

WEIZENBAUM JOURNAL OF THE DIGITAL SOCIETY
Volume 4 \ Issue 2 \ w4.2.4 \ 05-27-2024
ISSN 2748-5625 \ DOI 10.34669/WI.WJDS/4.2.4

Information on this journal and its funding can be found on its website:
<https://wjds.weizenbaum-institut.de>

This work is available open access and is licensed under Creative Commons Attribution 4.0 (CC BY 4.0):
<https://creativecommons.org/licenses/by/4.0/>

KEYWORDS

feedback loop
echo chamber
filter bubble
reinforcing process
self-regulating system
algorithm
recommender system

RESEARCH PAPER

Communicative Feedback Loops in the Digital Society

Damian Trilling 

Vrije Universiteit Amsterdam
d.c.trilling@vu.nl

ABSTRACT

Contemporary communication is often characterized using metaphors as “echo chambers” or “filter bubbles.” Despite the popularity of these terms, however, growing concerns about vague definitions are complemented by empirical evidence that contests their widespread existence. Nonetheless, today’s media environment offers ample opportunities for human and/or algorithmic selection processes with undesirable outcomes. To reconcile these two observations, I propose taking a feedback-loop perspective. Such a perspective explains how processes can reinforce themselves without producing catastrophic consequences. This can solve the paradox that while extrapolating from typical filter bubble and/or echo chamber models results in full radicalization within a short period, this outcome has not, despite the long-standing presence of the relevant technologies, become omnipresent. After mapping different types of feedback loops in communication research, I review various empirical approaches, discuss how they can improve research on feedback-loop phenomena, and consider how this can enable us to build better theories of communication in the digital society.

1 Introduction

Seldom do communication theories and models become subjects of media discussion, public debate, and common knowledge. One exception is the concept of the “filter bubble.” Coined by internet activist Eli Pariser in 2011, the term is nowadays mentioned in countless newspaper articles, political debates, and educational contexts. Slightly less popular, but still widely known, is the metaphor of the echo chamber, popularized by law scholar Cass Sunstein (2001). The idea that modern citizens increasingly live in “bubbles” and hear their own voices reflected by “echo chambers” speaks to the imagination of many, and anecdotal evidence is easy to find. Some scholars, however, have argued that these metaphors are ill-defined, misleading, and hence too weak to serve as foundations for communication science theory (e.g., Bruns, 2019; Dahlgren, 2021; Jungherr et al., 2020; Jungherr & Schroeder, 2021; Rau & Stier, 2019). Indeed, both metaphors are “frequently employed as quasi-synonymous terms” (Vaccari & Valeriani, 2021, p. 84), but their emphases are different. For example, Bruns (2019) proposes the following distinction: “An echo chamber comes into being when a group of participants choose to preferentially *connect* with each other, to the exclusion of outsiders. [...] A filter bubble emerges when a group of participants, independent of the underlying network structures of their connections with others, choose to preferentially *communicate* with each other, to the exclusion of outsiders” (p. 29). We may also observe, orthogonally to Bruns’ proposed distinction, that work on echo chambers often emphasizes psychological mechanisms and individuals’ choices, potentially aided by algorithmic filtering, while work on filter bubbles places emphasis almost exclusively on the consequences of algorithmic filtering.

Despite their fuzziness, the popularity of these metaphors makes it difficult to ignore them. Moreover, in them lies some element of truth: People can radicalize one another in small, like-minded social media groups, and algorithmic recommendation systems can influence which news is seen by whom. To solve this impasse, I propose to focus on the underlying dynamic processes: feedback loops.

Readers click on news articles selected by a recommender system and, in doing so, influence future recommendations. Journalists write about what they have read on social media, sparking new discussions on those very platforms. Some social media users have ties to conspiracy circles; therefore, they meet more such people and begin following them, tightening the circle. Such situations can be described as *feedback loops*, as “parts joined so that each affects the other” (Ashby, 1956, p. 54). The input influences the output, which, in turn, influences the input. We know this process from the audio feedback loop, where a microphone too close to a speaker amplifies the noise.

