The complex link between filter bubbles and opinion polarization

Marijn Keijzer[a] and Michael Mäs[a,1]
[a] *University of Groningen, Department of Sociology, ICS*

Abstract. There is public and scholarly debate about the effects of personalized recommender systems implemented in online social networks, online markets, and search engines. On the one hand, it has been warned that personalization algorithms generate homogenous information diets that tend to confirm previously held attitudes and beliefs. Opinionated social media posts, shared news items, and online discussion could fragment social groups, alienate users with different political views, and ultimately foster opinion polarization. On the other hand, critics of this "personalization-polarization hypothesis" argue that the effects of personalization algorithms on information diets are too weak to have meaningful effects. Here, we argue that contributions to both sides of the debate fail to consider the complexity that arises when large numbers of interdependent Internet users interact and exert influence on one another in algorithmically governed communication systems. Reviewing insights from the literature of opinion dynamics in social networks, we demonstrate that opinion dynamics can be critically influenced by mechanisms active on three levels of analysis: the individual, local, and global level. We show which theoretical and empirical research on these three levels is needed to answer the question whether personalization fosters polarization or not, advocating an approach that combines rigorous theoretical modeling with the emergent field of data science.

Keywords. Personalization, recommender systems, opinion polarizations, filter bubbles, complexity, opinion dynamics, social networks

1. Introduction

Political events such as the Brexit referendum, the election of Donald Trump, and the success of other populist politicians in democratic elections have sparked an intensive public and scholarly discussion about the effects of online communication technology on public debate and collective decision-making. One of the most prominent warnings is that personalization algorithms installed in online social networks, search engines, and online stores contribute to the formation of so-called "filter bubbles" [1]. These bubbles create echo chambers, isolating users from information that might challenge their views and exposing them to online content that is in line with their views, and, thus, reinforces their opinions. Experts, pundits, and scholars have warned that this contributes to opinion polarization, a dynamic where competing political camps develop increasingly opposing political views. Public attention is enormous. Newspapers regularly cover the topic [e.g.

[1] Corresponding Author, Department of Sociology, University of Groningen. Grote Rozenstraat 31, 9712 TG Groningen, The Netherlands; E-mail: m.maes@rug.nl.

2,3]; leading politicians echo the warning [4,5]; and various initiatives have been undertaken to fight filter bubbles and polarization [6]. Here, we summarize the key arguments underlying the hypothesis that personalization algorithms contribute to opinion polarization and reflect on existing scientific research. While we echo the warning that personalization might have serious effects on societal processes, we also point to gaps in the theoretical and empirical literature that need to be filled before one can draw conclusions about whether or not personalization is indeed responsible for increasing polarization. Unlike other recent contributions [7], we do not argue that personalization is an innocent technology, but conclude that experts, politicians, and also scientists leap to conclusions when they propose that personalization is responsible for increased polarization. Accordingly, we call for more research on communication in online environments, pointing to the potential of approaches that combine theoretical modeling with the emerging field of data science.

Our analysis is inspired by the complexity approach [8–10] and builds on a rich literature in the field of opinion dynamics in social networks. This work departed in the 1950s in the social sciences and today profits from contributions from disciplines as diverse as physics, computer science, mathematics, economics, philosophy, sociology, political science, and complexity research [11–13]. In this literature, formal models of social networks have been developed, where network nodes exert social influence on the opinions of their contacts. These models allow one to understand the rich and intricate opinion dynamics that arise from social influence and to identify the conditions under which repeated social influence fosters the formation of opinion consensus, the fragmentation of the network into multiple clusters with competing opinions, or even opinion polarization. Decades of modeling work with analytical and computational methods have demonstrated that even seemingly innocent changes in models' assumptions can have profound effects on the outcomes of social influence processes, which shows that drawing conclusions about real complex systems, such as online communication systems, requires a formal model that is informed by detailed empirical research. This model is not available, to date, as we show here.

In a nutshell, we argue that the current public and scholarly debate about the personalization-polarization hypothesis has been paying too little attention to two important aspects. First, many contributions do not acknowledge the complexity of online social networks arising from repeated social influence between users. Complexity arises when a system consists of multiple micro-entities (users) that do not act in isolation but exert influence on each other [8,10]. In online social networks millions of users with a large number connections communicate with weak constraints on time and space, making these systems a very typical example of a complex system. Interdependency between users can generate chains of reaction such that even rare idiosyncratic events can have profound impact on the system as a whole [14,15]. So far, most contributions to the public and scholarly debate about the personalization-polarization-hypothesis are based on informal theoretical arguments and anecdotal evidence, and thus fail to address system complexity. We do not argue that the conclusions drawn from these contributions are necessarily false, but we discuss findings from complexity research that demonstrate how conclusions can change when a system's complexity is considered.

Second, we argue that contributions to the current debate tend to lean heavily on empirical and theoretical research on communication in offline worlds. We review insights from the opinion-dynamics literature showing that there may be differences between online and offline interaction that can critically alter opinion dynamics. In particular, we distinguish three levels of communication networks on which these differences typically reside: the individual level, the local level, and the global level.

The remainder is organized as follows. In the following section, we summarize the central theoretical, empirical, and political arguments underlying the scholarly and public debate about the effects of personalization on polarization. Next, we identify gaps in these debates, reviewing findings from the literature on opinion dynamics in social networks. In the concluding section, we sketch an agenda for future research, advocating an approach to data science that combines empirical research with rigorous theoretical modeling.

2. The debate about the personalization-polarization hypothesis

Personalization is ubiquitous on the Internet. Providers of Internet services seek to tailor their products to the needs and interests of individual users. Search engines, for instance, rank the results of users' search queries according to the interests of the individual user. When the authors of the present article google the term "polarization", for example, websites discussing political polarization should be ranked higher than websites of manufacturers selling "polarized" sunglasses, even though both websites contain the search term. Likewise, online markets recommend products based on the purchases of other customers who bought similar products in the past and online social networks sort incoming messages according to the similarity between the user and the source of the message. Personalization has tremendously improved online companies' services, making it easier for users to navigate the immense and rapidly growing amount of online content. Personalization has also turned into a multibillion-dollar business area, increasing engagement on online platforms using this technology, and allowing advertisers to directly target potential customers.

Despite these immense technological advances, there is growing concern about unintended negative consequences of web personalization. For many users, the Internet is an important source for information on political, social, and cultural topics [16]. Criticizing personalization in this context, observers of the Internet warned that users are less exposed to content that challenges their own political opinions. Being insulated from competing views, you get "stuck in a static, ever-narrowing version of yourself – an endless you-loop" [1]. Users of online social networks complained that their online communities have turned into cocoons consisting exclusively of likeminded friends, which makes online communication increasingly boring [1].

Scholars have echoed this concern, adding that personalization also intensifies processes of opinion polarization, the development of antagonistic groups, where opinion differences between groups intensify and positions between the two extremes of an opinion spectrum are increasingly sparsely occupied [17,18]. Personalization algorithms increase homophily, the degree to which users communicate with others who share similar views [19,20]. Research has shown that on Facebook personalization increases the degree to which Internet users are exposed other users who hold similar political opinions [21]. Informed by social-psychological research [22,23], it has been further proposed that this can intensify users' opinions, as they are mainly exposed to online content containing persuasive information that reinforces their initial opinions. As opinions of users form the left end of the political spectrum grow more leftist and users identifying with rightist political views grow more extreme, opinion differences between the political camps increases. Here, we refer to this conjecture as the *personalization-polarization hypothesis*.

The warning that personalization fosters polarization needs to be taken seriously, as opinion polarization has been argued to endanger societal cohesion [22,24–27] or cause cultural conflicts [28,29]. Opinion polarization might also pose challenges for political decision making in general [30] as it impedes political consensus formation also on otherwise non-controversial issues [28,29].

Political decision makers have echoed the warnings. Very prominently, Barack Obama warned in his farewell address that "for too many of us, it's become safer to retreat into our own bubbles, whether in our neighborhoods or on college campuses, or places of worship, or especially our social media feeds, surrounded by people who look like us and share the same political outlook and never challenge our assumptions. [..] And increasingly, we become so secure in our bubbles that we start accepting only information, whether it is true or not, that fits our opinions, instead of basing our opinions on the evidence that is out there." [4] Frank-Walter Steinmeier, Germany's president, took this argument even further, linking personalization with adverse societal outcomes. In his 2018 Christmas message, he argued that "more and more people are sticking with their own kind, living in self-made bubbles where everyone always agrees one hundred percent […]. What happens when societies drift apart, and when one side can barely talk to the other without it turning into an all-out argument, is all too evident in the world around us. We have seen burning barricades in Paris, deep political rifts in the United States and anxiety in the United Kingdom ahead of Brexit. Europe is being put to the test in Hungary, Italy and other places" [5].

