of them.

We distinguished four modes for the generation of new bits of information. (1) In the *individual* mode, each individual receives one new individual bit of information and tries to integrate it. (2) In the *central* mode, all individuals try to integrate one central new bit of information. This represents mass media input from one central, unbiased channel (*mainstreaming*; Griffin, 2012). In the two remaining modes, (3) *select close info-bits* and (4) *select distant info-bits*, a new random info-bit is created and presented to each individual analogously to the individual mode until the total number of info-bits is equal to the number of individuals. When the number of info-bits is equal to the number of individuals, each individual is presented a random existing info-bit that is either inside (in the mode *select close info-bits*) or outside (in the mode *select distant info-bits*) of its latitude of acceptance. This represents a recommendation algorithm that aims to present info-bits that the receiver will integrate with a probability higher than 0.5 (*select close info-bits*) or, respectively, an info-bit that confronts the individual with very different information that is unlikely to be integrated. The latter two modes represent influences of different *technological* filters.

The concept of *social filters* is implemented through the possibility that agents post one of their bits of information (randomly selected) to all of their friends. This represents a social media mechanism. In Scenarios 1 to 8, we study the four modes of information generation once with social posting and once without social posting. Further on, Scenarios 9 and 10 show what happens when individuals sometimes drop friendships and add a new friend from their friends of friends. In Scenario 9, each friendship is up for dropping with a probability of 0.01, and in Scenario 10, each friendship is up for dropping with certainty. Whether a friendship up for dropping is kept or dropped is a probabilistic event analogue to the integration of information. Friends with very different average attitudes are dropped more likely. The evolution of the friendship networks in Scenarios 9 and 10 thus follows the same mechanism but at much faster pace in Scenario 10.

The concept of *cognitive filters* is always present through the mechanism that information is more likely integrated if it fits closely to pre-existing attitudes. This mechanism is active in all scenarios of the model, since such biases cannot be easily switched off. In most simulations, we used a latitude of acceptance of $D = 0.3$. In Scenarios 11 and 12, we tested which changes a larger latitude of acceptance $D = 0.5$ elicits. We always used a sharpness parameter $\delta = 20$. This strong sharpness implies that integration is very likely when the info-bit is at a distance smaller than five in attitude space and very unlikely otherwise. We used this comparatively high sharpness value because it leads to relatively stable outcome states and less fluctuation of different simulation runs under the same configuration. We made exploratory simulations and observed that the effects also emerge when the latitude of acceptance is less sharp. A detailed parameter study for the effects of D and δ is beyond the scope of this paper.