Similarly, “amplification” is used to describe retweeting (Gallagher et al., 2021), the granting of media coverage to messages voiced by extremist groups (Kaiser & Rauchfleisch, 2019), algorithmic recommendations of extremist content (Whittaker et al., 2021), and the general effect of recommendations on preferences (Kalimeris, 2021). Jungherr and Schroeder (2021) summarize the point as follows:

Publics amplify information by publicly interacting and sharing it, digital platforms amplify information by algorithmically deciding which information to push and which to hide, they also allow the originators of information to pay for their content being displayed prominently and thereby amplify it, while news organizations and political actors use multiple channels available to them through their social position and authority to amplify selected information and voices. (p. 5)

Content production and amplification can thus be performed either by the same or by different actors (see also Starbird et al., 2018). Most amplification processes can be described as feedback loops, as sharing usually leads to further sharing. Two mechanisms can explain this, either separately or in conjunction: The (displayed) number of shares can serve as a popularity cue and motivate additional individuals to share the content; and frequently shared content may be ranked more highly by ranking algorithms, leading to increased exposure and thus more opportunities for sharing (Trilling et al., 2022). Crucially, however, and in sharp contrast to commonly held ideas, the existence of such feedback loops does not necessarily imply catastrophic consequences. As we shall see, there are both dampening and reinforcing feedback loops, and these coexist with other processes that, taken together, form a complex system.

The main purpose of this article is to argue that a feedback loop perspective can improve communication research in the digital age. Below, I present a typology of different feedback loops and show how they can help us understand contemporary challenges. I subsequently suggest concrete methodological avenues that reconcile the theoretical and conceptual complexity implied by a feedback-loop approach with empirically feasible research designs and offer suggestions for researchers who wish to expand their empirical toolbox.

2 Feedback Loops in Communication Research

2.1 The History of Feedback Loops

Originating in the discipline of cybernetics in the 1950s (see, e.g., Ashby, 1956), feedback loops are central to complexity science and systems theory. The so-called dynamic-transactional approach (Früh & Schönbach, 1982) explicitly studies the feedback loops between media producers and media recipients, proposing “that it takes two to generate media effects, and that the relationship between the two actors may change during an effects process” (Schönbach, 2017, p. 8).

As early as 1957, Westley and MacLean acknowledged that the concept of *gatekeeping* is not mono-directional: Journalists adapt their selection process based on audience feedback, and sources adapt their communications based on feedback from journalists. Feedback, in this early context, was often indirect and incomplete. In today’s digital society, by contrast, to borrow the vivid phrasing of Shoemaker and Vos (2009), “the dotted line [in Westley and MacLean’s model] representing a weak audience feedback loop in mass communication models can now be made solid” (p. 7). Indeed, Boczkowski and Mitchelstein (2013) describe the interaction between what journalists publish and what users click, share, or comment on as a “dynamic system.”

Figure 1: A system of two simple feedback loops. The content that the journalist delivers is shaped by audience feedback on that content, but also by the feedback that sources (such as interview partners) receive from the journalist.

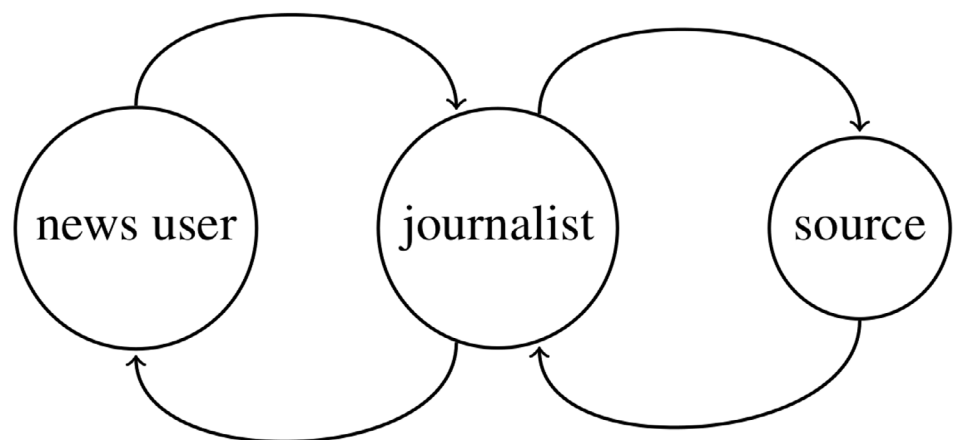


Figure 1 illustrates such a system, containing two simple feedback loops. Questions remain, however, as to the precise nature of the influences and the long-term development of the system. It can be quite an undertaking to formalize such a model and translate it into empirical research. It is, nonetheless, common to do so. Slater (2007), for instance, draws a parallel between cross-lagged panel designs and feedback loops:

