Journal of Computational Social Science (2025) 8:88
https://doi.org/10.1007/s42001-025-00422-7

RESEARCH ARTICLE

Filter bubbles as a vector of tradition? Decoding opinion dynamics with agent-based modelling

Benoît Desmarchelier[1] · Faridah Djellal[2] · Faïz Gallouj[2]

Received: 9 January 2025 / Accepted: 24 July 2025 / Published online: 17 August 2025

Abstract
This paper investigates the relationship between online filter bubbles and the emergence of echo chambers. In particular, building on an agent-based simulation model, we find that the narrower the filter bubble, the larger the number of echo chambers, and the lower the degree of consensus among the population of agents as a whole. According to the existing literature, echo chambers present risks of polarization, extremism and separatism. Yet, we do not recommend reinforcing the regulation of filter bubbles. Indeed, our model suggests that the presence of sources of serendipitous knowledge, like for instance in the form of third places, can prevent the apparition of echo chambers, even in the presence of narrow filter bubbles. Going back to Gabriel Tarde's classical distinction between opinion, tradition, and reason, we argue that sources of serendipitous knowledge favor the development of a society of opinion, while echo chambers resemble micro-societies of tradition. In this perspective, current developments of information technologies produce a tension between tradition and opinion.

Keywords Network · Filter bubble · Echo chamber · Agent-based modelling · Serendipity

✉ Benoît Desmarchelier
benoit.desmarchelier@univ-paris13.fr
https://scholar.google.com/citations?user=B23MJV4AAAAJ&hl=en&oi=ao

Faridah Djellal
faridah.djellal@univ-lille.fr
https://scholar.google.com/citations?hl=en&user=uJcB8mAAAAAJ

Faïz Gallouj
faiz.gallouj@univ-lille.fr
https://scholar.google.com/citations?hl=en&user=jWNvUf4AAAAJ

1 Sorbonne Paris-Nord University, Villetaneuse, France

2 University of Lille, Lille, France

1 Introduction

Dismal predictions about the future of the debate of ideas are multiplying over the past few years. Recently, the House of Lords [20] pointed out three interlinked problematic trends in the evolution of media.[1]

First, there is the widespread use of online recommendation systems. These are algorithms embedded into online platforms, such as Twitter, Facebook, Google News and the likes, to rank items so as to attract users' attention [34]. Feezell et al. [13] distinguish two types of such algorithms: the "user-driven" ones, that use a person's online history as input to make recommendations, and the "socially driven" algorithms, based on the social network of the user. Both types are filtering online content so as to avoid confronting users to objecting views [37], therefore creating "filter bubbles" [17, 34]. To the best of our knowledge, this term was first coined by Pariser [34], who defines it as "a unique universe of information for each of us […] which fundamentally alters the way we encounter ideas and information". (p. 17). This personalization of news threads is indeed suspected to reinforce one's views, and therefore to produce polarization, instead of promoting a widely shared opinion across the society [10, 24].

A second noteworthy development is the use of social media and their growing influencer culture. In a recent report, Ofcom,[2] the broadcasting regulator in the UK, indicates that the audience of traditional, generalist, media sources is declining. For instance, BBC One used to be watched by 58% of adults in the UK in 2019, against 43% in 2024. In parallel, the audience of social media is rising: while inexistant in 2019, TikTok is used as a media source by 11% of the adults in the UK in 2024. Youtube figures switched from 6 to 19% during that time span (p. 10). By age group, Instagram, Youtube, Facebook, TiTok and X form the top 5 news sources for the 16–24 years old in the country in 2024 (p. 11). The problem is that, with these platforms, "the traditional role of news in 'convening society' is becoming inverted, as users consume disparate stories and follow opinion-shapers on personalized news feeds" [20], p. 43). In particular, "younger audiences [show] preference for 'authentic' content rather than 'authoritative' sources" (p. 43). As such, everybody constructs his/her own definition of the truth, making productive debates harder to organize. A particular issue is the development by the users themselves of a "self-selective exposure" called "echo chamber" [10]. These "are user-made homophilous clusters where most users share the same views about an issue and where individuals primarily share information that supports their views" (p. 2). Jacob and Banisch [21] speak of "structural polarization", in the sense that, when forming an echo chamber, people with similar opinions disconnect themselves from the rest of the society. In addition to separatism, these echo chambers might fuel extremism. Indeed,

[1] We thank the four anonymous reviewers, who—with the insightfulness of their comments—have greatly contributed to improving our manuscript.

[2] https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/tv-radio-and-on-demand-research/tv-research/news/news-consumption-2024/news-consumption-in-the-uk-2024---supporting-data.pdf?v=379623 [last access: Dec. 17th, 2024].

Baumann et al. [5] suggest that interaction of like-minded individuals is conducive to strengthening their opinions.

The third trend is only emerging, it is the development of artificial intelligence (AI). “Whether the AI provides one answer to a question about current affairs, or a variety of news, may shape the average casual user’s views on a topic” [20], p. 32). If AI tools are trained to please their users, like the recommendation algorithms, then they can lock individuals into an information bubble shaped by a confirmation bias.[3]

The dangers that have been listed so far—i.e. polarization, separatism and extremism—motivate some researchers to propose guidelines in the design of recommendation systems, so as to force them to promote "exposure to diversity" [19]. But homophily—i.e. a tendency to interact primarily with others sharing a certain degree of homogeneity with oneself, including in terms of opinions [27]—is not a new bias in human behavior: it is at the basis of the emergence of cultures and subcultures [2] and it explains the diffusion of many fads into society [35]. In fact, homophily is a central assumption in many models of social influence [8, 14, 30]. Yet, echo chambers go beyond homophily, because they mean a complete “insulation from rebuttal” [1], p. 10).

There is an ongoing debate about whether recommendation systems create filter bubbles and favor opinion fragmentation. For instance, using a nation-wide survey conducted in the United States, Feezell et al. [13] find no evidence that algorithmic news generates opinion polarization. Similarly, using interaction patterns observed on Facebook, Bakshy et al. [3] find only a marginal effect of algorithmic rankings on the users’ exposure to "cross-cutting content" (p. 1131). However, such results can be platform dependent. Accordingly, Kaiser and Rauchfleisch [23] highlight that channels recommended by Youtube are creating homophilous communities, especially around far-right themes. For their part, de Zúñiga et al. [12] identify that social media and recommender systems favor the emergence of a new type of behavior, labelled "news find me perception", which consists in the belief of being "well informed about public affairs and political news without being actively seeking news information anymore" (p. 1). One can reasonably suspect that filter bubbles, when coupled with this behavior, can result in echo chambers.

The present paper is an attempt to contribute to this ongoing debate on whether filter bubbles can be conducive to the emergence of echo chambers. We address this question by means of an agent-based simulation model. In particular, we add filter bubbles to the classical model of social influence proposed by Carley [8]. Simulations confirm that these filter bubbles produce the appearance of disconnected clusters of agents characterized by a higher degree of consensus than what is observed in the population as a whole. This result, which stands as a distinctive feature of our model, suggests the presence of echo chambers. In addition, we find that narrower filter bubbles provoke the multiplication of echo chambers and a decline in the level of consensus within the population.