Results

Results for Scenarios 1 to 4

Figure 3 shows the positions of individuals in the attitude space and their info-links after stabilization. The info-bits are not shown in the figure because they would often cover the individuals. Their location can be assessed by the other empty end of the info-links. Friend links and info-sharer links are not shown for similar reasons. Nevertheless, they are part of the simulation. The quantitative characteristics are summarized in Table 1.
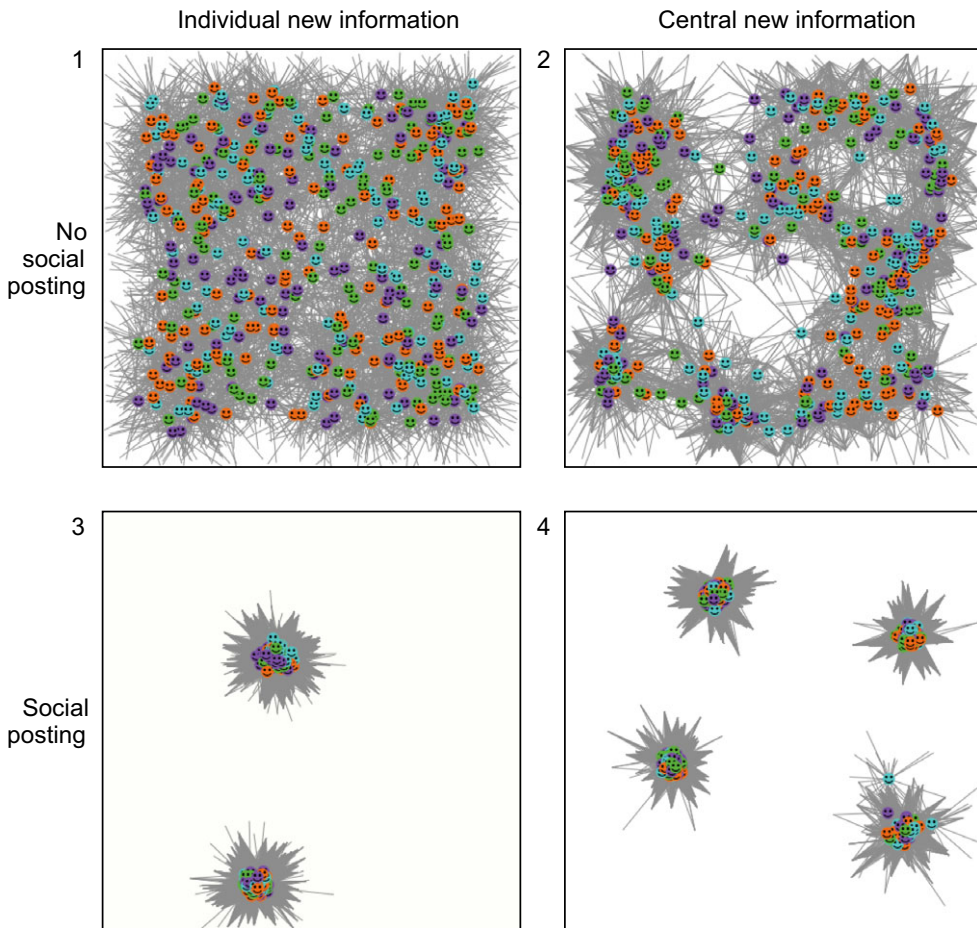
Figure 3. Individuals and their info-links after stabilization for Scenarios 1, 2, 3, and 4. The colour of individuals determines their group. On average, 80% of an individual's friends were of the same colour. [Colour figure can be viewed at wileyonlinelibrary.com]

In Scenarios 3 and 4 with social posting, the bubble of people with whom information was shared (as indicated by mean distance info-sharer) was smaller than the information bubble itself (as indicated by mean distance info-bits). This means that these individuals might have perceived strong attitude homogeneity with the people they shared information with and at the same time had the perception that they held diverse info-bits. This was different without social posting in Scenarios 1 and 2. Under the condition of pure individual info-bits (as in Scenario 1), there was no clustering and no info-sharing. Under the condition of central information propagation (Scenarios 2 and 4), there was some clustering, but the info-bridges between different clusters remained (individuals also shared information with individuals from other clusters). In our baseline case, the availability of social posting enforced strong clustering into sharers of information, who operated in slightly wider information bubbles with no informational contact to the other communities. These clusters evolved even though there was continual inflow and exposure to info-bits from the whole attitude space, because through social posting each individual became exposed to on average twenty bits of information from the social

network, in addition to the one new bit of information with a random position. The attitudinal position of the information from the social networks was not randomly distributed, but was, in each time step, based upon the current distribution of info-bits. This implies that a region in attitude space where the concentration of individuals is slightly higher by chance can self-reinforce through the propagation of the information these individuals post. In that way, they attract other individuals to move towards these regions. This is the mechanism how social posting together with the cognitive filter creates the pronounced and disconnected clusters. The square geometry of the attitude space and the level of the latitude of acceptance then determined how many such concentrated regions ultimately emerge and remain.

A second result was that the friend network had no effect on attitude clustering. The bubble of friends maintained its large attitude radius and clusters in no way self-sorted with respect to friendship communities. Figure 3 shows the communities of individuals by their colour. It is clearly visible that the info-sharer bubbles were composed of members of all four friendship communities. Actually, the formation of info-sharer bubbles would evolve as it would do with only one friendship community. Thus, attitude clustering was possible even though all individuals continuously received strongly differing info-bits from many of their friends, who held different average attitudes. They just did not integrate this information.