What is more, there are already initiatives to break filter bubbles. Software developers, for instance, proposed novel personalization algorithms ranking higher content that challenges the opinions of the user [31,32]. In addition, Bozdag and Van den Hoven [6] distinguish two types technological solutions: those that make the user aware of their own bias, and those that show the users the opinion diversity for a given topic. The first type includes online tools that help users quantify and visualize the degree to which their news consumption is biased. Awareness of the composition of their information diet should then make users more open to other views. Second, there are electronic tools seeking to make users aware of the existing opinion diversity that they may overlook from the limited perspective of their bubble. Some of these tools use questionnaires to plot opinion distributions or allow users to list and share pro and con arguments they consider relevant for given issues. Other tools alert users when they visit a website that has been disputed on the web. Other initiatives seek to foster offline discussion between individuals with opposing views. In multiple national and international events, *mycountrytalks.org* motivated thousands of participants to first indicate their political opinions in online surveys to be then electronically matched for face-to-face discussion with users holding maximally opposite opinions.

While the public debate about the link between personalization and polarization is mainly based on anecdotal evidence, also outcomes of scientific research echoed the warnings. First, modelers of social-influence processes in networks have developed formal models mimicking communication on the web, showing that the theoretical reasoning underlying the personalization-polarization hypothesis is logically valid [17,18,33]. These models assume that individuals adjust their political opinions as a result of communication with network contacts. When two agents hold similar opinions, their opinions are reinforced because they provide each other with new persuasive arguments supporting their views. In line with the informal reasoning underlying the personalization-polarization hypothesis, these models show that opinion polarization is more likely to emerge when agents are mainly communicating with likeminded individuals. Recent modeling work based on alternative assumptions about communication found similar dynamics [34,35].

Second, researchers have collected ample empirical evidence for the central assumptions underlying the formal models. There is a rich empirical literature documenting that humans have a strong tendency to interact with similar others [20,36] and to selectively consume media that supports their own political views [37–39]. In search of evidence for the existence of echo chambers on the web, these tendencies have been observed in online settings too [21,40–44]. Online social networking platforms further promote homophilic interactions through personalization algorithms [21]. There is also strong empirical evidence for the second critical model assumption: opinion reinforcement by communication with likeminded individuals [18,22,45–47]. Recently, empirical research in online contexts also supported this assumption [48].

There is, however, also considerable skepticism about the personalization-polarization hypothesis. In an interview with the New York Times, Mark Zuckerberg, the CEO of Facebook, responded that it is a "good-sounding theory, and I get why people repeat it, but it's not true" [49]. More importantly, however, there is also empirical evidence that might challenge the personalization-polarization hypothesis. For instance, analyzing users' browser histories, researchers found that a large part of online news is still being consumed on news websites that do not filter content on the personal level, which should temper the effects of personalization of other web services [50]. Some scholars even argue that "social media usage […] reduces political polarization" [51]. Barabera's analyses, for instance, suggest that most Twitter users are still exposed to diverse content and that exposure to diverse content fosters moderate rather than polarized opinions. Similar observations led Axel Bruns to conclude that even if personalization did foster the creating of filter bubbles, the "the disconnect […] is too mild to create any deleterious effect" [7].

Likewise, empirical research on the collective level has not yet painted a clear picture. On the one hand, research has documented that opinion distributions have polarized in many western countries since the Internet has become a dominating communication platform [27,52–54]. On the other hand, it is debated whether the Internet is actually responsible for this trend. One could argue that the more time users spend on the Internet the easier it is for them to escape their filter bubbles. A Facebook user, for instance, who does not only read the top-ranked messages of her news feed will also be exposed to online content challenging her views. In fact, a prominent study found that opinions amongst young people – the demographic subgroup that spends most time on the Internet and in social networks – are the least polarized of all age cohorts [53].

In sum, the personalization-polarization hypothesis has received a lot of attention but research has so far not been able to provide conclusive evidence supporting or falsifying it. In the following section, we reflect on reasons why studying this hypothesis is challenging, pointing to aspects of online communication that are highly complex but hardly understood.

3. The complexity perspective on the personalization-polarization hypothesis

Answering the question whether personalization technology fosters polarization is an ideal-typical research problem requiring a complexity perspective, as it is concerned with the two defining ingredients of complexity. First, a complex system consists by definition of multiple levels of analysis [8,9]. In the case of the personalization-polarization hypothesis, there is the level of the individual user who consumes, shares, adjusts, and generates content; and there is the collective level, the Internet. Both personalization and polarization are collective phenomena. For instance, an individual user cannot be polarized, but the distribution of users' opinions may be. The second defining ingredient of a complex system are interdependencies between the entities on the microlevel. On the Internet, users do not act in isolation but they share information, respond to each other, and exert influence on each other's opinions. In fact, the core argument underlying the personalization-polarization hypothesis proposes that personalization manipulates who is interacting with whom, changing the structure of interdependencies between users. This suggests that the analytical tools developed by complexity researchers have the potential to generate critical insight into personalization effects.

Research in various fields has demonstrated that complex systems can generate so-called "emergent phenomena", collective patterns that are a consequence of the behavior of the individual-level entities but that are external to the behavioral patterns of these individual-level actors [8–10,55]. In the social sciences, for instance, Schelling and Sakoda demonstrated that cities can segregate into black and white districts even when all inhabitants are tolerant [56–58]. In their models, agents accept to live in neighborhoods where their own ethnic group is in the minority. They leave their homes only when, for example, more than seventy percent of their neighbors belong to the other ethnic group. Cities segregate, despite this high degree of tolerance, because agents do not act in isolation. Whenever an agent moves, she changes her old and her new neighborhood, making her own group less represented in her old and more represented in her new neighborhood. These changes in the composition of her neighborhoods might convince her old and new neighbors who used to be satisfied with their neighborhood's composition to also move away. Thus, every moving has the potential to spark chains of reaction that intensify the ethnic homogeneity of neighborhoods and foster differences between neighborhoods to a degree that is not intended by the individuals that give rise to this pattern.

Also opinion polarization can be an emergent phenomenon, according to theories underlying the personalization-polarization hypothesis [17,18]. These theories do not assume that Internet users intend to live in a polarized world or that personalization increases their motivation to intensify opinion differences to other users. In contrast, these models assume that users seek to be positively influenced by their communication partners. However, personalization algorithms increase the degree to which they are communicating with likeminded individuals who likely expose them to information that reinforces their opinions. Thus, polarization is an unintended consequence of communication in a personalized world.

While complexity science appears to contribute a critical perspective on the personalization-polarization hypothesis, the public and scholarly debate largely ignores the complexity of online communication. We argue here that two typical characteristics of complex systems are largely overlooked. First, a typical characteristic of many complex systems is that even small and seemingly innocent aspects of a system can have immense impact on system behavior. In fact, theoretical as well as empirical research demonstrates that complex social systems can be in a state where even rare and random events can alter collective outcomes [15,59]. The segregation models by Schelling and Sakoda, for instance, generate higher segregation when small amounts of randomness are added to the behavior of the agents. That is, it is added that also agents who are satisfied with their neighborhood may move and that the agents who are dissatisfied happen to refrain from moving. It turns out that this randomness increases segregation, because every random moving by an agent has the potential to motivate further moving decisions by her old and new neighbors, potentially sparking a new cascade of segregation increasing moving sequences [60]. In the remainder of this section, we will provide examples of seemingly unimportant differences between communication in online and offline systems and illustrate why these differences might have important implications for the effects of personalization. Accordingly, we criticize contributions to the debate on the personalization-polarization hypothesis that are based on theoretical and empirical research in offline settings, as they might overlook important implications of communication in online systems.

A second typical characteristic of complex systems is that dynamics can be highly nonlinear. A typical example of a nonlinear dynamic on the Internet is the phenomenon that sometimes information goes “viral” [61,62]. In such an event, content is suddenly shared by a huge number of users and diffuses through the network at immense speed, creating bursts of attention that are notoriously hard to predict [63]. There is also a debate about the linearity of the effect of personalization. In their study of Facebook users, Bakshy et al. [21] found that the homophily generated by Facebook’s personalization algorithms is considerably smaller than the homophily resulting from users’ own tendency to select content that supports their political orientation. This may suggest that personalization is an innocent technology, but in a complex system this may not be true [33,64]. Increasing the temperature of water by one degree, for instance, usually does not have meaningful consequences, but it can trigger of a transition from liquid to gas when the temperature increases from 99 to 100 degrees Celsius. Likewise, it has been demonstrated that homophily has a nonlinear effect on systems tendencies towards polarization [33]. A slight increase in the already high degree of homophily on the Internet may be enough to tip the system over, and cause polarization. This is because algorithmically increasing homophily has an effect on many users. What is more, even when only a few users were directly affected by personalization algorithms, the change in the information diet of these users will indirectly affect the information diet of their friends and the friends of their friends.

The following subsections, we review central insights from complexity research on opinion dynamics in networks and conclude that the existing research on the personalization-polarization hypothesis is not sufficient. In particular, we show that the complexity of opinion dynamics can arise on three levels of analysis: the individual, the local, and the global level. We show that empirical and theoretical research on these levels is needed to test the personalization-polarization hypothesis. Table 1 summarizes the three levels of analysis.