[3] A recent tragic example of such an information bubble is reported by CNN, according to which a teenager’s personal AI assistant encouraged him to murder his parents. https://edition.cnn.com/2024/12/10/tech/character-ai-second-youth-safety-lawsuit/index.html [last access: Dec. 17th 2024].

Taking into account the fact that individuals are usually embedded into offline networks of interaction [10] and that they might attend to third places (Oldenburg and Brisset, 1982), such as cafés and associations, that may be sources of serendipitous knowledge, we also add a probability that agents get exposed to unexpected discovery to the model. This amendment succeeds at conciliating narrow filter bubbles with the emergence of a society-wide discussion network, characterized by the disappearance of all forms of unanimity, and the emergence of a moderately high form majority opinion (shared by around 60% of the agents). By analogy with Gabriel Tarde's distinction between opinion, tradition, and reason (1901), we advocate that serendipitous knowledge favors the development of a society of opinion, while echo chambers stand as micro-societies of traditions.

The rest of the paper is organized as follows. The Sect. 2 reviews existing models on social influence, paying particular attention to the modelling of opinion polarization occurring in social media. Section 3 presents the simulation model. Simulation results are detailed in Sect. 4. The Sect. 5 provides more perspective to the results by interpreting them using Tarde's distinction between the tradition, the opinion, and the reason. Finally, we provide a conclusion to the paper in Sect. 6.

2 An overview of the literature on social influence

The literature on social influence provides many models that can serve as a basis for modelling filter bubbles and investigating their consequences for the dynamic of opinions. Flache et al. [14] identify three classes of such models: (i) the models of "assimilative social influence", (ii) the models of "similarity biased influence", and (iii) those of "repulsive influence" (p. 7). We start by reviewing these three categories of models, then we review how polarization on social media is usually modelled.

2.1 Models of assimilative social influence

French [15], DeGroot [9] and Friedkin [16] are prime examples of this line of models, where an agent's opinion typically evolves as a linear combination of neighbors' opinions. More specifically, in DeGroot [9], the opinion of an agent i at time step 2, $F_{i,2}$, is given by the Eq. (1), where $p_{i,j}$ represents the weight attributed by the agent i on the opinion of agent j, and $F_{j,1}$ stands as the opinion of j at the previous time step (see DeGroot, [9], p. 119).

$$F_{i,2} = \sum_{j=1}^{k} p_{i,j}.F_{j,1} \tag{1}$$

As the Eq. (1) suggests, unless in the case where $p_{i,j} = 0$, each agent i integrates the opinions of all other k agents. Thus, there is no homophily in this model. As might be expected, if the network is sufficiently well connected—that is when most weighting coefficients are such as $p_{i,j} > 0$—then agents' opinions converge towards some consensus around a central value of F.

Exploring further this model, Golub and Jackson [18] attribute an initial information to each agent, augmented by an error term. Then, they search for the network structure that ensures the convergence of opinions towards the true information signal. Their experiments indicate that a necessary and sufficient condition is that "the influence of the most influential agent [should] vanish [...] as the society grows" (p. 114). Otherwise, opinions converge towards the one of the most influential agents. This point is of particular relevance to us, as the "influencer culture" [20] implies that few individuals get disproportionate popularity. In this context, Golub and Jackson [18] suggest that people's opinion should converge towards those of these influencers. Flache et al. [14] add that the emergence of a consensus requires a high connectivity in the network.

Assimilative models can therefore be of some relevance for the study of opinion dynamics in the era of modern internet. But this line of models present several drawbacks. Firstly, the assumption of a highly connected network might not be realistic, and it is not adapted to the simulation of echo chambers, which require an interaction network made of several components. Secondly, the averaging process only applies to continuous opinions.

2.2 Models with a similarity bias

This type of models assumes that individuals prefer interacting with those displaying some similarity with them—i.e. they exhibit some degree of homophily (Mc Pherson et al., 2001). It has to be noted that "homophily has [often] been used in ABM studies to emulate the effect of social media algorithms" [21], p.2). This assumption is therefore of particular interest to tackle the question of the emergence of echo chambers.

The aggregate consequences of homophily have been explored in models where agents are spatially situated. Schelling [36] initiated the spatial models with his famous problem of segregation. Here, agents belonging to two different communities decide where to live on a grid, given that they exhibit a weak preference for their own community with regards to the ethnic composition of their neighborhood. As agents move freely on the grid, the system eventually stabilizes when the two communities are highly segregated.

Axelrod [2] adopts a similar spatial approach to study the apparent paradox of the coexistence of a culture homogenization at the aggregate level, with the persistence of subcultures at the microeconomic level. The model takes the form of a grid in which each cell represents an agent with four neighbors (except for the corner cells, which have two neighbors). An agent's culture is represented by a list of digits. At every step, one agent selected at random picks up one of his/ her neighbors at random. The probability that an interaction actually takes place between them is proportional to their cultural proximity and, in case of interaction, the calling agent adopts one of the cultural traits of his partner.

At first sight, one may consider that this model uses a regular network, where each agent has four neighbors. But disconnection can actually happen when two neighbors have completely dissimilar cultures. As such, the simulations succeed at producing an overall homogenous culture, with some small irreducible clusters—which may be assimilated to echo chambers.

By reducing dissimilar agent's willingness to interact, homophily seems to favor the emergence of clustering. We also note that Axelrod [2] moves away from the continuous opinions used in averaging models. However, the variety of network configurations permitted by this model remains limited: because the agents cannot move freely on the grid, they cannot interact with others, located at distance. Further, the exchange process is not reciprocal, since the calling agent adopts a cultural trait of his/ her counterpart, while this latter remains unaffected. We also note that only one interaction takes place at a time. In contrast, in the real world, online discussions might involve individuals located at distance, several interactions can happen at the same time, and one can assume that both agents can alter their opinions within the interaction process.

Carley [8] proposes a model that displays such characteristics (i.e. bidirectional exchange, and simultaneity of several discussion dyads). Here, the agents are not located in a grid. Also, it is assumed that there are $K = 1...k$ facts that can either be known or ignored by an agent i at time t—i.e. $F_{i,k}(t) = 1$ or $F_{i,k}(t) = 0$. In accordance with the hypothesis of similarity bias, an agent i chooses another agent j to interact with with a probability given by the share of knowledge facts they have in common. If the chosen agent j is not already busy interacting with someone else, then both i and j select one fact at random among those they already know, and their partner becomes knowledgeable about it.

As might be expected, since there are not contradictory opinions in the model, the overall system tends towards a situation where all agents know everything. The objective of the model, as it was formulated by Carley [8], is therefore less to study the possibility of a consensus than to compute the speed of its emergence. The model nonetheless presents a number of important advantages. (i) It operationalizes homophily through a form of random process. (ii) Agents' interactions are not bounded by some geographical location, nor by some arbitrary initial network configuration decided by the modeller. (iii) The interaction network emerges endogenously from agents' partially homophilous search.