The impact of technological filters: Scenarios 5 to 8

We repeated the design above and analysed effects of the technological filter mechanisms for the exposure of individuals with new bits of information. The *select close info-bits* mode appears in Scenarios 5 and 7, and the *select distant info-bits* mode in Scenarios 6 and 8. Scenarios 5 and 6 are without social posting, and 7 and 8 with. Results are depicted in Figure 4, and their quantitative characteristics are summarized in Table 1.

The technological filter selecting close info-bits had an effect similar to social posting, even when posting was disabled. Individuals formed tight info-sharer bubbles with almost no shared info-bits between the bubbles. A filter selecting distant info-bits was able to sustain a fully connected info-sharer network. When social posting was switched on, the final outcome was very similar to the former scenarios (3 and 4) with social posting: Eight info-sharer bubbles evolved. The only difference is that it took much longer to reach stability when only distant info-bits were selected by the technological filter.

The impact of refriending: Scenarios 9 and 10

People not only map real-life friendship and family networks onto online social networks; they also defriend and refriend online contacts. The refriend mechanism in our model makes defriending more likely when the attitudinal distance is greater, while new friends are random friends of friends. In previous simulations, we had found that existing community structures in the friendship networks of information sharing with social posting led to info-sharer bubbles with people from all friendship communities (Scenarios 3 and 4). Scenarios 9 and 10 in Figure 5 show the situation after stabilization when new bits of information appear individually and with social posting enabled as in Scenario 3, but additionally with refriend probabilities of 0.01 and one.

This result implies that echo chambers evolve when refriending happens in addition to social posting, but refriending is not the driving force for the formation of disconnected clusters of individuals with similar attitudes.
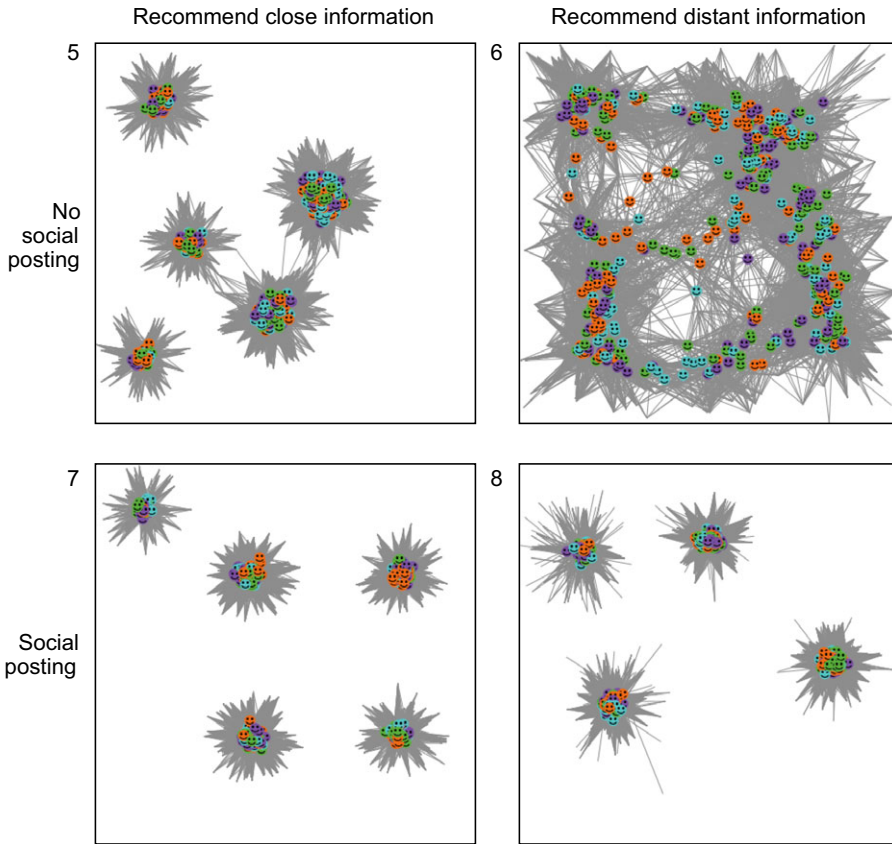
Recommend close information   Recommend distant information

Figure 4. Individuals and their info-links after stabilization for Scenarios 5, 6, 7, and 8. [Colour figure can be viewed at wileyonlinelibrary.com]