Table 1. Levels of analysis on the personalization-polarization hypothesis

Level of analysis	Definition	Important open questions
Individual	The individual level relates to aspects of communication that affect processes internal to the sender and receiver of content.	- Who expresses their views online and do individuals express their opinions online in the same way as in offline interaction? - What is being communicated online and do individuals communicate different content online than offline? - Is content communicated differently in an online than in an offline setting? - How do individuals adjust their opinions after communication online and are opinions changed in the same way as after offline communication?
Local	The local level relates to aspects of communication that affect who is when encountering content emitted by whom.	- To which degree is polarization intensified when there is one-to-many communication rather than one-to-one communication? - To which degree is polarization weakened when forwarding content allows individual to exert direct influence on users they are not directly connected to?
Global	The global level relates to the structural characteristics of the communication network that affect individuals' content diet	- How does personalization change the structure of the communication network? - How do these changes affect the diffusion of online content in the network?

3.1. The individual level

The level of analysis that has certainly received most attention in the literature is the individual level. It is concerned with all processes that act within the sender and the receiver of communication in online social-networks. That is, it focused on who is emitting what content, to whom, and when. In addition, it matters who is when exposing herself to online content and how this content affects the opinions of the target of communication.

It turns out that different assumptions about how users update their opinions can lead to markedly different conclusions about whether web personalization increases or decreases polarization, as models of opinion dynamics demonstrate [18,33]. In particular, reinforcement models [17,18,65] and rejection models [66–69] imply competing predictions about the conditions under which polarization emerges.

The central assumption of reinforcement models is that individuals with opinions leaning towards one of the poles of the opinion scale will develop more extreme views after communication with a likeminded individual [17,18,65]. One theory supporting this assumption is Persuasive-Argument Theory [18,22,23], a psychological theory assuming that humans communicate arguments underlying their opinions. Individuals may hold a nuanced opinion themselves, but can only convey arguments that support or oppose an issue. During communication with likeminded individuals, users of online social networks will be mainly exposed to arguments in line with their own opinions. This, it is argued, reinforces their views and, thus, leads to more extreme opinions. Communication with users holding opposing opinions, in contrast, leads to opinion shifts in the opposite directions, as users are exposed to arguments challenging their opinions. The reinforcement of opinions also follows from biased-assimilation theory [17] and reinforcement-learning theory [65].

Reinforcement of opinions is a central assumption underlying the personalization-polarization hypothesis [17,33]. As personalization of online services increases the exposure to likeminded users and content that is in line with one's own views Internet users with opinions leaning towards the left end of the opinion spectrum would develop more leftist opinions and users with rightist opinions shift further towards the right. On the global level, this aggregates to increasing levels of opinion polarization, in line with the personalization-polarization hypothesis.

Rejection models, on the other hand, make different assumptions and also imply different macro-predictions [66,68,69]. Similar to the reinforcement models, rejection models also assume that individuals generally tend to grow more similar to likeminded individuals, an assumption that is usually implemented as averaging [13]. These models typically assume that users convey their exact position on an opinion continuum rather than exchanging arguments as is assumed by reinforcement models. Furthermore, it is added that individuals tend to dislike communication partners holding very distant views. Seeking to increase dissimilarity to, or distance themselves from persons they dislike, individuals adjust their opinions away from their communication partner, an opinion shift that is labeled "rejection" [70,71].

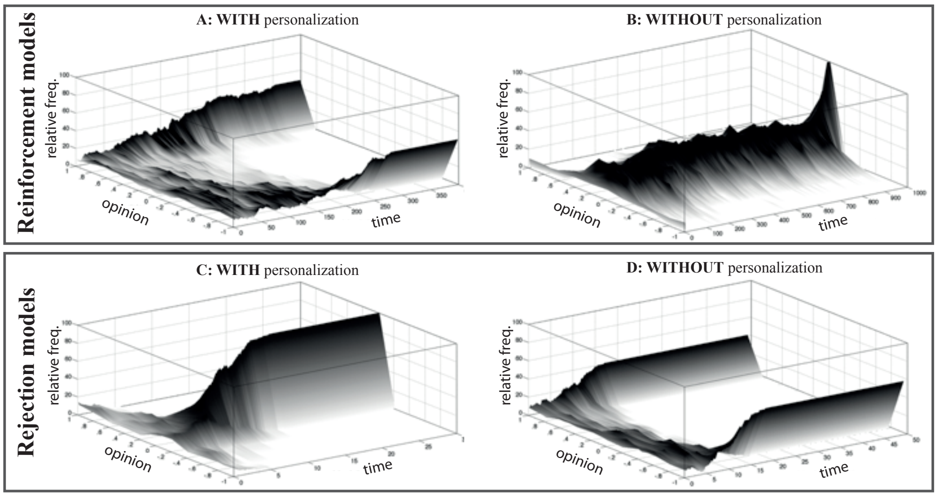

Figure 1. Predictions of reinforcement and rejection models

Rejection models contradict the personalization-polarization hypothesis [33]. As personalization leads to fewer encounters between users who hold opposing views, rejection is an increasingly unlikely event. Over time, users who hold the most extreme opinions engage in interactions with communication partners who are similar, but a bit less extreme, little by little pulling even the most extreme agents towards consensus. Rejection models thus predict that an increase in web personalization will decrease opinion diversity over time.

Figure 1 illustrates the contradicting predictions of reinforcement and rejection models, showing the distribution of opinions over time in two scenario's; with and without personalization. The figures in the top row show typical simulation runs with a reinforcement model and were generated with a model assuming persuasive-argument communication [18]. In the bottom row of the figure, we show two typical runs with a rejection model [67].

In a nutshell, depending on whether one assumes rejection models or reinforcement models, one will come to the conclusion that personalization either decreases or increases polarization. Empirical research on social influence, however, is inconclusive in that it does not inform about which of the two models or which combination of the two models is empirically more accurate. On the one hand, social-psychological research suggests that online communication should reduce rejection between members of different demographic subgroups or different political camps. As group memberships are not observed in online communication, group boundaries that might cause rejection effects in offline settings could turn irrelevant online [72]. On the other hand, there is also research pointing in the opposite direction. In qualitative research, it has been observed that online communication is often "unregulated by social context cues" [73]. In e-mails, users therefore use various tactics to allow receivers to better understand the meaning of their messages. Online social networks, however, restrict communication to relatively short messages, which makes communicating meaning and nuance more complicated. This, it has been observed, can cause confusion and rejection when receivers misinterpret messages [73,74]. Also experimental research on online social networks provided competing evidence for rejection [48,75]. Research on the persuasive-argument communication did provide ample of empirical support for reinforcement models, but this research is has been conducted in offline settings [22,23]. In sum, it remains an open empirical question whether users of online social networks emit and receive persuasive arguments as described by reinforcement models, in particular because communication in these settings is often restricted to very short messages.

In addition to individual responses to political messaging, personalized online environments may also affect senders' communication decisions. Recently, researchers reported that the personalized design of online platforms contributes to political outrage, rather than actual opinion shifts within individuals [76]. Predominantly communicating with likeminded contacts, users may experience outrage when content challenging their views enters their filter bubble [74]. Furthermore, users may misrepresent their opinions to obtain credibility among likeminded others, communicating more extreme views than they actually hold [77,78]. Since, in addition, extreme, moral, and emotional content tends to spread more easily on online social media [79] and since computer mediated communication decreases empathy on the sender's side [77], political debate within filter bubbles can grow more heated than users' actual opinions would suggest.

In conclusion, alternative theories of the individual-level processes in communication network make opposing predictions about whether the personalization-polarization hypothesis is true or false. In reality, online communication may be best described by a hybrid of assumptions from rejection and reinforcement models, but without empirical information about which theory is true under what conditions it seems hardly possible to derive reliable predictions about the consequences of web personalization.

3.2. The local level

The local level of observation is concerned with all mechanisms that govern the diffusion of information in individual's direct network neighborhoods. In the context of online social networks, this refers mainly to the technical implementation of communication and personalization. Unlike individual-level factors, local-level aspects are external to the individual sender or receiver. That is, these technical aspects do not affect how senders of communication emit online content and how receivers respond to communication. Local-level aspects change who is when encountering online content emitted by another user. It turns out that even seemingly small technical aspects have the potential to generate very different opinion dynamics than communication in offline systems. Most modeling work and the public and scholarly debate about the personalization-polarization hypothesis, however, tend to be based on theoretical models representing communication in offline worlds.

Here is a first example. One difference between communication in many online communication systems and offline face-to-face interaction is that in the online realm users often emit messages to all of their "friends" or "followers" at the same time. This so-called "one-to-many" communication differs from the "one-to-one" communication implemented in most models of social-influence [80,81]. On the one hand, the difference between one-to-one and one-to-many communication seems to be small, as a one-to-many communication-event is the same as a sequence of one-to-one communication events. On the other hand, modeling work with Axelrod's model of cultural dissemination demonstrated that one-to-many communication can foster opinion fragmentation in personalized systems [81].