2.3 Models of repulsive influence

Here, agents may decide to differentiate their opinions from those of their dissimilar counterparts. Jager and Amblard [22] propose a model able to account for both attraction and repulsion among a population of N agents. Each agent i is described by a set of three parameters (x_i, u_i, t_i), with $x_i \in [-1,1]$ the opinion of the agent i, u_i a "latitude of acceptance" (p. 296) of opinion differences, and $t_i > u_i$ a threshold of rejection. Agents interact at random and, in case of encounter, the opinion of the agent i evolves as described in the Eqs. (2) and (3) (see [22], p. 296).

$$If\ |x_i - x_j| < u_i\ then\ dx_i = \mu.(x_j - x_i) \tag{2}$$

$$If\ |x_i - x_j| > t_i\ then\ dx_i = \mu.(x_i - x_j) \tag{3}$$

In Eq. (2), the agent i approaches his/her views to those of the agent j—we can speak of an attraction effect. Alternatively, in the case of Eq. (3), the agent i moves further away from the views of the agent j, it is a repulsion effect.

Depending on the distribution of the values of u_i and t_i, the model can produce bi-polarization, convergence, and clustering of opinions. Also, the authors demonstrate that these results are robust to the integration of the agents into a grid, therefore to the constraint of localized interactions.

2.4 Modeling polarization in social media

These three mechanisms—i.e. assimilative influence, similarity bias, and repulsion—have been used to simulate the polarization of opinions occurring in social media.

Keijzer et al. [24] represent a filter bubble by a mechanism similar to Carley's (1991) matching algorithm: an agent receives the opinions of only a subset of the population, constituted by like-minded agents. The larger this subset, the larger the filter bubble. In this point of view, the filter bubble operates like homophily. Indeed, a homophilous person would only consider interacting with a narrow subset of very similar agents—the stronger the homophily, the narrower this subset.

In terms of opinion dynamics, Keijzer et al. [24] blend attraction and repulsion: both the sender and the receiver approach their views when their opinions are close enough, while they adjust in the opposite direction if their difference in opinion is deemed too large. This dual process of attraction–repulsion creates convergence of opinions within filter bubbles, and divergence between these bubbles. As such, filter bubbles can create echo chambers.

The relationship between filter bubble and polarization is also explored by Del Cerro [10] in the context of the Catalan separatism. Here, agents are positioned into a multiplex network: an offline network of relations that takes the form of a small-world [39], partly intertwined with a scale-free online network [4]. Information sharing behavior differs in these two networks: in an online context, agents share information with all their social ties, whereas offline they only share it with one of their contacts, chosen at random. The model can be run with a filter bubble—where agents receive biased information, supporting their prior beliefs. Agents hold a "national identity score", which is updated only when the information received is not too different from the agent's convictions. Del Cerro [10] tests for scenarios with and without filter bubbles, combined with random and homophilous social networks. With this model, she finds that polarized opinions require both homophilous networks and filter bubbles to emerge.

Jacob and Banisch [21] further develop the idea of multiplex networks. In their model, agents are positioned in a fixed random graph representing the offline world. These agents select one of m virtual worlds based on the degree of their agreement with these worlds' members.[4] Another originality of the model is that agents can hold a private opinion, different to the one they express in public. An agent reveals his hidden opinion when he becomes sufficiently confident. Also, the authors distinguish "un-structural polarization"—i.e. when polarisation does not translate into a disconnection of the social network into separate hubs—with "structural polarization" (p. 2), where effective disconnection occurs. Simulations reveal that homophily can result in structural polarization online, and un-structural polarization offline, although it is not clear if this later result is not the product of the assumption of a fixed offline network. A similar conclusion is reached by Baumann et al. [5]. In their model, agents' opinions can be either positive or negative, without upper or lower limits, allowing for radicalization when interacting with a similar minded partner. On the other hand, the model assumes that interacting with opposite-minded agents produces convergence in opinions.

Ng and Carley [32] bring the discussion of opinion change at the individual level, as they search on Twitter for the characteristics of people who are more likely to change their stance about the vaccination against the covid-19. Notably, they find that "two degrees of neighbors and the connectivity of neighbors contribute significantly to the influence of an agent's stance" (p. 13). They also distinguish between two types of agents: the non-bots and the bots, these later being more likely to change their stance on a given issue. The authors suspect this difference to be the expression of bots' strategy to influence others' opinion. This hypothesis is investigated by Carragher et al. [7]. Through the simulation of various strategies of opinion changes and friends selection, they establish that the best way to influence others' opinion is to start by approaching the most popular views of the close network. They also find support for the hypothesis that a tipping point is necessary to provoke large opinion change within a population of agents.

All these models uncover specific factors at play in the polarization of opinion in social media: filter bubbles—which are offering agents with an homophilous sub-set of choice alternatives -, multiplex network structures, specific strategies adopted by bot agents, and various human behaviors of opinion convergence and differentiation when interacting with others. This discussion confirms that filter bubbles are a candidate to explain the emergence of echo chambers. In particular, the model from Keijzer et al. [24] integrates filter bubbles of various sizes, but with specific rules of attraction and repulsion in agents' opinion changes behavior. As such, it remains difficult to untangle the specific role of filter bubbles. In that respect, the model in Carley [8] presents several distinctive features:

- The model is of relative simplicity, which makes it more easily amendable. Simplicity also means that we can isolate with more confidence the effect of a specific change on the model's overall dynamic.

[4] For a model representing the functioning of a virtual world, see Murdock et al. [29].

- Like in the online world, agents' interactions in Carley [8] are not constrained by a pre-existing physical network or by some characteristics of the geography.
- The model is based on the principle of a probabilistic similarity bias that resembles the way filter bubbles operate. Indeed, in this model agents rank all the others based on a similarity assessment. This process resembles page ranking algorithms on the Internet. In this context, a filter bubble could be easily implemented in the form of a limitation in the number of alternatives to choose from. Like in Keijzer et al. [24], we could then simulate filter bubbles of various sizes.

For these reasons, we decide to use Carley [8] as a basis for our model of opinion change. Details of the model are provided in the next section.[5]

3 Presentation of the simulation model

A drawback of the model in Carley [8] is that agents do not formulate attitudes towards knowledge facts. Indeed, in this model, either $F_{i,k}(t) = 1$ or $F_{i,k}(t) = 0$, i.e. a fact k is either known or unknown by the agent i. Consequently, the first amendment we make to the model is to allow for opinion divergence. As such, agents can now hold a negative view on a fact k, ignore it, or hold a positive view on it, that is $F_{i,k}(t) \in (-1,0,1)$.

The second amendment we bring to Carley [8] consists in the introduction of filter bubbles of various sizes. Let N_{Pop} be the number of agents in the model, and m, the number of recommendations made by the filter bubble. When the filter bubble is activated, agents choose randomly with whom to interact among the $m < N_{Pop}$ most similar other agents. As such, tuning the m parameter allows us to simulate filter bubbles of various sizes.

Carley [8] computes an index of "culture homogeneity" (p. 336), which is the percentage of facts for which there is complete consensus within the society. For that reason, we rather call it the unanimity index. Be UI_t the unanimity index at time t, its formula is given in the Eq. (4), with N_F the number of facts to be known, $UF_j = 1$ if all agents share the same opinion about the fact j, and $UF_j = 0$ otherwise.

$$UI_t = \frac{1}{N_F} \times \sum_{j=1}^{N_F} UF_j \tag{4}$$

In order to better capture the level of convergence or divergence of agents' opinions, we also calculate another index: MS_t, the average market share of the most frequent (or majority) opinion. Be, $MaxCount_j$ the number of agents who share the most frequent opinion about the fact j, the formula for MS_t is provided in the Eq. (5).