Mutually reinforcing processes, as opposed to self-regulating processes, might be expected to spin out of control or move to some extreme value – *a positive feedback loop* [italics original]. It should be noted that a loop is a graphic convenience for illustrating a process ignoring the time dimension; if time-ordered processes are illustrated, a loop would presumably tend to be illustrated instead in a way similar to Figure 1. (p. 288)

Slater's Figure 1 depicts a standard cross-lagged model in which two variables, "media use" and "belief, attitude, behaviour," reinforce each other across three time points.

"Reinforcing processes" that "spin out of control": We may detect here the familiar sound of the filter bubble and/or echo chamber argument. Nevertheless, there are at least three arguments as to why we can accept the idea of feedback loops without also accepting the conclusion that an algorithmic funnel inevitably confines people in partisan filter bubbles, or that people self-sort into increasingly like-minded echo chambers.

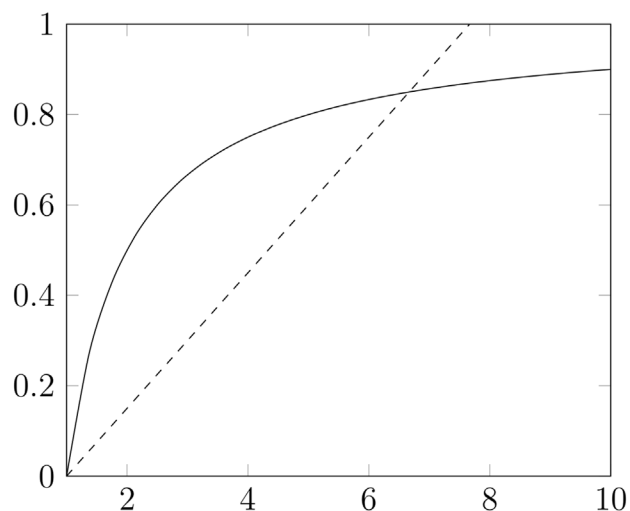
First, alongside *positive*, reinforcing feedback loops, there also exist *negative*, self-regulating feedback loops. The classic example is the thermostat that turns down the heating when a certain temperature is reached. We may also think of the social media user who becomes bored of recommendations that become too similar, or the news site user who misses being surprised (Schoenbach, 2007; Kiddle et al., 2023) by content produced by a journalist who is increasingly driven by audience metrics (see, e.g., Diakopoulos, 2019) – both users will decrease their exposure in favor of other outlets. Hence, many communicative feedback loops are self-regulating: Once we reach extreme outcomes, the relationship will break or even be reversed until acceptable levels are reached again. To be clear, feedback loops can spin out of control, sometimes with disastrous consequences. The Rohingya genocide in Myanmar (Whitten-Woodring et al., 2020) and the January 6 United States Capitol attack (Ng et al., 2023) have both been linked to social media dynamics. However, the existence of such system failures (which caused suspensions of Parler by major technology companies and court cases against Facebook) does not allow us to draw conclusions about the typical behavior of the feedback loops we encounter in daily life.

Second, even though cross-lagged models are often estimated using linear models (which may be a reasonable approximation for a limited number of time points), the underlying process is most likely non-linear. The number of minutes spent on some medium simply cannot increase by x minutes every few months, and there is a limit to how extreme an individual's attitudes can become. Hence, either the effect must be so small that even during an individual's lifetime, the maximum is not reached or, more likely, the effect levels off at some point. Figure 2 depicts a simple linear function and a simple non-linear function that approaches, but never exceeds, a given maximum value. Arguably, the effect of media use on an individual's attitude over time more closely resembles the solid than the dashed line. The first time someone consumes content about some issue or position may have a large impact. When consuming further content, however, the user becomes more certain or knowledgeable, and the effect levels off (Chong & Druckman, 2010). The shape of media effects over time can take many forms: Shehata et al. (2021) distinguish and visualize a wide variety of models – but none of these are linear. We may thus state the following:

Proposition 1: Communicative feedback loops cannot be modeled as a linear relationship over time. Once a reinforcing relationship leads to extreme outcomes, the relationship will break or be reversed.