Figure 2 illustrates why one-to-many communication might foster polarization. Assume that there are four users who "follow" each other on Twitter. Each user has a stance on three issues illustrated by their color (black or white), shape (circle or box), and letter (A or B). In Panel a of Figure 2, the number of lines connecting two users corresponds to the number of issues where users agree at the outset of the communication process. In a personalized system, this overlap will affect how likely an emitted piece of information will be consumed by the other user. The two users on the right, for instance, have zero opinion overlap and are, therefore, not exposed to each other's tweets.

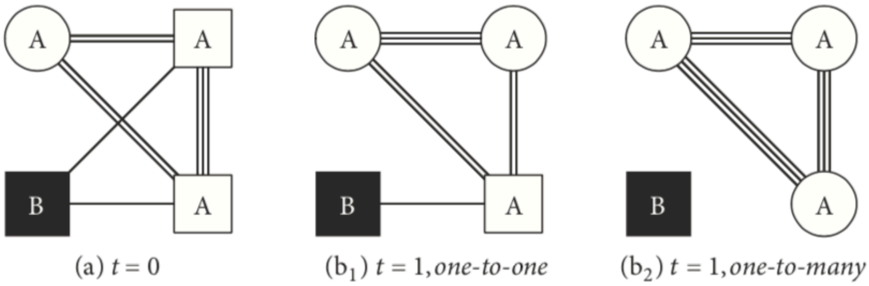

Figure 2. Illustration of the intuition that one-to-many communication fosters isolation [81]

Next, assume that the top-left user communicates her shape. Under the one-to-one communication regime, this trait may, for instance, be received by the top-right user, who adopts it and grows more similar to the sender as Panel b_1 of the figure illustrates. This instance of communication also changed the overlap between the receiver of the communication and the two remaining users, as a side effect. Nevertheless, the network remains connected and further communication between the two users on the right or the two users on the bottom can increase similarity between these users again.

Panel b_2 shows what happens under one-to-many communication when again the top-left user emits her shape trait to all of her followers and all followers with a non-zero overlap adopt her shape. As the bottom-left user does not share a trait with the sender, the personalization algorithm will not expose the bottom-left user to the message. This form of communication has two effects, as Panel c shows. First, a homogenous cluster formed because the communication did not only increase overlap between sender and each receiver. In addition, also the overlap between the two receivers increased. Second, the bottom-left user ended up isolated, as she no longer shares any trait with the three others. Communicating her shape, the sender did not only increase the overlap between herself and the two users on the right. In addition, the sender "pulled" these two users away from the bottom-left user. As a consequence, they will not interact with the isolated agent anymore.

The case shown in Figure 1 is the simplest scenario where the difference between one-to-one and one-to-many communication can be illustrated. Modeling work, however, demonstrated robust differences between one-to-one and one-to-many communication also in much bigger networks, in particular in networks characterized by high transitivity and high node degrees [81]. One-to-many communication increases the chances that individual agents are isolated and that multiple internally homogenous but mutually distinct subgroups form.

Personalization actually intensifies the difference between the two communication regimes, as it increases homophily. The central mechanism underlying the polarizing effect of one-to-many communication is that the sender of a message concurrently pulls away joint friends from users who disagree. Personalization decreases the probability that these friends will be influenced by the disagreeing user in the future, which can foster the fragmentation and polarization of opinions.

In sum, one-to-many communication has the potential to intensify clustering and polarization in personalized communication systems. To our knowledge, however, public and scholarly contributions to the debate about the personalization-polarization hypothesis have so far failed to consider this aspect.

A second potentially important aspect of communication in online social networks is that content can be easily forwarded to users who have no direct connection to the sources of the content. While sharing online content by forwarding incoming messages is very prominent on online social networks, it is unclear how opinion dynamics will be affected. Consider, for illustration, a simple line-network with three agents: A, B, and C. In this network, B is connected to A and C, but there is no link between A and C. The opinion scale ranges from minus one to plus one and A holds an initial opinion of $o_{A,t=0} = 0.6$. B holds and opinion of $o_{B,t=0} = 0$ and the opinion of agent C is $o_{C,t=0} = -0.6$. Furthermore, presume that agents are positively influenced by similar nodes. A simple formalization of this social-influence model would assume that agents always adopt the *weighted* average of their own opinion and the opinion communicated by their contact. Implementing positive social influence, influence weights adopt the value one, when the opinion differences between two actors do not exceed one (half of the opinion scale's range). Otherwise, weights adopt the value 0.5, which implements that influence decreases when actors disagree too much. What will happen in an offline setting where A first exerts influence on B and B subsequently influences C? First, as a result of the influence form A, B's opinion will shift in the direction of the opinion of A and will adopt the value of $o_{B,t=1} = \frac{(1\cdot 0.6\ +\ 0.5\cdot 0)}{(1+1)} = 0.3$. Second, having been exposed to B's updated opinion, also the opinion of C is adjusted from $o_{C,t=0} = -0.6$ to $o_{C,t=1} = \frac{(1\cdot 0.3\ +\ 1\cdot\ )}{(1+1)} = -0.15$. However, opinion shifts differ when the three agents communicate in an online setting and B forwards the message received from A, exposing C not to her own updated opinion but the initial opinion of A. In this case, C's updated opinion will be $o_{C,t=1} = \frac{(0.5\cdot 0.6\ +\ 1\cdot -0.6)}{(0.5+1)} = -0.2$, which is a bigger shift in the direction of the other two agents. This suggests that forwarding should foster the formation of a consensus.

This conclusion does not hold, however, when a slightly different influence model is assumed. To see this, assume that agents reject opinions that differ too much from their own view [66–69]. That is, assume that weights adopt a value of -0.5 (rather than 0.5) when the opinion differences exceed half the range of the opinions scale. According to this model, Agent C will be negatively influenced when receiving a forwarded message from Agent A, increasing the opinion differences in the network. According to this model, forwarding does not foster consensus formation but increases chances that dynamics generate a polarized opinion distribution.

In a nutshell, there are local-level aspects of communication in online social networks that have the potential to generate different opinion dynamics than communication in offline settings. So far, however, the debate about effects of personalization on polarization does not take into account these aspects. A key roadblock is a lack of empirical research on how users of online social networks adjust their opinions after communication, as the effects of local-level aspects can critically depend on this.

3.3. The global level

The global level refers to all structural elements of the communication network as a whole. For example, one characteristic of a network's structure that has been shown to have strong effects on opinion dynamics is *network clustering*, the degree to which connected nodes in a graph share other connections forming densely connected groups [80,82–84]. Consider the illustration in Figure 3 that shows two networks with 120 nodes and different degrees of clustering [85]. To generate them, we arranged nodes in a circle and created undirected links between each agent and their five nearest neighbors to the right and the five nearest neighbors to the left. The resulting network has 600 edges and is shown in panel a of Figure 3. It is characterized by very high clustering because this method of generating a network ensures a high number of triads, sets of three connected nodes. The transitivity coefficient – the number of realized triads over all possible triads – in this network amounts to .67. In contrast, the network shown in panel b of Figure 3 has a much lower degree of clustering. To generate it, we departed from the same circle network, but randomly rewired 35% of the links [86]. As a consequence, the number of links in the network and the number of links each agent has remained unaffected, but the transitivity coefficient dropped to .22.

In order to illustrate that network clustering affects opinion dynamics, we studied the dynamics generated by one of the most prominent social-influence models, the bounded-confidence model [87,88]. We chose this model, as it already has been used to derive hypotheses about the effects of personalization on opinion dynamics [34,35]. However, unlike earlier implementations of the bounded-confidence model, we assumed one-to-many communication, as this communication regime better mimics communication in online social networks [81].

To implement the bounded-confidence model, we assigned every agent a random initial opinion drawn from a uniform distribution ranging from zero to one. Dynamics were then broken down into a sequence of discrete events. At every event, a randomly picked agent exerted influence on each of her network neighbors. That is, the program selected always one agent i who then communicated her opinion to all of her network neighbors j. When the opinion difference between the source of communication and the respective target was smaller than the so-called "bounded-confidence threshold" ε, then the opinion of the target agent was updated according to Equation 1. Parameter μ represents how open agents are to social influence and was set to a value of .5.

$$o_{j,t} = o_{j,t} + \mu\left(o_{i,t} - o_{j,t}\right) \qquad (1)$$

This model assumes that agents can exert only positive influence on each other, which is implemented as opinion averaging [11,13]. However, two agents can only exert influence on each other when two conditions are met. First, the two agents need to be directly connected by a network link. Second, agents' opinions must be sufficiently similar, a simple representation of personalization [34,35]. Small values of the bounded-confidence threshold ε imply that agents are only influenced by very similar network contacts, which represents that the influence from network neighbors with dissimilar views is suppressed by a personalization algorithm. Higher values represent that agents are also exposed to influence by neighbors who hold relatively different opinions. This represents that personalization algorithms have a weaker effect. We ran all simulations until a state of equilibrium was reached in that further communication would not have led to opinion adjustments because all connected agents either held identical opinions or held opinions that were too different to result in social influence.