[5] The model has been written in C++. The source code can be downloaded freely on the CoMSES platform: https://www.comses.net/codebases/c240dcc6-1746-4cc4-b7a3-c8630a100418/releases/1.0.0/ (last access: May 7th 2025).

$$MS_t = \frac{1}{N_F} \times \sum_{j=1}^{N_F} \left(\frac{MaxCount_j}{N_{Pop}} \right) \quad (5)$$

Further, we record every interaction happening throughout the simulation runs so as to reconstruct the emerging networks.[6] With these networks, we identify echo chambers with two criteria: (i) the presence of several components[7] and (ii) the fact that opinions—measured by the UI_t and MS_t indexes—are more homogenous within these components than within the population of agents as a whole.

A typical time step of the model works as follows:

1. Assign agents a random order of passage.[8]
2. Each agent i computes the percentage of shared knowledge with each of the other agents j and ranks them in decreasing order of opinion similarity.
3. Agents call their interaction method (see Fig. 1 for details).
4. Record the number of interactions, increment the network file, compute UI_t and MS_t.

Figure 1 details agents' interaction method. First, in order to discuss with someone else, the agent i has to know something. With K_i the number of facts known by the agent i, interaction is possible only when $K_i > 0$. If this condition is met, and without filter bubble, the agent i selects an interaction partner j at random, with a probability of choice proportional to the degree of shared knowledge. This selection routine, taken from Carley [8], operationalizes the principle of homophily. As mentioned earlier, partner's selection is not constrained by geography or by an existing network configuration. When there is a filter bubble, then the agent i selects randomly an agent j among the m potential partners that are being proposed to him by the online filter. These are the m most similar agents to i in terms of shared knowledge. If j knows something and is not busy interacting with someone else, then an interaction takes place. Concretely, both agents i and j select a fact at random among those they know. Then, the agent i adopts the view of the agent j and vice versa. Note that this information exchange mechanism does not presume opinion homogenization. Such homogenization would occur only if one of the two agents shares an opinion that has already been adopted by the other agent.

This procedure for interaction reveals several important parameters of the model:

[6] The model generates network files in Pajek format, compatible with the most popular network analysis softwares.

[7] A component is "a subset of the vertices of a network such that there exists at least one path from each member of that subset to each other member" [31], p. 142). If a network consists in a single component, then all agents are at least indirectly reachable by any other agent. At the other hand, if a network contains two or more components, then some agents cannot be linked together, even indirectly.

[8] Agents operate sequentially in the model. Under these conditions, the first agents to intervene are those most likely to find unoccupied partners. As such, we randomize agents' order of passage to limit the bias that could result from this situation.

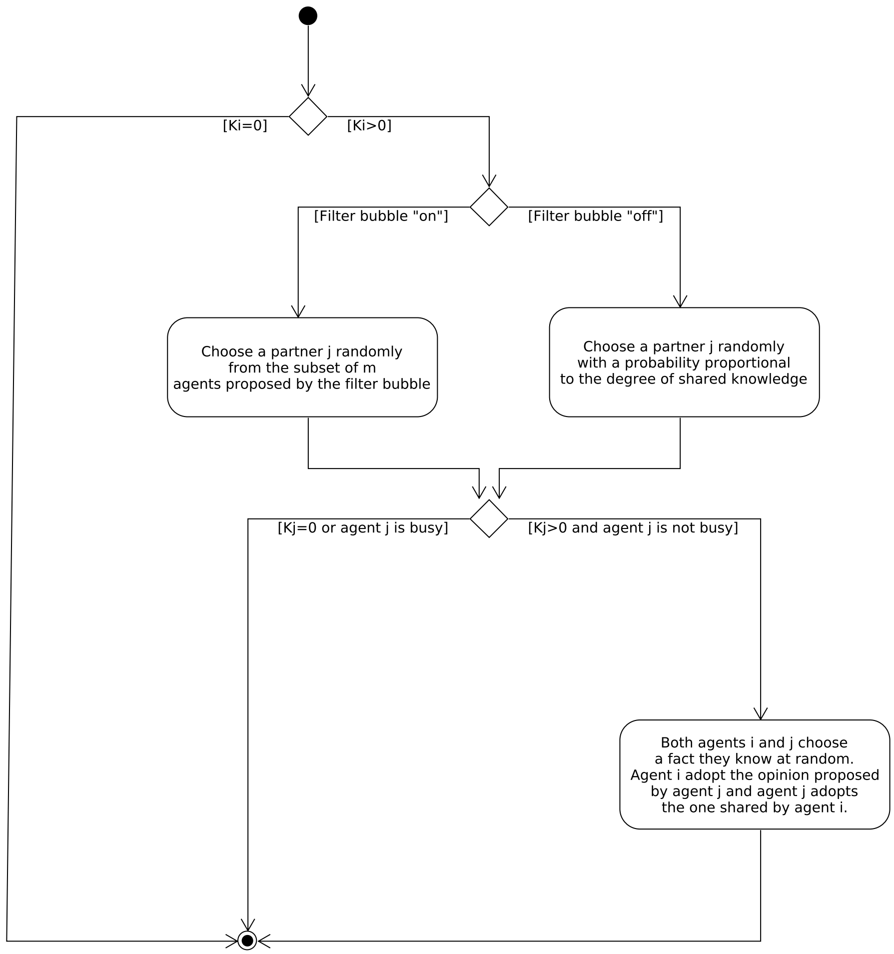

Fig. 1 Agents' interaction method

- N_{Pop}, the number of agents.
- K_0 the initial percentage of knowledge facts for which agents have a native opinion at $t = 0$[9].
- N_F, the number of knowledge facts to be known.
- m, the size of the filter bubble.

[9] In the model, there would be no interaction at all if agents had no prior knowledge (i.e. if K0 = 0). This condition is less artificial than one may think. Indeed, when discussing the difference between human beings and systems of artificial intelligence, Le Cun [26] and Mitchell [28] indicate that humans always approach novel problems with a certain degree of prior knowledge, for instance acquired during early socialization, which they call "common sense". Thus, we can interpret the K0 parameter in our model as the level of agents' common sense at the start of a simulation.

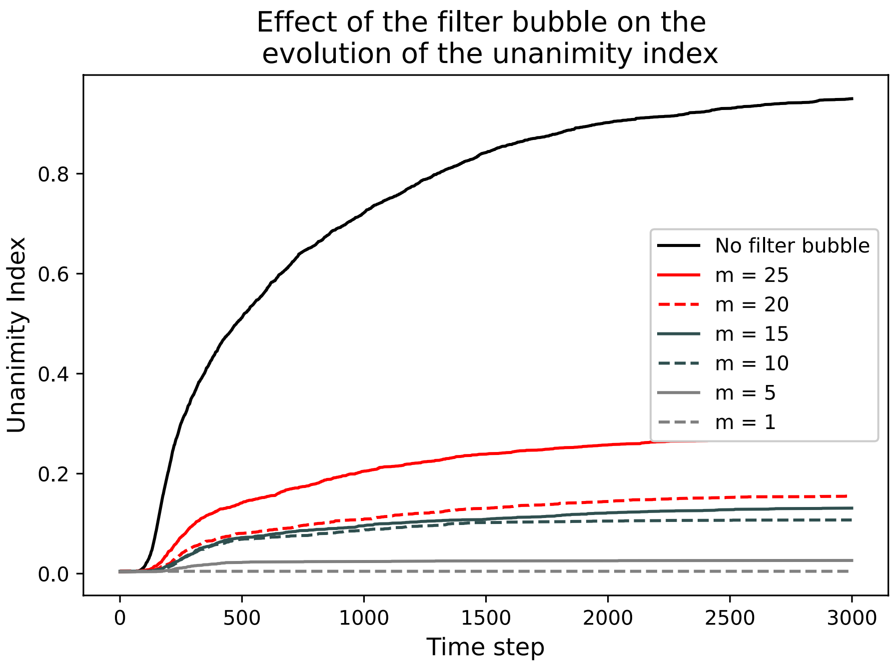

Fig. 2 Effect of the filter bubble on the evolution of the unanimity index (average trajectories over 100 simulation runs). (In this figure, m is the number of alternatives proposed by the filter bubble.)