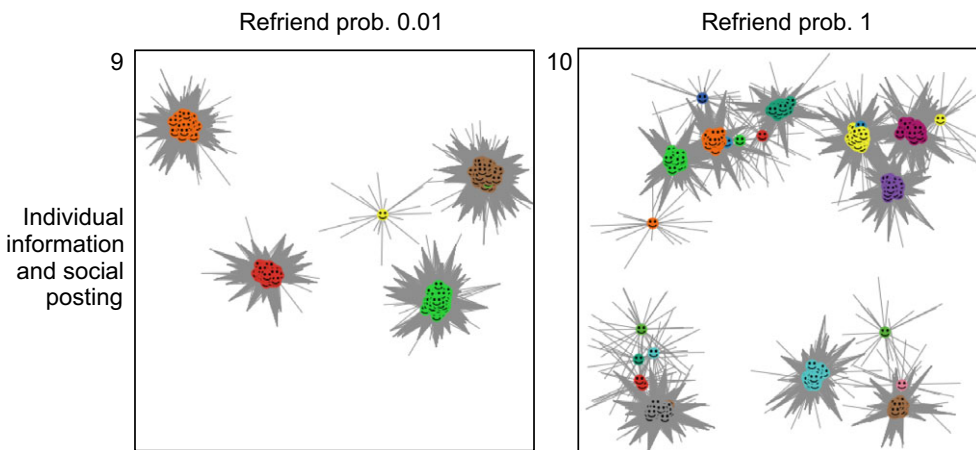Refriend prob. 0.01   Refriend prob. 1

Figure 5. Scenarios 9 and 10, individual new-info-mode with social posting and a refriend probability of 0.01 (Scenario 9) or 1 (Scenario 10). Colours represent connected components of the evolving friendship network. [Colour figure can be viewed at wileyonlinelibrary.com]

Larger latitudes of acceptance: Scenarios 11 and 12

We also tested the impact of larger latitudes of acceptance. In particular, we studied Scenario 2 (central new information without social posting) and Scenario 3 (new information individually with social posting) with a latitude of acceptance of 0.5 instead of 0.3. Figure 6 shows the results.

Interestingly, the larger latitude of acceptance led to a large consensual cluster with social posting, while much more diversity including some clustering remained with central information without social posting. This suggests that social media could also have the potential to bring about a societal consensus that would not happen without. On the other hand, as we saw before, social media could also cause strong cohesive clusters maintained without any shared information between clusters.

Scenario 11 also shows interesting transient dynamics until the large consensual cluster forms. This can be observed running the simulation (Lorenz *et al.*, 2018). In transient, two or three clusters evolve with only very little shared info-bits. Over time, the number of shared bits of information increases slowly and at some point a certain tipping point is reached and clusters converge rapidly to one cluster. The parameter constellation is thus prone to rapid social change. We conjecture that there are parameter configurations, which always lead to the same characteristic outcomes, while others are more prone to path dependence. A further exploration of this is beyond the scope of this paper, but proposed for future research.

Discussion

Main findings

Without central information propagation, social posting, or recommender systems, no echo chambers emerged (Scenario 1). Individuals spread out evenly in the attitude space (except for the extreme fringes). With central information propagation but without social posting or recommender systems, distinct echo chambers emerged, but individuals still shared some information with people outside their respective echo chambers (Scenario 2). Without central information propagation, but with social posting and without recommender systems, distinct echo chambers without links between them emerged
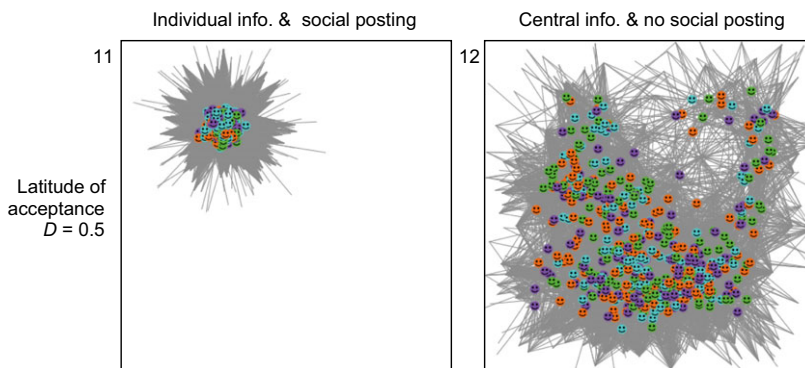