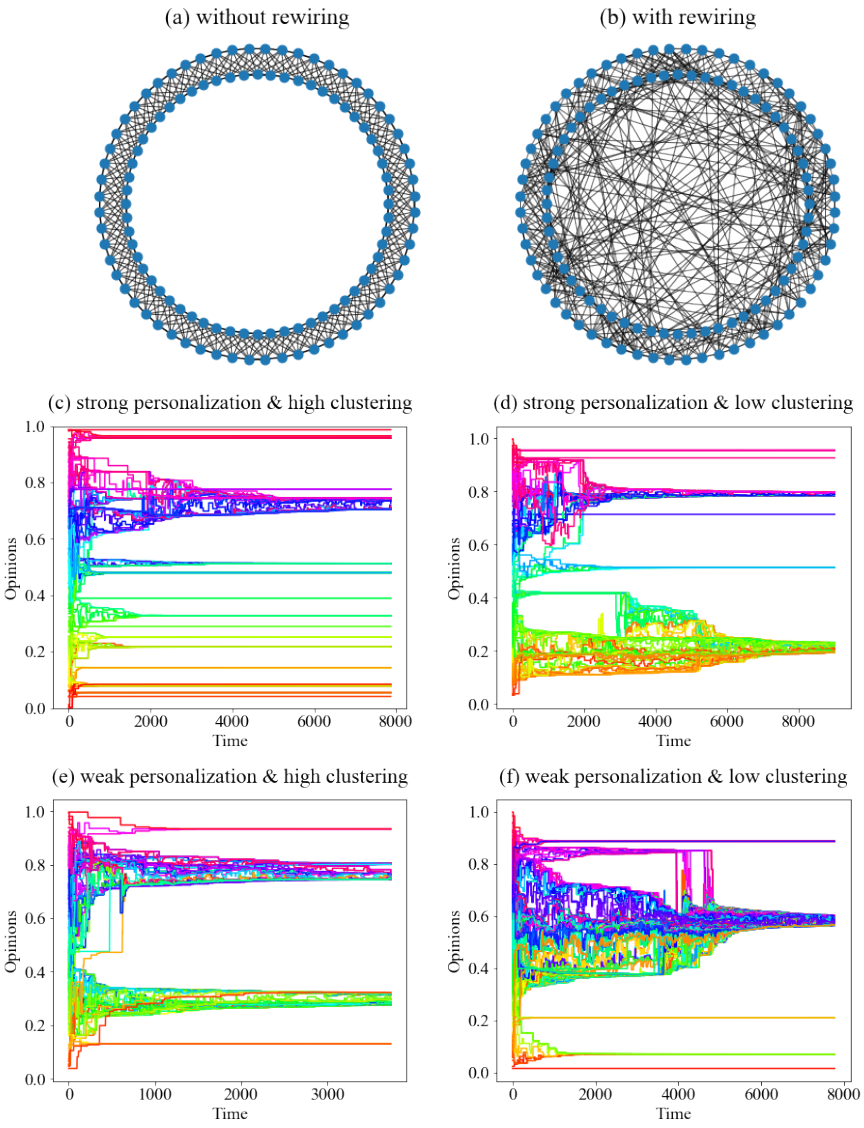

Figure 3. Effect of network clustering and personalization on opinion fragmentation.

In Figure 3, the four panels below the two network graphs show typical opinion dynamics in networks with high and low clustering and with strong or weak personalization. In each panel, we plot the trajectories of all 120 agents' opinions. Initially, all four opinion distributions were uniform, but dynamics always led to the formation of subgroups. Comparison of the dynamics on the left-hand side with those on the right-hand side shows that opinion dynamics resulted in the formation of a higher number of subgroups when network clustering was high. That is, highly clustered networks tend to fall apart into a larger number of homogenous but mutually distinct subgroups. Agents belonging to a subgroup hold identical opinions but the opinion differences to their network neighbors who do not belong to the same subgroup are too high to allow for more influence. Note that the bounded-confidence model, unlike the models studied in Section 3.1, fails to generate increasing opinion differences between subgroups if no further assumptions are added [89]. The model does, however, allow one to study the conditions of opinion fragmentation, the emergence of multiple subgroups.

Network clustering promotes opinion fragmentation because network clusters hamper the growth of subgroups. If, for instance, three agents are connected by two links and, thus, form a line network, then social influence will lead to opinion convergence if their opinions do not differ too much. A third link that would close the triad will in most cases not affect opinion dynamics in this small group. If this third link, however, has been rewired, there is a good chance that it connects the three agents to another agent with an opinion similar enough to make her join the subgroup.

Figure 3 also suggests that personalization fosters the formation of opinion subgroups, according to the bounded-confidence model. This effect obtains because personalization decreases the number of neighbors that agents exert influence on. Those neighbors who do influence each other, form homogenous groups, pulling agents who could have acted as bridges between groups towards the group's opinion average until they have grown too different from other groups to exert influence on them. When personalization is strong, agents exert influence on fewer neighbors. As a consequence, the network falls apart into a larger number of subgroups.

Panel a of Figure 4 shows that network clustering intensifies the effects of personalization on the emergence of subgroups according to the bounded-confidence model. The figure is based on a simulation experiment in which we experimentally varied network clustering and the strength of personalization. We studied the same circle networks as shown in Figure 3, including networks without rewiring (clustering = .67), networks with moderate clustering (105 rewiring iterations, average clustering = .38, sd = .02), and networks with strong clustering (210 rewiring iterations, average clustering = .22, sd = .02). In addition, we studied three levels of personalization, simulating dynamics under ε=.1 (strongest personalization), ε=.2, and ε=.3 (weakest personalization). For each of the nine experimental treatments, we studied 100 independent simulations runs and always counted the number of distinct opinion subgroups in equilibrium.

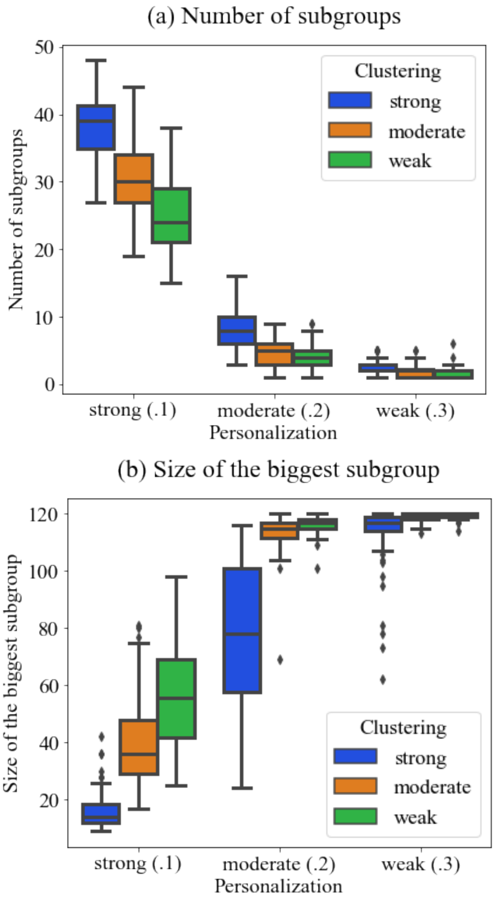

Figure 4. The effect of network clustering and personalization on opinion fragmentation measured by the number of subgroups and by the size of the biggest subgroup

Panel a of Figure 4 shows for all three personalization treatments that more distinct subgroups formed when the network was characterized by higher clustering. Poisson regressions revealed that the effect of the number or rewiring iterations on the number of subgroups observed in equilibrium was statistically significant in each personalization treatment (minimal z-value was -6.38). In addition, the effect of network clustering was strongest in the treatment with strong personalization. In fact, in a Poisson regression, there is a strong and significant interaction effect between the number of rewiring iterations and personalization on the number of subgroups in equilibrium (*b* = -3.49, *SE* = 1.20, *p* = .004, full model in appendix A).

Panel b of Figure 4 shows results from the same simulation experiment but reports the size of the biggest subgroup in the network as the outcome variable, revealing another interesting difference between the moderate and the weak personalization treatment. While panel a of Figure 4 depicts that the number of subgroups formed was relatively similar, panel b of Figure 4 shows that under weak personalization there tends to be one very big subgroup and a number of smaller subgroups. Under moderate personalization, the average number of subgroups increases from 1.98 to 5.69, but the size of the biggest group tends to be considerably smaller than under the low personalization treatment, showing that groups of more similar size had formed.

The presented analysis of the effects of network clustering illustrates, in a nutshell, that the structure of the communication network can affect opinion dynamics and the degree to which personalization technology affects the outcomes of social-influence processes. Obviously, network clustering is just one of many potentially important global aspects, as the opinion-dynamics literature demonstrates. Other global aspects are demographic diversity [67,90,91], network segregation [83,92], the number of bridges connecting otherwise disconnected network clusters [90], and the existence of agents with many connections [82,93].