In what follows, we choose $N_{Pop} = 50$, $N_F = 30$ and $K_0 = 10\%$. Also, a typical simulation run lasts 3000 time steps.[10] In what follows, we present average statistics generated by the model over 100 repetitions without filter bubble, and with filter bubbles of various sizes—i.e. $m = (25{,}20{,}15{,}10{,}5, or 1)$. By doing so, we can highlight the effect of filter bubbles on the model behavior, especially with regard to the emergence of echo chambers.

4 Simulation results

The section presents the results of our simulations. We start by identifying the emergence of echo chambers, and then we propose a potential solution.

[10] Multiple simulation runs revealed that increasing or decreasing parameters Npop, Nf and K0 only accelerate or decelerate the dynamic of the model, without important consequences for its emergent properties. Also, in most cases, the model reaches its final equilibrium state by time step 2000. Likewise, average statistics do not change significantly above 30 repetitions.

4.1 From filter bubbles to echo chambers

Figure 2 presents the evolution of the unanimity index, UI_t, over time depending on the size of the filter bubble. We observe that without a filter bubble, opinions are gradually moving towards complete consensus. This result is similar to Carley [8], although consensus is certainly slower to emerge in our version of the model, because we introduced a wider opinion choice set for the agents. This general pattern changes dramatically with the introduction of a filter bubble. Indeed, even with the largest bubble—i.e. when $m = 25$ recommendations out of a total population of 50 agents—the model produces unanimity for only around 20% of the knowledge facts. This number gradually declines as the filter bubble shrinks.

The absence of unanimity does not necessarily imply complete disagreement. An illustration of this is provided by the Fig. 3, which displays the evolution of MS_t, the average market share of the most frequent opinion, depending on the size of the filter bubble. Without filter, MS_t tends asymptotically towards 1. This result is coherent with the unanimity found in Fig. 2 for that configuration of the model. Introducing a filter bubble reduces the equilibrium level of MS_t (Fig. 3), but to a lower extent than in the case of the unanimity index (Fig. 2). Indeed, with a filter bubble of a size $m = 25$, the most frequent opinion is still shared, on average, by 94.6% of the agents at $t = 3000$ (Fig. 3). This number drops to 75% and to 28.4% for $m = 5$ and $m = 1$,

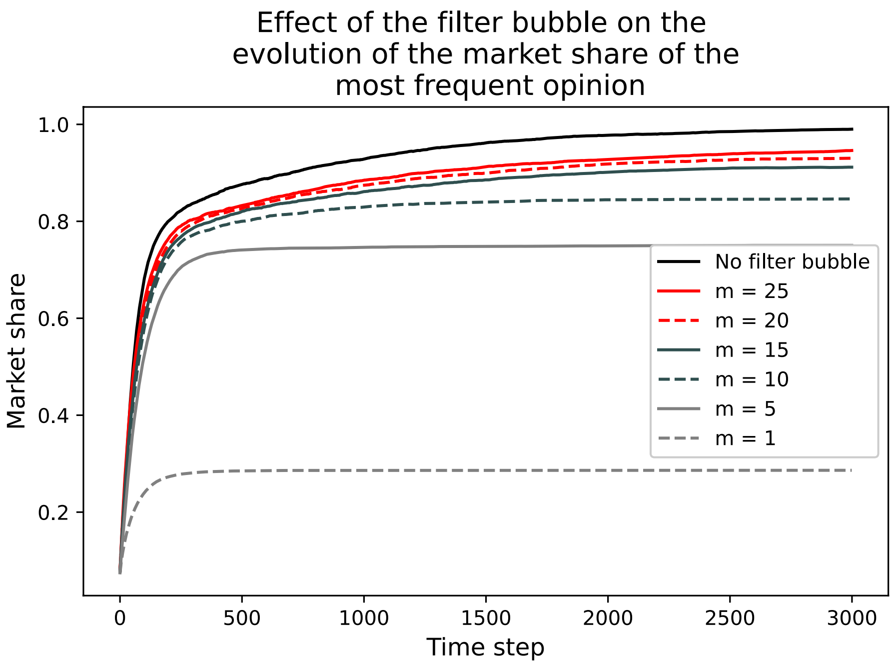

Fig. 3 Effect of the filter bubble o,n the evolution of the market share of the most frequent opinion (average trajectories over 100 simulation runs). (In this figure, m is the number of alternatives proposed by the filter bubble.)

Table 1 Summary of intra-components statistics (m is the number of alternatives proposed by the filter bubble)

		Mean	Std
No bubble	Average number of components	1.03	0.17
	Average intra-component UI_t	0.98	0.028
	Average intra-component MS_t	0.989	0.052
m = 25	Average number of components	2.35	1.258
	Average intra-component UI_t	0.95	0.038
	Average intra-component MS_t	0.965	0.118
m = 20	Average number of components	2.45	1.058
	Average intra-component UI_t	0.944	0.037
	Average intra-component MS_t	0.955	0.132
m = 15	Average number of components	2.63	1.212
	Average intra-component UI_t	0.927	0.048
	Average intra-component MS_t	0.95	0.125
m = 10	Average number of components	2.79	1.258
	Average intra-component UI_t	0.834	0.108
	Average intra-component MS_t	0.925	0.116
m = 5	Average number of components	2.99	1.307
	Average intra-component UI_t	0.702	0.186
	Average intra-component MS_t	0.889	0.127
m = 1	Average number of components	10.67	2.629
	Average intra-component UI_t	0.917	0.056
	Average intra-component MS_t	0.942	0.149

respectively (Fig. 3). Thus, in the model, opinions remain mostly shared by individuals, even with a filter as narrow as $m = 5$.

Both indicators (i.e. UI_t and MS_t) lean towards the same conclusion: narrower filter bubbles reduce the extent of agreement within the population. We are now wondering if filter bubbles can generate echo chambers. Following the concept of structural polarization [21], we characterize an echo chamber by two criteria: it is an isolated sub-network of discussion that presents a higher degree of agreement than what is observed at the level of the entire population.

To identify the presence of echo chambers, we record every interaction happening at each time step of the model.[11] This allows us to reconstruct the entire discussion network. We then proceed as follows:

- We count the number of components of the network.
- We compute the within-component UI_t and MS_t indexes based on agents' opinions at $t = 3000$.

[11] At the start of a simulation run, the network file produced by the model is empty. It is then filled-in progressively over the course of the simulation. When an interaction occurs between agents i and j, a link is recorded between them.