Figure 6. Scenarios with latitude of acceptance $D = 0.5$, which is higher than the baseline case of $D = 0.3$ used in all other scenarios. Left: individual new-info-mode with social posting (analogue to Scenario 3); right: central new-info-mode without social posting (analogue to Scenario 2). [Colour figure can be viewed at wileyonlinelibrary.com]

(Scenario 3). This indicates strong attitude group polarization. With central information propagation and social posting, but without recommender systems, distinct echo chambers emerged as well (Scenario 4).

Taken together, this shows that in our ABM filter bubbles and echo chambers evolved from individual cognitive processes (modelled in all scenarios) in combination with central news sources that reach almost everyone alone, even without any social (posting or refriending) or technological (recommender systems) processes involved (Scenario 2). If, however, additional social posting processes occurred (simulating many-to-many communication; Scenarios 3 and 4), these echo chambers became more distinct and less interconnected; this would lead to even more fragmentation and polarization of society. Thus, these scenarios resemble the supposed impact of social media on strongly polarizing, political events such as the election of Donald Trump as US President or the British people's decision on Brexit.

In Scenarios 5 to 8, recommender systems were used to present new information to individuals. We found that recommendation of close info-bits had the same effect as social posting even without social posting, while recommendation of distant info-bits could maintain a connected info-sharer network at least for some time. Social posting had the same effects as before.

The triple-filter-bubble framework

The triple-filter-bubble framework takes into account information filtering processes on the individual, the social, and the technological level. While the filtering mechanisms on the different levels had already been identified and described in previous research, their complex combination in a joint framework is novel. Results of our simulations show that the different filters interact and have effects on individual and social conceptual phenomena in ways that are not trivial: Disconnected echo chambers of individuals with almost identical attitudes based on information that nobody outside shares evolve through social posting and the cognitive filter given a relatively small latitude of acceptance. Interestingly, existing communities of friends that were initially not based upon shared attitudes did not have any substantial effect on attitude clustering. No self-sorting with respect to attitudes into existing friendship communities happened, and individuals typically ended up in different echo chambers from many or even most of their friends within the social network. With refriending, the clustering into echo chambers is just accelerated. Without social posting, bridges of information between such clusters of individuals could survive.

Our simulations yielded results that bear resemblance to phenomena that can be observed in real-life contexts as well, such as the emergence of homophilous social networks and, under some circumstances, the emergence of detached echo chamber formations. The complexity of the framework might seem like a disadvantage; however, colleagues who are interested in running simulations as means of testing more specific predictions, can build upon the face validity of our ABM and adjust it to their purposes. [The model can be downloaded at https://doi.org/10.5281/zenodo.1407733. To run the model, download the latest (free) version of NetLogo and open the downloaded file.] Of course, they are invited as well to validate or invalidate our initial findings.

Limitations and directions for future research

With regard to our ABM, the results presented here are not a complete analysis of the behaviour of the model. Next steps of interest departing from our baseline case would be,

for example, to study the effect of memory size or different settings of the latitude of acceptance. Potential sensitivity analyses should address the questions of whether the effects are similar in attitude spaces with one or multiple dimensions. We expect that increasing the number of dimensions would have a strong effect on the number of evolving info-sharer bubbles. Furthermore, other distributions for the random appearance of the bits of information could be tested. Additionally, a birth–death mechanism for individuals could be used to model the possibility of adding additional dynamics to the network by having individuals leave it or join it and to explore how robust our findings would be with such a turnover (Kurahashi-Nakamura, Mäs, & Lorenz, 2016).