To date, however, there is a lack of empirical research on the structure of online communication networks, which makes it hard to evaluate the personalization-polarization-hypothesis. There are three central roadblocks. First, even gathering data about online social networks is very challenging [94] and data allowing to quantify the structure and the evolution of communication networks is available only for very few networks [95–97]. Second, too little is known about the overlap between different networks. Critics of the personalization-polarization hypothesis do admit that online communication network can be segregated into clusters, but they also point to the fact that users tend to be active in various online and offline networks [7]. This, it is argued, creates crosscutting that allow information and arguments to travel from cluster to the other and decreases opinion polarization. Whether this is actually the case, however, is an empirical question that requires more research. In particular, it remains open whether and how individuals exert influence on each other's opinions in each communication network. For instance, users may use Twitter to communicate about political issues and focus on Facebook on entertainment and leisure. As a consequence, network overlap may have only limited impact on opinion dynamics. Third, personalization can also affect the structure of the interaction network. For instance, if personalization algorithms intensify the degree to which users are exposed to other users holding similar views, then they can also increase the degree to which the social network is clustered [98]. Assume, for illustration, that user A and user B are friends on Facebook and hold similar opinions. If Facebook's algorithms tend to propose creating links to users who hold similar views, then they may propose to both A and B to create a link to the same user C. While both links would result from the intention to create ties to likeminded users, an unintended consequence would be that A, B, and C form a triangle and, thus, contribute to network clustering. The analyses presented in this section have demonstrated that an increased degree of network clustering can further intensify processes of opinion polarization.

4. Conclusion

There is a public and scholarly debate about the hypothesis that the personalized technology of online services contributes to the polarization of political opinions. On the one hand, experts, scholars, and political decision makers warn that personalization creates echo chambers where users' opinions are reinforced as they are mainly exposed to content that does not challenge their views. On the other hand, there are skeptical contributions arguing that the homophily generated by personalization may be too mild to generate these undesired effects.

Both positions in this debate appear to leap to conclusions, from the perspective of researchers studying the complexity emerging from social-influence dynamics in social networks. We summarized insights from research on opinion dynamics in networks to show that more empirical and theoretical research needs to be conducted before one can arrive at reliable predictions about the effects of personalization. In particular, we argued that the opinion dynamics created by personalization critically dependent on aspects on the system's individual, local, and global level. To date, there is a lack of research into these aspects, which makes it impossible to reliably conclude whether or not personalization breeds polarization.

To be sure, we do echo the warning that personalization may have detrimental effects on public opinion formation and democratic decision making. These warnings need to be taken very seriously as democratic societies rely on an open public debate and a population's ability to find collective consensus. Although so far based on informal reasoning and anecdotal evidence, it is not an option to simply neglect the warnings.

The current state of the debate is worrisome for two reasons. First, the fact that there are theoretical arguments for and against negative effects of personalization allows stakeholders to cherry-pick arguments that support their interests. In his 2017 community address, for instance, Mark Zuckerberg referred to the rejection assumption, arguing that "ideas, like showing people an article from the opposite perspective, actually deepen polarization by framing other perspectives as foreign" [99]. In fact, Zuckerberg might be correct but so far research has not demonstrated this. Second, there are already various attempts to break filter bubbles with the help of sophisticated technology and international events creating debate between individuals holding opposite views (see Section 2). The problem is that designing a successful intervention requires a proper understanding of the opinion dynamics on personalized communication networks. If, for instance, opinion dynamics are better described by rejection models than reinforcement models, then interventions trying to expose users more to content challenging their views might increase rather than decrease opinion polarization (see Section 3). Interventions that are based on a false theory about how users exert influence on each other's opinions can backfire.

We advocate here an approach that combines formal theoretical modeling with empirical research. On the one hand, a purely empirical approach to testing the personalization-polarization hypothesis can lead to false conclusions. Assume, for instance, that an empirical study quantified the degree of personalization-induced homophily in various settings and found no correlation with opinion polarization in these settings. This finding certainly challenges the personalization-polarization hypothesis. However, in complex systems effects can take very long to unfold and can then be very abrupt and strong. In Panel A of Figure 1, for instance, polarization remained low for a long time, until it grew rapidly [33]. In addition, personalization algorithms are still being improved. The fact that they have not contributed to opinion polarization so far, does not imply that further advances in personalization will also remain without negative effects [64]. This suggests that the empirical observation that personalization so far appears to be relatively mild and its effects on opinions modest [7,21], should not lead one to conclude that personalization will remain an innocent technology in the future. On the other hand, also a purely theoretical approach will fail to generate reliable predictions about personalization effects, even when analytical and computational tools are used to derive predictions. Our review of the opinion-dynamics literature provided several examples of modeling decisions that can have big impact on the model's predictions. As a consequence, models relying on assumptions that have not been backed up by rigorous empirical research in the context of online social networks may fail to make true predictions and, in addition, will not be considered reliable tools for anticipating future opinion dynamics.

From our perspective, the most promising approach to deriving predictions about the future effects of personalization on opinion polarization is to develop empirically calibrated models, an endeavor that requires empirical and theoretical research [11]. Theoretical research is needed to identify those theoretical assumptions that have a critical impact on model predictions, as these assumptions need to be put to the test by empirical research. Our review has covered several aspects that require empirical investigation, but this list is not conclusive. To identify the most important mechanisms, modelers should invest more into comparing the predictions of alternative models [18,33,100–102]. Unfortunately, a recent review of the literature concluded that many contributors fail to highlight the similarities and differences between the model underlying their work and existing models [11], hampering the field's ability to accumulate knowledge and move forward. To improve, modelers should invest more into identifying these critical model assumptions, understanding why their model generates outcomes that other models do not. Furthermore, theoretical work should not only derive predictions about when a given model generates certain outcomes, but should find conditions under which different models provide different predictions. These insights will point empirical researchers to the empirical settings where competing models can be tested against each other, which in turn will help modelers develop validated models.

The emerging field of data science provides novel computational tools, sources of data, and methods of analysis to study opinion dynamics in online environments. Without proper theoretical foundations, however, attempts to empirically quantify the amount of online polarization or network segregation will remain underutilized [103–105]. Informing research on the individual level, many online services offer application programming interfaces (APIs) that provide researchers with information about the content that users share online. In tandem with novel methods of sentiment analysis and topic modeling, this may allow testing assumptions about who is communicating what content to whom on the Internet [74,106]. In addition, controlled online experiments shed light on how users adjust their opinions as a result of online communication [48,75,107–109]. On the local level, models need to be enriched with empirical information on how often users are exposed to online content on different online platforms and when they decide to contribute to online debates. Finally, there have been advances in gathering, storing, and analyzing detailed information about global-level factors [95–97]. In particular, there is considerable research on the structure of online communication networks, which make it possible to directly implement or regrow realistic communication networks in models of opinion dynamics [110–112].

Empirically validated models of social influence dynamics will not only make it possible to predict the consequences of web personalization, but they can also serve as a powerful tool to optimize personalization algorithms. Theoretically informed and empirically grounded computational models allow programmers to experiment with alternative specifications of personalization algorithms and analyze when they outperform each other on dimensions such as accuracy, scalability, user experience, and computational efficiency. In addition, validated models will make it possible to predict undesired effects of personalization technology on societal processes such as public debate, opinion polarization, and political decision-making, providing new tools to design algorithms that generate personalized services for individual users without harming societal dynamics. As communication technology is critical to democracy, such tools for the prediction of technological consequences are urgently needed.

References

[1] Pariser E. The Filter Bubble: What the Internet Is Hiding from You. New York: Penguin Press HC; 2011.

[2] Lapowsky I. How'd the Cohen Hearing Go? That Depends on Your Filter Bubble | WIRED [Internet]. 2019. Available from: https://www.wired.com/story/cohen-hearing-filter-bubbles/

[3] Chapin S. Who's Living in a 'Bubble'? - The New York Times [Internet]. 2018. Available from: https://www.nytimes.com/2018/12/11/magazine/whos-living-in-a-bubble.html

[4] Obama B. Farewell adress [Internet]. 2017. Available from: https://obamawhitehouse.archives.gov/farewell

[5] Steinmeier F-W. 2018 Christmas Message [Internet]. 2018. Available from: https://www.bundespraesident.de/SharedDocs/Reden/EN/Frank-Walter-Steinmeier/Reden/2018/12/181225-Christmas-message.html

[6] Bozdag E, van den Hoven J. Breaking the filter bubble: democracy and design. Ethics Inf Technol. 17(4):249–65. Available from: http://link.springer.com/10.1007/s10676-015-9380-y

[7] Bruns A. Are filter bubbles real? [Internet]. John Wiley & Sons; 2019. 144 p. Available from: https://www.amazon.com/Filter-Bubbles-Real-Axel-Bruns/dp/1509536442

[8] Bar-Yam Y. Dynamics of complex systems. Westview Press; 2003. 848 p.