Results are summarized in Table 1. In terms of connectivity, we observe that networks with no filter bubble are mostly made of a single component. This does not mean that all agents interact with everyone, but that opinions can circulate within the entire population. In this respect, the introduction of filter bubbles produces a noticeable change. Indeed, even in the larger bubble (i.e. $m = 25$), networks are on average made of 2.35 components, with a wider dispersion of this statistic (Table 1). In fact, the narrower the filter bubble, the larger the number of components. Thus, the extreme case with $m = 1$ produces on average 10,67 distinct components.

With regards to agents' opinions within these components, unanimity is achieved—on average—for 94% of facts when $m = 25$, and for 70% of the facts when $m = 5$ (Table 1). The corresponding figures for the entire population were $UI_t = 27\%$ and $UI_t = 2.6\%$ when $m = 25$ and $m = 5$, respectively (see Fig. 2). This larger propention for unanimity inside components pleads for the presence of echo chambers. The market share of the most frequent opinion, MS_t, goes into the same direction, with intra-component values of 96% and 88,9% for $m = 25$ and $m = 5$, respectively (Table 1), against 94% and 75% for the entire population (see Fig. 3). We can therefore consider that the model succeeds at generating echo chambers, and that filter bubbles stand as a determinant factor in their appearance.

The general pattern, both within-components and within the entire population, is of a decline in agreement when filter bubbles become narrower, except when $m = 1$, where intra-components $UI_t = 91.7\%$ and $MS_t = 94.2\%$ (Table 1). We explain this surprising result by the fact that, when $m = 1$, several components are made of a single agent who never interacted with anyone.

4.2 Sources of serendipitous knowledge as a potential solution

The fact that narrow filters provoke the multiplication of echo chambers is a pessimistic result, as current AI models typically aim at making ever more precise recommendations [26, 28]. An obvious solution can be to manipulate the filter bubbles, for instance by imposing a minimum threshold in the number of recommendations they make. Helberger et al. [19] go in this direction when they suggest to force filters to pursue "exposure to diversity".

The problem posed by filter bubbles is certainly attenuated by the fact that people do not spend all their life online. Indeed, they also take part in many offline networks—at the workplace for instance—that include sources of serendipity, that is of unexpected discoveries [6]. There are also online initiatives that promote users' exposure to a plurality of perspectives, like the My Country Talks[12] and the Echo Chamber Club[13] websites [24]. Another source of unexpected knowledge can be the reading of a newspaper. Finally, one can mention the existence of many third places, like the coffee houses, pubs, fablabs, or even public libraries [11, 25]. Third places are places of "pure sociability" [33], p. 270), i.e. where human relations are not constrained by status, nor by any specific organizational

[12] https://www.mycountrytalks.org/ (last access: May 16th 2025).

[13] https://archive.echochamber.club/who-are-we.html (last access: May 16th 2025).

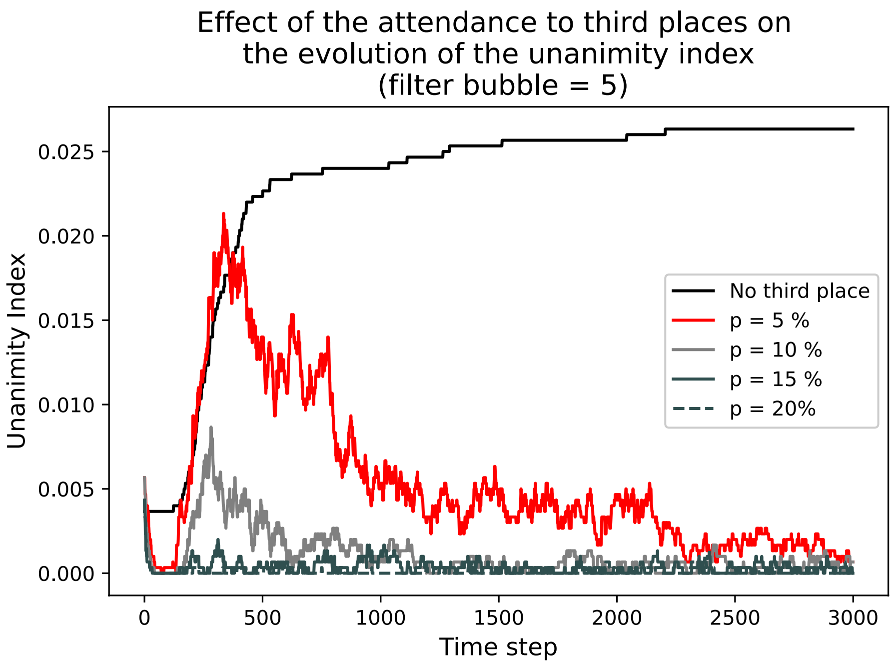

Fig. 4 Effect of the probability p of the attendance to third places—i.e. a source of unexpected discovery—on the evolution of the consensus index with a narrow filter bubble ($m = 5$)

objective. They permit casual discussions between strangers on equal terms, and as such, they provide access to a "diversity and novelty" of opinions and ideas (Oldenburg and Brisset, 1982, p. 273). In particular, third places "which are not insulated by formal membership requirements, often uniquely provide a common meeting ground for people with diverse backgrounds and experiences. Depending upon when a person stops in [...] he may chance to meet the friend of a friend; someone's visiting relative; someone new to the area, and perhaps just some of the regulars" (p. 275).

Can all these sources of serendipity-offline networks, diversity-promoting websites, and third places—counter the emergence of echo chambers when there are narrow filter bubbles? Within the framework of our model, we address this question by modifying agents' interaction method as follows:

- When an agent i does not know anything—i.e. when $K_i = 0$ (see Fig. 1)—or when the chosen partner j is already busy interacting with someone else, then the agent i attends to a third place with a probability p.
- When that happens, one of his knowledge bits gets modified at random.

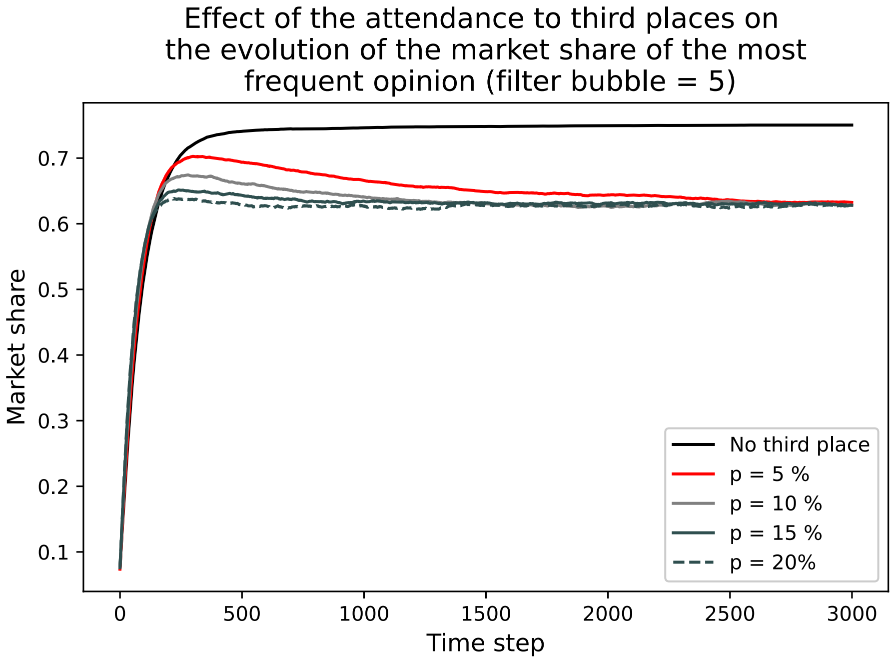

Fig. 5 Effect of the probability p of attendance to third places—i.e. a source of unexpected discovery—on the evolution of the market share of the most frequent opinion, when the filter bubble is narrow (m = 5)