Another direction for future research will be to compare our results with those of studies using actual behavioural data. Our Scenario study admittedly uses a relatively stylized configuration of isolated dynamical mechanisms of micro–macro interaction. Nevertheless, the set-up of the model matches well to digital behavioural data from social media platforms. In line with previous studies that have analysed opinion dynamics using 'big data' techniques (e.g., Bakshy *et al.*, 2015; Del Vicario *et al.*, 2016), the importance of the technical infrastructure and of the properties of the respective social networks on the emergence of filter bubbles and echo chambers can in principle also be studied using authentic data from social media users. For example, our simulations indicate that enabling serendipity in recommender systems (Herlocker *et al.*, 2004) may attenuate the emergence of filter bubbles and echo chambers (Scenarios 6 and 8). However, we would assume that in view of the sheer number of information that is shared by friends and acquaintances in social media, the effectiveness of such technological countermeasures could be only marginal (cf. Scenario 8). The first challenge for such studies would be to assess the relevant societal attitudinal space and to develop methods of measuring it.

One typical concern about ABM is that researchers might relatively easily create any desired model outcome by trying out different rules and settings until their model fits their theory. Jackson *et al.* (2017) and ourselves do not agree with this concern in its generality. To counter it, we provide all details including the simulation code and intuitively understandable buttons to reproduce our findings. A critical reader can thus check, refute, or validate the findings and their robustness and sensitivity with respect to additional or other theoretical assumptions about the behaviour of individuals.

Conclusion

Modern technology cannot be stopped; people like to share their experiences digitally, and tech giants will further professionalize recommender systems to maximize the time users spend on their sites. On an individual level, these processes may lead to reassurance and enhancement of individually existing attitudes, behaviours, and identities. They increase individual attitudinal stability, and, thus, individual certainty and security. On a societal level, however, these processes are prone to increase attitudinal differences between opinion groups and individuals and to cut communication ties between them, leading to attitude clusters, societal fragmentation, and polarization. This poses a problem for modern democracies that rely on conflict resolution and reaching consensus through processes of democratic discourse. For the democratic process, it is necessary to be able to hear people express different opinions, to be willing to listen to them, and to engage in mutual discussion. So the digital world presents a genuine dilemma, where *positive* individual effects go along with *negative* societal effects. What can be done?

Generally, remedies to these issues on the individual, social, technological, and societal level have been proposed: On the individual level, knowledge about the

processes leading to filter bubbles, or more generally, *media competence*, might mitigate these effects. On a social level, alternative mechanisms for debate, discussion, and creation of consensus are proposed (possibly using social media). On the technological level, means of increasing the serendipity of recommender systems are currently being discussed (Zhang, Séaghdha, Quercia, & Jambor, 2012) and will hopefully be implemented in the future. On the societal level, the deletion of *fake news* or unwanted content from the Internet, that is censorship (as recently enforced in Germany for private companies like Facebook and the other tech giants), or the institutionalization of the latter is proposed as solutions. However, since these measures limit the human right of free speech and damage free discourse, they may finally turn out to be more harmful than useful to a democratic society.

We believe that it is necessary to not only study the effects of interventions on these different phenomena separately. Such studies should be complemented by attempts to employ a methodology that allows for gauging possible non-linear effects of different combinations of these factors and that allows for the development of theories that integrate findings from as different fields as sociology, politology, computer engineering, and psychology into a common framework like our triple-filter-bubble model.

Acknowledgements

References

Abelson, R. P. (1964). Mathematical models of the distribution of attitudes under controversy. In L. R. Tucker (Ed.), *Contributions to mathematical psychology* (pp. 142–160). New York, NY: Holt, Rinehart & Winston.

Anderson, N. H. (1971). Integration theory and attitude change. *Psychological Review*, 78, 171–206. https://doi.org/10.1037/h0030834

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web* (pp. 519–528). ACM. https://doi.org/10.1145/2187836.2187907

Bazzan, A. L., & Klügl, F. (2014). A review on agent-based technology for traffic and transportation. *The Knowledge Engineering Review*, 29(3), 375–403. https://doi.org/10.1017/S0269888913000118

Boutyline, A., & Willer, R. (2017). The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political Psychology*, 38(3), 551–569. https://doi.org/10.1111/pops.12337

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*, 17(5), 475–482. https://doi.org/10.1177/0146167291175001

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331–370. https://doi.org/10.1023/A:1021240730564

Castells, M. (2010). *End of millennium: The information age: Economy, society, and culture*. Oxford, UK: Wiley-Blackwell.