[9] Mäs M. The Complexity Perspective on the Sociological Micro-Macro-Problem. SSRN Electron J. Available from: https://www.ssrn.com/abstract=3129362

[10] Page SE. What Sociologists Should Know About Complexity. Annu Rev Sociol. 41(1):21–41. Available from: http://www.annualreviews.org/doi/10.1146/annurev-soc-073014-112230

[11] Flache A, Mäs M, Feliciani T, Chattoe-Brown E, Deffuant G, Huet S, et al. Models of social influence: towards the next frontiers. Jasss-the J Artif Soc Soc Simul. 20(4). Available from: http://jasss.soc.surrey.ac.uk/20/4/2.html

[12] Mason WA, Conrey FR, Smith ER. Situating social influence processes: Dynamic, multidirectional flows of influence within social networks. Personal Soc Psychol Rev. 11(3):279–300.

[13] Friedkin NE, Johnsen EC. Social Influence Network Theory. New York: Cambridge University Press; 2011.

[14] Mäs M, Helbing D. Random Deviations Improve Micro–Macro Predictions: An Empirical Test. Sociol Methods Res.

[15] Macy MW, Tsvetkova M. The Signal Importance of Noise. Sociol Methods Res. 44(2):306–28.

[16] Smith A, Anderson M. Social Media Use in 2018. Pew Res Cent. (March):1–17.

[17] Dandekar P, Goel A, Lee DT. Biased assimilation, homophily, and the dynamics of polarization. Proc Natl Acad Sci U S A. 110(15):5791–6.

[18] Mäs M, Flache A. Differentiation without distancing. explaining bi-polarization of opinions without negative influence. PLoS One. 8(11).

[19] Lazarsfeld PF, Merton RK. Friendship and Social Process: A Substantive and Methodological Analysis. In: Berger M, Abel T, Page CH, editors. Freedom and Control in Modern Society. New York, Toronto, London: Van Nostrand; 1954. p. 18–66.

[20] McPherson M, Smith-Lovin L, Cook JM. Birds of a Feather: Homophily in Social Networks. Annu Rev Sociol. 27:415–44.

[21] Bakshy E, Messing S, Adamic LA. Exposure to ideologically diverse news and opinion on Facebook. Science (80-). 348(6239):1130–2.

[22] Myers DG. Polarizing Effects of Social Interaction. In: Brandstätter H, Davis JH, Stocker-Kreichgauer G, editors. Group Decision Making. London: Academic Press; 1982. p. 125–61.

[23] Vinokur A, Burnstein E. Depolarization of Attitudes in Groups . J Pers Soc Psychol. 36(8):872–85.

[24] Bryson B. `Anything but heavy metal': Symbolic exclusion and musical dislikes. Am Sociol Rev. 61(5):884–99.

[25] DiMaggio P, Evans J, Bryson B. Have American's Social Attitudes Become More Polarized? Am J Sociol. 102(3):690.

[26] Esteban J-M, Ray D. On the measurement of polarization. Econom J Econom Soc. 62(4):819–51.

[27] Evans J. Have Americans' Attitudes Become More Polarized?-An Update. Soc Sci Q. 84(1):71–90.

[28] Hunter JD. Culture Wars. New York: Basic Books; 1991.

[29] Hunter JD. Covering the Culture War: Before the Shooting Begins. Columbia J Rev. (July/August):29–32.

[30] Brewer PR. Polarisation in the USA: climate change, party politics, and public opinion in the obama era. Eur Polit Sci. 11(1):7–17.

[31] Zhou T, Kuscsik Z, Liu J-G, Medo M, Wakeling JR, Zhang Y-C. Solving the apparent diversity-accuracy dilemma of recommender systems. Proc Natl Acad Sci U S A. 107(10):4511–5. Available from: http://www.pnas.org/content/107/10/4511.abstract

[32] Lü L, Medo M, Yeung CH, Zhang Y-C, Zhang Z-K, Zhou T. Recommender systems. Phys Rep. 519(1):1–49. Available from: http://www.sciencedirect.com/science/article/pii/S0370157312000828

[33] Mäs M, Bischofberger L. Will the Personalization of Online Social Networks Foster Opinion Polarization? SSRN Electron J. Available from: http://papers.ssrn.com/abstract=2553436

[34] Geschke D, Lorenz J, Holtz P. The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. Br J Soc Psychol. 58(1):129–49. Available from: http://doi.wiley.com/10.1111/bjso.12286

[35] Sîrbu A, Pedreschi D, Giannotti F, Kertész J. Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. Gargiulo F, editor. PLoS One. 14(3):e0213246. Available from: http://dx.plos.org/10.1371/journal.pone.0213246

[36] Byrne D. The Attraction Paradigm. New York, London: Academic Press; 1971.

[37] Iyengar S, Hahn KS. Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use. J Commun. 59(1):19-U6.

[38] Morris JS. The Fox News factor. Harvard Int J Press. 10(3):56–79.

[39] Stroud NJ. Media use and political predispositions: Revisiting the concept of selective exposure. Polit Behav. 30(3):341–66.

[40] Adamic LA, Glance N. The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. In: 3rd International Workshop on Link Discovery, Association of Computing Machinery, August 21-25. Chicago, Illinois; 2005. p. 36-43.

[41] Johnson TJ, Bichard SL, Zhang WW. Communication Communities or "CyberGhettos?": A Path Analysis Model Examining Factors that Explain Selective Exposure to Blogs. J Comput Commun. 15(1):60–82.

[42] Vespignani A. Modelling dynamical processes in complex socio-technical systems. Nat Phys. 8(1):32–9. Available from: http://www.nature.com/articles/nphys2160

[43] Schmidt AL, Zollo F, Scala A, Betsch C, Quattrociocchi W. Polarization of the vaccination debate on Facebook. Vaccine. 36(25):3606–12. Available from: https://www.sciencedirect.com/science/article/pii/S0264410X18306601

[44] Nikolov D, Lalmas M, Flammini A, Menczer F. Quantifying Biases in Online Information Exposure. J Assoc Inf Sci Technol. 70(3):218–29. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.24121

[45] Isenberg DJ. Group Polarization: A Critical Review and Meta-Analysis. J Pers Soc Psychol. 50(6):1141–51.

[46] Sunstein CR. The law of group polarization. J Polit Philos. 10(2):175–95.

[47] Vinokur A, Burnstein E. Depolarization of Attitudes in Groups. J Pers Soc Psychol. 36(8):872–85.

[48] Guilbeault D, Becker J, Centola D. Social learning and partisan bias in the interpretation of climate trends. Proc Natl Acad Sci.

[49] Manjoo F. Can Facebook Fix Its Own Worst Bug? The New York Times Magazine. Available from: https://www.nytimes.com/2017/04/25/magazine/can-facebook-fix-its-own-worst-bug.html

[50] Flaxman S, Goel S, Rao JM. Filter Bubbles, Echo Chambers, and Online News Consumption. Public Opin Q. 80(S1):298–320. Available from: https://academic.oup.com/poq/article-lookup/doi/10.1093/poq/nfw006

[51] Barberá P. How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S. [Internet]. 2015. Available from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.658.5476

[52] Abramowitz AI, Saunders KL. Is polarization a myth? J Polit. 70(2):542–55.

[53] Boxell L, Gentzkow M, Shapiro JM. Greater Internet use is not associated with faster growth in political polarization among US demographic groups. Proc Natl Acad Sci U S A. 114(40):10612–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28928150

[54] DiMaggio P, Evans J, Bryson B. Have Americans' social attitudes become more polarized? Am J Sociol. 102(3):690–755.

[55] Camazine S, Deneubourg JL, Franks N, Sneyd J, Bonabeau E, Theraulaz G. Self-Organization in Biological Systems. Princeton, New Jersey: Princeton University Press; 2001.

[56] Schelling TC. Dynamic models of segregation†. J Math Sociol. 1:143–86.

[57] Sakoda JM. The checkerboard model of social interaction. J Math Sociol. 1(1):119–32. Available from: http://www.tandfonline.com/doi/abs/10.1080/0022250X.1971.9989791

[58] Hegselmann R. Thomas C. Schelling and James M. Sakoda: The Intellectual, Technical, and Social History of a Model. J Artif Soc Soc Simul. 20(3):15. Available from: http://jasss.soc.surrey.ac.uk/20/3/15.html

[59] Mäs M, Helbing D. Random deviations improve micro-macro predictions. An empirical test. Sociol Methods Res.

[60] van de Rijt A, Siegel D, Macy M. Neighborhood Chance and Neighborhood Change: A Comment on Bruch and Mare. Am J Sociol. 114(4):1166–80.

[61] Weng L, Menczer F, Ahn Y-Y. Virality Prediction and Community Structure in Social Networks. Sci Reports 2013 3. 3:2522. Available from: https://www.nature.com/articles/srep02522

[62] Cihon P, Yasseri T. A Biased Review of Biases in Twitter Studies on Political Collective Action. Front Phys. 4:34. Available from: http://journal.frontiersin.org/Article/10.3389/fphy.2016.00034/abstract

[63] Goel S, Anderson A, Hofman J, Watts DJ. The structural virality of online diffusion. Manage Sci. 62(1):180–96.