Table 2 Summary of intra-components statistics when a source of serendipitous knowledge has been introduced in the model (in this Table, m = 5 and p is the probability to attend a third place)

		Mean	Std
$p=0.05$	Average number of components	1.00	0.00
	Average intra-component UI_t	0.00	0.00
	Average intra-component MS_t	0.63	0.02
$p=0.10$	Average number of components	1.00	0.00
	Average intra-component UI_t	0.00	0.00
	Average intra-component MS_t	0.63	0.02
$p=0.15$	Average number of components	1.00	0.00
	Average intra-component UI_t	0.00	0.00
	Average intra-component MS_t	0.63	0.02
$p=0.20$	Average number of components	1.00	0.00
	Average intra-component UI_t	0.00	0.00
	Average intra-component MS_t	0.63	0.01

We test this new behavioral routine with various values of p in the presence of a narrow filter, that is when $m = 5$.[14] We start by exploring the consequences of the third place for the emergence of consensus within the population as a whole. Then, we investigate the presence of network components.

With no third place and $m = 5$, the population reaches unanimity for 2.6% of the knowledge facts on average at $t = 3000$ (see Figs. 2 and 4). This low proportion drops asymptotically to zero when third places are introduced, that is for all values of p (Fig. 4). Thus, unexpected discoveries prevent unanimity. But this does not mean that all forms of consensus have disappeared. Indeed, for all values of p, the most frequent opinion remains shared by around 60% of the population on average, at the equilibrium of the model (Fig. 5)—a number that remains inferior to that without access to third places.

In previous simulation runs, the relation that emerged was that narrower filters produce lower degrees of consensus and the multiplication of separated components in the interaction network. Are the relatively low consensus levels in the presence of third places favoring the emergence of echo chambers? To respond to this question, Table 2 provides summary statistics about the networks' components. We observe that for all values of p, networks are always made of a single component. This implies that, in the model, sources of serendipitous knowledge prevent echo chambers. Because interaction networks are made of a single component, UI_t and MS_t statistics reported in Table 2 are those of the entire population. Consistently with Fig. 4, we observe in Table 2 that unanimity has disappeared ($UI_t = 0$), and consistently with Fig. 5, we find that the majority opinion is shared by slightly more than 60% of the agents (Table 2).

5 Discussion

To label the various forms of consensus produced by the model, we refer to the work of Gabriel Tarde, an early sociologist, who witnessed the emergence of public opinion in his classic book The Opinion and the Crowd (1901).

Tarde distinguishes three kinds of shared ideas, or "fractions of the social spirit" (p. 64): tradition, reason, and opinion. Reason is "the personal, relatively rational, though often unreasonable, judgments of an elite that isolates itself and thinks and steps outside the popular current in order to stem or direct it" (p. 64). It is produced by academics, philosophers and, in general, an intellectual elite. For its part, tradition is defined as "the condensed and accumulated extract of what was the opinion of the dead" (p. 64). It is a minimal and relatively stable form of consensus, which is acquired by the young through early education and socialization. Tradition is a stable, more or less innate, and highly cohesive form of consensus.

In the context of our model, tradition first refers to situations of high UI_t and MS_t indexes. It is achieved in two cases: when there are no filter bubbles nor sources of serendipitous knowledge—and within the echo chambers produced after the introduction of filter bubbles. Among these two situations, we consider

[14] We do not consider m = 1, because we found that, in this case, the interaction networks are typically made of components with a single agent, which biases the statistics about consensus.

that only the echo chambers meet Tarde's definition of the tradition. Indeed, without filter bubble, consensus is obtained after a sustained deliberation among all the agents—revealed by the progressive growth of the UI_t and MS_t indexes in Figs. 2 and 3. For their part, echo chambers stand as associations of like-minded individuals a priori: in Figs. 2 and 3, UI_t and MS_t indexes reach their equilibrium level relatively fast when $m = 5$, which reveals that interactions quickly stop, while at the same time intra-components levels of consensus are important in this case (Table 1). To put it another way, within the echo chambers of the model, local consensus requires only a few interactions to be reached.

Echo chambers of the model stand as local forms of consensus, which make them comparable to traditional societies. Interestingly, in Tarde's view, to be maintained and shared, traditions require some physical proximity between individuals, while by definition, online echo chambers abstract from geography. As such, we can say that filter bubbles promote the emergence of an abstract form of traditional societies, freed from locational boundaries.

For Tarde [38], opinion is "a temporary and more or less logical group of judgments which, responding to problems currently posed, are found reproduced in numerous people of the same country, of the same time, of the same society". (p. 68). Compared with tradition, the opinion is shared by a broader group of persons, and it is a more temporary form of consensus. Historically, it appeared thanks to the mass media, via printed books and, more importantly, with the daily press and railways, because with these two innovations, the same news could be read at the same time by a large and physically dispersed audience.

Tarde's public opinion is "temporary" (p. 68), that is subject to changes. Within our model, such a situation cannot be unanimity (i.e. high UI_t). Indeed, in principle in the model, agents change their opinions by adopting those of someone else (see Fig. 1). In that context, unanimity would mean that opinions cannot change anymore. Yet, public opinion stands as the view of a majority—we therefore need $MS_t > 50\%$. This situation of low UI_t and moderately high MS_t is achieved when a source of serendipitous knowledge is introduced (i.e. when $p > 0$ in Figs. 4 and 5). Indeed, in that case, there is no unanimity at all ($MS_t = 0$), and a majority opinion ($UI_t \simeq 60\%$). The last criterion is the size of the discussion network: the view of the majority should be widely shared among all agents. Here again, this condition is met when $p > 0$, because in that case agents' interactions occur on a single component (see Table 2). We can therefore say that, in the model, the introduction of an external source of knowledge produces a society of opinion—even in the presence of a narrow filter bubble (i.e. $m = 5$ in Figs. 4 and 5, and in Table 2).

Overall, our simulation results suggest that filter bubbles promote the maintenance of micro-societies of diverse traditions which take the form of echo chambers, while third places—or any other source of serendipitous knowledge—favor the emergence of a society of opinion. Interestingly, this society of opinion emerges even in the presence of narrow online filters. For that reason, it appears more effective to promote the multiplication of third places than to promote the control of online filter bubbles if we wish to prevent the disappearance of a society of opinion, and therefore of discussion and debate.

6 Conclusion

In writing this paper, we aimed to investigate the consequences of online filter bubbles for the dynamic of opinions, especially in terms of the emergence of echo chambers. These echo chambers are indeed being suspected to be a source of opinion polarization—hence of disagreement—and of extremism. We proceed by adding filter bubbles to the agent-based model proposed in Carley [8]. In our view, this model is a good starting point, because it produces unanimity among all of its agents. Also, it has the advantage of producing endogenous interaction networks, instead of relying on ex ante hypotheses about the structure of such networks. This point is of importance, because echo chambers take the form of isolated clusters of homogeneous individuals.