Deffuant, G., Neau, D., Amblard, F., & Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems*, *3*, 87–98. https://doi.org/10.1142/S0219525900000078

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., . . . Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113* (3), 554–559. https://doi.org/10.1073/pnas.1517441113

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Fisher, S., & Lubin, A. (1958). Distance as a determinant of influence in a two-person serial interaction situation. *The Journal of Abnormal and Social Psychology, American Psychological Association*, *56*, 230–238. https://doi.org/10.1037/h0044609

Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., & Lorenz, J. (2017). Models of social influence: Towards the next frontiers. *Journal of Artificial Societies & Social Simulation*, *20*(4), 1–31. https://doi.org/10.18564/jasss.3521

Garimella, V., & Weber, I. (2017). *A long-term analysis of polarization on Twitter*. https://arxiv.org/pdf/1703.02769.pdf

Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication*, *14*(2), 265–285. https://doi.org/10.1111/j.1083-6101.2009.01440

Geschke, D. (2017). The RE-CO-KIT: A REality COnstruction KIT – technological, group dynamic, cognitive and motivational aspects. Poster presented at the *18th General Meeting of the European Association of Social Psychology*, Granada, Spain, July 5–8th, 2017.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380. https://doi.org/10.1086/225469

Griffin, E. (2012). *Communication communication communication* (pp. 366–377). New York, NY: McGraw-Hill.

Happer, C., & Philo, G. (2013). The role of the media in the construction of public belief and social change. *Journal of Social and Political Psychology*, *1*, 321–336. https://doi.org/10.5964/jspp.v1i1.96

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

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, *22*(1), 5–53. https://doi.org/10.1145/963770.963772

Hermida, A., Fletcher, F., Korell, D., & Logan, D. (2012). Share, like, recommend: Decoding the social media news consumer. *Journalism Studies*, *13*, 815–824. https://doi.org/10.1080/1461670X.2012.664430

Holtz, P., Wagner, W., & Sartawi, M. (2015). Discrimination and immigrant identity: Fundamentalist and secular Muslims facing the Swiss Minaret Ban. *Journal of the Social Sciences*, *43*(1), 9–29.

Hosanagar, K., Fleder, D., Lee, D., & Buja, A. (2013). Will the global village fracture into tribes? Recommender systems and their effects on consumer fragmentation. *Management Science*, *60* (4), 805–823. https://doi.org/10.1287/mnsc.2013.1808

Huang, Q., Parker, D. C., Filatova, T., & Sun, S. (2014). A review of urban residential choice models using agent-based modeling. *Environment and Planning B: Planning and Design*, *41*(4), 661–689. https://doi.org/10.1068/b120043p

Hunter, J. E., Danes, J. E., & Cohen, S. H. (1984). *Mathematical models of attitude change*. Cambridge, MA: Academic Press.

Jackson, J. C., Rand, D., Lewis, K., Norton, M. I., & Gray, K. (2017). Agent-based modeling: A guide for social psychologists. *Social Psychological and Personality Science*, *8*, 387–395. https://doi.org/10.1177/1948550617691100

John, N. A., & Dvir-Gvirsman, S. (2015). "I Don't Like You Any More": Facebook unfriending by Israelis during the Israel-Gaza Conflict of 2014. *Journal of Communication*, *65*(6), 953–974. https://doi.org/10.1111/jcom.12188

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, *80*(4), 557. https://doi.org/10.1037/0022-3514.80.4.557

Knobloch-Westerwick, S., Mothes, C., & Polavin, N. (2017). Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Communication Research*, 0093650217719596 (online first). https://doi.org/10.1177/0093650217719596

Krause, U. (2000). A discrete nonlinear and non-autonomous model of consensus formation. In S. Elaydi, G. Ladas, J. Popenda & J. Rakowski (Eds.), *Communications in Difference Equations* (pp. 227–236). Amsterdam, The Netherlands: Gordon and Breach. https://doi.org/10.1201/b16999

Kurahashi-Nakamura, T., Mäs, M., & Lorenz, J. (2016). Robust clustering in generalized bounded confidence models. *Journal of Artificial Societies and Social Simulation*, *19*(4), 7.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*(1), 75–98. https://doi.org/10.1037/0033-2909.116.1.75