[64] Lazer DMJ. The rise of the social algorithm. Science (80-). 348(6239):1090–1.

[65] Banisch S, Olbrich E. Opinion polarization by learning from social feedback. J Math Sociol. 43(2):76–103. Available from: https://www.tandfonline.com/doi/full/10.1080/0022250X.2018.1517761

[66] Macy MW, Kitts JA, Flache A, Benard S. Polarization in Dynamic Networks: A Hopfield Model of Emergent Structure. Breiger R, Carley K, Pattison P, editors. Dyn Soc Netw Model Anal. (January 2003):162–73.

[67] Flache A, Mäs M. How to get the timing right. A computational model of the effects of the timing of contacts on team cohesion in demographically diverse teams. Comput Math Organ Theory. 14(1):23–51.

[68] Salzarulo L. A Continuous Opinion Dynamics Model Based on the Principle of Meta-Contrast. J Artif Soc Soc Simul. 9(1).

[69] Mark NP. Culture and Competition: Homophily and Distancing Explanations for Cultural Niches. Am Sociol Rev. 68(3):319–45.

[70] Festinger L. A Theory of Cognitive Dissonance. Evanston, White Plains: Row, Petersen and Company; 1957.

[71] Tajfel H, Turner JC. The Social Identity Theory of Intergroup Behavior. In: Worchel S, Austin WG, editors. Psychology of Intergroup Relations. Chicago: Nelson-Hall Publishers; 1986. p. 7–24.

[72] Postmes T, Spears R, Sakhel K, De Groot D. Social Influence in Computer-Mediated Communication: The Effects of Anonymity on Group Behavior. Personal Soc Psychol Bull. 27(10):1243–54.

[73] Menchik DA, Tian X. Putting Social Context into Text: The Semiotics of E-mail Interaction. Am J Sociol. 114(2):332–70. Available from: https://www.journals.uchicago.edu/doi/10.1086/590650

[74] Lin TZ, Tian X. Audience Design and Context Discrepancy: How Online Debates Lead to Opinion Polarization. Symb Interact. 42(1):70–97. Available from: http://doi.wiley.com/10.1002/symb.381

[75] Bail CA, Argyle LP, Brown TW, Bumpus JP, Chen H, Hunzaker MBF, et al. Exposure to opposing views on social media can increase political polarization. Proc Natl Acad Sci U S A. 115(37):9216–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30154168

[76] Brady WJ, Crockett MJ, Van Bavel JJ. The MAD Model of Moral Contagion: The role of Motivation, Attention and Design in the spread of moralized content online. PsyArXiv. 2019.

[77] Crockett MJ. Moral outrage in the digital age. Nat Hum Behav. 1:769–771.

[78] Jordan JJ, Hoffman M, Bloom P, Rand DG. Third-party punishment as a costly signal of trustworthiness. Nature. 530(7591):473–6.

[79] Brady WJ, Wills JA, Jost JT, Tucker JA, Van Bavel JJ. Emotion shapes the diffusion of moralized content in social networks. Proc Natl Acad Sci. 114(28):7313–8.

[80] Flache A, Macy MW. Local Convergence and Global Diversity: From Interpersonal to Social Influence. J Conflict Resolut. 55(6):970–95.

[81] Keijzer MA, Mäs M, Flache A. Communication in online social networks fosters cultural isolation. Complexity. :1–20.

[82] Castellano C, Fortunato S, Loreto V. Statistical physics of social dynamics. Rev Mod Phys. 81(2):591–646.

[83] Feliciani T, Flache A, Tolsma J. How, When and Where Can Spatial Segregation Induce Opinion Polarization? Two Competing Models. J Artif Soc Soc Simul. 20(2):6.

[84] Perra N, Rocha LEC. Modelling opinion dynamics in the age of algorithmic personalisation. Sci Rep. 9(1):7261. Available from: http://www.nature.com/articles/s41598-019-43830-2

[85] Watts DJ, Strogatz S. Collective dynamics of "small-world" networks. Nature. 393(6684):440–2.

[86] Maslov S, Sneppen K. Specificity and stability in topology of protein networks. Science (80-). 296(5569):910–3.

[87] Hegselmann R, Krause U. Opinion Dynamics Driven by Various Ways of Averaging. Comput Econ. 25:381–405.

[88] Deffuant G, Huet S, Amblard F. An Individual-Based Model of Innovation Diffusion Mixing Social Value and Individual Benefit. Am J Sociol. 110(4):1041–69.

[89] Hegselmann R, Krause U. Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation. J Artif Soc Soc Simul. 5(3).

[90] Mäs M, Flache A, Takács K, Jehn K. In the short term we divide, in the long term we unite: Demographic crisscrossing and the effects of faultlines on subgroup polarization. Organ Sci. 24(3):716–36.

[91] Flache A, Mäs M. Why do faultlines matter? A computational model of how strong demographic faultlines undermine team cohesion. Simul Model Pract Theory. 16(2):175–91.

[92] Grow A, Flache A. How attitude certainty tempers the effects of faultlines in demographically diverse teams. Comput Math Organ Theory. 17(2):196–224.

[93] Suchecki K, Eguíluz VM, San Miguel M. Voter model dynamics in complex networks: Role of dimensionality, disorder, and degree distribution. Phys Rev E. 72(3):036132. Available from: https://link.aps.org/doi/10.1103/PhysRevE.72.036132

[94] Russell MA, Klassen M. Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and More. O'Reilly Media, Inc.; 2018.

[95] Efstathiades H, Antoniades D, Pallis G, Dikaiakos MD, Szlavik Z, Sips R-J. Online social network evolution: Revisiting the Twitter graph. In: 2016 IEEE International Conference on Big Data (Big Data). IEEE; 2016. p. 626–35. Available from: http://ieeexplore.ieee.org/document/7840655/

[96] Stella M, Ferrara E, De Domenico M. Bots sustain and inflate striking opposition in online social systems. 115(49):12435–40.

[97] Myers SA, Sharma A, Gupta P, Lin J. Information network or social network? In: Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion. New York, New York, USA: ACM Press; 2014. p. 493–8. Available from: http://dl.acm.org/citation.cfm?doid=2567948.2576939

[98] Grund TU, Densley JA. Ethnic Homophily and Triad Closure: Mapping Internal Gang Structure Using Exponential Random Graph Models. J Contemp Crim Justice. 31(3):354–70. Available from: http://journals.sagepub.com/doi/10.1177/1043986214553377

[99] Zuckerberg M. Building Global Community [Internet]. 2017. Available from: https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634/

[100] Flache A, Macy MW, Takács K. What sustains cultural diversity and what undermines it? Axelrod and beyond. In: Takahashi S, editor. Advancing Social Simulation: Proceedings of the First World Congress on Social Simulation. Kyoto, Japan, Japan: Springer; 2006. p. 9–16.

[101] Kurahashi-Nakamura T, Mäs M, Lorenz J. Robust clustering in generalized bounded confidence models. J Artif Soc Soc Simul. 19(4):7.

[102] Mäs M, Flache A, Kitts JA. Cultural Integration and Differentiation in Groups and Organizations. In: Dignum V, Dignum F, Ferber J, Stratulat T, editors. Perspectives on Culture and Agent-based Simulations. New York: Springer; 2013.

[103] Golder SA, Macy MW. Digital Footprints: Opportunities and Challenges for Online Social Research. Annu Rev Sociol. 40(1):129–52. Available from: http://www.annualreviews.org/doi/10.1146/annurev-soc-071913-043145

[104] Lazer D, Pentland A, Adamic L, Aral S, Barabasi AL, Brewer D, et al. Computational Social Science. Science (80-). 323(5915):721–3.

[105] Conte R, Gilbert N, Bonelli G, Cioffi-Revilla C, Deffuant G, Kertesz J, et al. Manifesto of computational social science. Eur Phys J Spec Top. 214:325–46. Available from: https://link.springer.com/content/pdf/10.1140%252Fepjst%252Fe2012-01697-8.pdf

[106] Cohen R, Ruths D. Classifying Political Orientation on Twitter: It's Not Easy! In: Seventh International AAAI Conference on Weblogs and Social Media. 2013. Available from: http://www.senate.gov

[107] Becker J, Porter E, Centola D. The wisdom of partisan crowds. Proc Natl Acad Sci U S A. 116(22):10717–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31085635

[108] Clemm von Hohenberg B, Mäs M, Pradelski BSR. Micro Influence and Macro Dynamics of Opinion Formation [Internet]. 2017. (SSRN). Available from: https://ssrn.com/abstract=2974413

[109] Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. Sociol Methods Res. 40(2):211–239.

[110] Holme P, Kim BJ. Growing scale-free networks with tunable clustering. Phys Rev ERev E. 65:026107.

[111] Newman MEJ. The structure and function of complex networks. Siam Rev. 45(2):167–256.

[112] Mislove A, Koppula HS, Gummadi KP, Druschel P, Bhattacharjee B. Growth of the flickr social network. In: Proceedings of the first workshop on Online social networks. 2008. p. 25–30.