Our simulations allow us to identify the following relation: narrower filters produce lower degrees of consensus and the multiplication of separated components in the interaction network. Confronted with this situation, we wondered if sources of serendipitous knowledge, like the third places, could break the echo chambers produced by filter bubbles. Thus, we added a probability of access to alternative views for the agents of the model. Despite the presence of a narrow filter bubble, this modification produces a society-wide interaction network characterized by the absence of consensus and a moderately high level of majority opinion. Going back to Gabriel Tarde's classical distinction between opinion, tradition, and reason (1901), we argue that filter bubbles promote the development of a society of traditions, while the sources of serendipitous knowledge nurture a society of opinions. In this perspective, current developments of information technologies produce a tension between tradition and opinion.

At the time of his writing, Tarde [38] witnessed the advent of a society of opinion due to advances in communication technologies. By a twist of history, the relationship seems to be reversed today, because advances in information technology, in this case online filter bubbles, rather seem to favor the establishment of more traditional societies. In this context, the maintenance of a society of opinion and debate seems to rest on the multiplication of sources of serendipitous knowledge, such as third places.

While interesting, one should keep in mind that these results rely on a very simplistic model of opinion changes, in which agents treat the information emanating from their partners as relevant. To be more realistic, one may consider adding opinion leaders, or repulsion mechanisms [22, 24]. Also, we are wondering what if agents were receiving multiple, and possibly contradictory, signals from the third place? These limitations of our model constitute avenues for future research.

Funding Funding was provided by European Research Executive Agency (Grant No. 101061516).

Data availability No Data associated in the manuscript. The computer code of the model is made available freely at the following address: https://www.comses.net/codebases/c240dcc6-1746-4cc4-b7a3-c8630a100418/releases/1.0.0/.

References

1. Arguedas, A. R., Robertson, C. T., Fletcher, R., & Nielsen, R. K. (2022). Echo chambers, filter bubbles, and polarization: A literature review. *The Royal Society*. https://doi.org/10.60625/risj-etxj-7k60
2. Axelrod, R. (1997). The dissemination of culture. A model with local convergence and global polarization. *Journal of Conflict Resolution, 41*(2), 203–226.
3. Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on facebook. *Science, 348*(6239), 1130–1132.
4. Barabási, A. L., & Alberty, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512.
5. Baumann, F., Lorenz-Spreen, P., Sokolov, I. M., & Starnini, M. (2020). Modeling echo chambers and polarization dynamics in social networks. *Physical Review Letters, 124*, 048301.
6. Busch, C. (2024). Towards a theory of serendipity: A systematic review and conceptualization. *Journal of Management Studies, 61*(3), 1110–1151.
7. Carragher, P., Ng, L. H. X., Carley, K. M. (2023). Simulation of stance perturbations. In *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation* (pp. 159–168). https://arxiv.org/pdf/2307.08511. Accessed May 20th 2025.
8. Carley, K. (1991). A theory of group stability. *American Sociological Review, 56*, 331–354.
9. De Groot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association, 69*(345), 118–121.
10. Del Cerro, C. C. (2024). The power of social networks and social media's filter bubble in shaping polarization: And agent-based model. *Applied Network Science, 9*, 1–32.
11. Desmarchelier, B., Djellal, F., & Gallouj, F. (2025). Public libraries as social innovators. *Public Library Quarterly, 44*(1), 1–20.
12. De Zúñiga, H. G., Cheng, Z., & Gonzalez-Gonzalez, P. (2022). Effects of the news find me perception on algorithmic news attitude and social media political homophily. *Journal of Communication, 72*(5), 578–591.
13. Feezell, J. T., Wagner, J. K., & Conroy, M. (2021). Exploring the effects of algorithm-driven news sources on political behavior and polarization. *Computers in Human Behavior, 116*, 106626.
14. Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., & Lorenz, J. (2017). Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*. https://doi.org/10.18564/jasss.3521
15. French, J. R. P. (1956). A formal theory of social power. *Psychological Review, 63*(3), 181–194.
16. Friedkin, N. E. (2012). A formal theory of reflected appraisals in the evolution of power. *Administrative Science Quarterly, X*, 1–29.
17. Geschke, D., Lorenz, J., & Holtz, P. (2019). The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *British Journal of Psychology, 58*, 129–149.
18. Golub, B., & Jackson, M. O. (2010). Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics, 2*(1), 112–149.
19. Helberger, N., Karppinen, K., & D'Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication and Society, 21*(2), 191–207.

20. House of Lords. (2024). The future of news, communications and digital committee, 1st Report of Session 2024–2025, House of Lords.
21. Jacob, D., Banisch, S. (2023). Polarization in social media: a virtual worlds-based approach. *Journal of Artificial Societies and Social Simulation*, 26(3), http://jasss.soc.surrey.ac.uk/26/3/11.html. Accessed May 20th 2025.
22. Jager, W., & Amblard, F. (2004). Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. *Computational and Mathematical Organization Theory, 10*, 295–303.
23. Kaiser, J., & Rauchfleisch, A. (2020). Birds of a feather get recommended together: Algorithmic homophily in youtube's channel recommendations in the United States and Germany. *Social Media + Society, 6*(4), 1–15.
24. Keijzer, M. A., Mäs, M., & Flache, A. (2024). Polarization on social media: Micro-level evidence and macro-level implications. *JASSS, 27*(1), 7.
25. Lankes, R. D., Silverstein, J., & Nicholson, S. (2007). Participatory networks: The library as conversation. *Information Technology and Libraries, 26*(4), 17–33.
26. Le Cun, Y. (2019). *Quand la machine apprend. La révolution des neurones artificiels et de l'apprentissage profond*. Odile Jacob.
27. McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415–444.
28. Mitchell, M. (2020). *Artificial intelligence. A guide for thinking humans*. Picador.
29. Murdock, I., Carley, K. M., Yağan, O. (2023). An agent-based model of reddit interactions and moderation. In *Proceedings of the international conference on advances in social networks analysis and mining* (pp. 195–202). https://dl.acm.org/doi/https://doi.org/10.1145/3625007.3627489. Accessed May 20th 2025.
30. Namatame, A., & Chen, S. H. (2016). *Agent-based modeling and network dynamics*. Oxford University Press.
31. Newman, M. E. J. (2010). *Networks, an introduction*. Oxford University Press.
32. Ng, L. H. X., & Carley, K. M. (2022). Pro or anti? A social influence model of online stance flipping. *IEEE Transactions on Network Science and Engineering, 10*(1), 3–19.
33. Oldenburg, R., & Brissett, D. (1982). The third place. *Qualitative Sociology, 5*(4), 265–284.
34. Pariser, E. (2011). *The filter bubble*. The Penguin Press.
35. Rogers, E. M. (1983). *Diffusion of innovations*. The Free Press.
36. Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1*, 143–186.
37. Sunstein, C. R. (2017). *# Republic. Divided democracy in the age of social media*. Princeton University Press.
38. Tarde, G. (1901). *L'opinion et la foule*. Félix Alcan.
39. Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature, 393*, 440–442.