Lorenz, J., Holtz, P., & Geschke, D. (2018). *janlorenz/TripleFilterBubble: TripleFilterBubble NetLogo Model (Version v1.0)*. Zenodo. https://doi.org/10.5281/zenodo.1407733

Mason, W., Conrey, F., & Smith, E. (2007). Situating social influence processes: Dynamic, multidirectional flows of influence within social networks. *Personality and Social Psychology Review*, *11*, 279–300. https://doi.org/10.1177/1088868307301032

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444. https://doi.org/10.3410/f.725356294.793504070

Nguyen, T. T., Hui, P. M., Harper, F. M., Terveen, L., & Konstan, J. A. (2014). Exploring the filter bubble: The effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*. ACM, pp. 677–686. https://doi.org/10.1145/2566486.2568012

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Noor, P. (2017). 'The fact that we have access to so many different opinions is driving us to believe that we're in information bubbles': Poppy Noor meets Michal Kosinski, psychologist, data scientist and Professor at Stanford University. *The Psychologist*, *30*, 44–47.

Ott, B. L. (2017). The age of Twitter: Donald J. Trump and the politics of debasement. *Critical Studies in Media Communication*, *34*(1), 59–68. https://doi.org/10.1080/15295036.2016.1266686

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. London, UK: Penguin UK.

Resnick, P., Garrett, R. K., Kriplean, T., Munson, S. A., & Stroud, N. J. (2013). Bursting your (filter) bubble: Strategies for promoting diverse exposure. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*. ACM, pp. 95–100. https://doi.org/10.1145/2441955.2441981

Ridings, C. M., & Gefen, D. (2004). Virtual community attraction: Why people hang out online. *Journal of Computer-Mediated Communication*, *10*(1), JCMC10110. https://doi.org/10.1111/j.1083-6101.2004.tb00229.x

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*(2), 143–186. https://doi.org/10.1080/0022250X.1971.9989794

Sherif, M., & Hovland, C. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change*. New Haven, CT: Yale University Press.

Smith, E. R., & Conrey, F. R. (2007). Agent-based modeling: A new approach for theory building in social psychology. *Personality and Social Psychology Review*, *11*, 87–104. https://doi.org/10.1177/1088868306294789

Sunstein, C. R. (2001). *Echo chambers: Bush v. Gore, impeachment, and beyond*. Princeton, NJ: Princeton University Press.

Sunstein, C. R. (2009). *Republic.com 2.0*. Princeton, NJ: Princeton University Press.

Sunstein, C. R. (2018). *# Republic: Divided democracy in the age of social media*. Princeton, NJ: Princeton University Press.

Swann, Jr., W. B., Pelham, B. W., & Krull, D. S. (1989). Agreeable fancy or disagreeable truth? Reconciling self-enhancement and self-verification. *Journal of Personality and Social Psychology*, *57*(5), 782. https://doi.org/10.1037//0022-3514.57.5.782

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, UK: Basil Blackwell.

Vaccari, C., Valeriani, A., Barberá, P., Jost, J. T., Nagler, J., & Tucker, J. A. (2016). Of echo chambers and contrarian clubs: Exposure to political disagreement among German and Italian users of twitter. *Social Media+ Society*, *2*(3), 2056305116664221. https://doi.org/10.1177/2056305116664221

Vallone, R. P., Ross, L., & Lepper, M. R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology*, *49*(3), 577. https://doi.org/10.1037/0022-3514.49.3.577

Vinokur, A., & Burnstein, E. (1978). Novel argumentation and attitude change: The case of polarization following group discussion. *European Journal of Social Psychology*, *8*(3), 335–348. https://doi.org/10.1002/(ISSN)1099-0992

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393* (6684), 440–442. https://doi.org/10.1038/30918

Weiman, G. (2006). *Terror on the internet*. The New Arena, the New Challenges, 147–171.

Wilensky, U. (1999). *NetLogo*. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Zhang, Y. C., Séaghdha, D. Ó., Quercia, D., & Jambor, T. (2012). Auralist: Introducing serendipity into music recommendation. In *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 13–22. https://doi.org/10.1145/2124295.2124300