This proposition inspires research questions and hypotheses that can be empirically tested, such as how quickly (and indeed whether) such an extreme is reached. One may hypothesize that some feedback loops operate over such a large time horizon that linear approximations remain appropriate for some practical purposes, while others may be much more short-lived. A slow feedback loop may be the relationship between quarterly circulation figures and the editorial line taken by a newspaper; a rapid feedback loop may be the interaction between a TikTok user and the platform's algorithms.

Figure 2: The linear function $y = 0.15 \times (x-1)$ (dashed) and the non-linear function $y=1-\frac{1}{x}$ (solid) plotted for $x=[1..10]$.



Third, it is important to acknowledge that feedback loops are not the only forces of interest in play. Slater (2007) points to “competing social, psychological, and environmental influences” (p. 288) that ensure that the system is not fully “closed.” For instance, the introductory example sketched in Figure 1 can be extended through the inclusion of multiple users, journalists, and sources, but also through the addition of organizational constraints (such as editorial policies). If we focus on a journalistic example, many (but not all) of the forces described in Shoemaker and Reese’s (2014) hierarchy-of-influences model can be seen as exogenous (i.e., they are not affected by the output); but they do play a role in explaining the overall behavior of the system. This complexity can set feedback-loop-based models apart from echo chamber and filter bubble models, which often involve only two (types of) actors, such as the algorithm and the user.

2.2 Types of Feedback Loops in the Digital Society

The digitalization of today’s society has spurred a revived interest in feedback loops. Feedback loops have been used to describe media hypes (Vasterman, 2005), the redistribution of messages in different media (Hewett et al., 2016), algorithmic news recommendations (Bodó et al., 2019; Jiang et al., 2019; Spinelli & Crovella, 2017), the interaction between news production and audience metrics (Hagar & Diakopoulos, 2019), and the relationship between populist actors, citizens, and the media (Reveillac & Morselli, 2022).

As this illustrates, in the digital society, feedback loops are omnipresent and occur between and within very different actors and on different levels. For instance, Slater’s (2007, 2015) reinforcing spiral model is largely concerned with cognitive processes and concerns the interrelation of media exposure, social identities, and attitudes. Such processes are increasingly important in online environments, but approaches focusing on cognitive approaches often do not explicitly address the role of algorithmic actors (such as recommendation systems or filtering and ranking algorithms). On the other hand, Pagan et al.’s (2023) classification of feedback loops integrates human and algorithmic actors, but the model is tailored toward a perspective of algorithmic fairness in automated decision-making.

Unfortunately, these literatures are fairly disjointed. Here, I take an integrative perspective and argue that many problems faced by the digital society can be formulated as research questions involving *communicative feedback loops* – a point that, as I will show, opens avenues for empirical research. I propose a categorization that can help identify similarities and differences between types of feedback loops. In particular, across these types, we see different mechanisms at play: These can include social–cognitive mechanisms such as the desire to make identity-confirming choices; sociological mechanisms that explain, for example, how the role of a journalist can require acting according

to the organization’s need to respond to market mechanisms; or technical arguments about the behavior of a specific metric or algorithm. Table 1 provides an overview, upon which I will elaborate in the following sections.

Table 1: Different (ideal) types of communicative feedback loops.

	human feedback loops		algorithmic feedback loops	
	within-human feedback loop	between-human feedback loop	human–algorithm feedback loop	pure algorithmic feedback loop
involved actors	1 human	≥2 humans	≥1 human and ≥1 algorithm	≥2 algorithms
research drawing on	(social–) psychological mechanisms	sociological and/or social–psychological mechanisms	multidisciplinary approaches, e.g., human–computer interaction	computer science
practical example	selection of attitude-conforming media content ↔ attitude reinforcement	content creators adapting their content to user feedback	radicalization due to recommendation of increasingly extreme content	two chatbots talking to each other; AI trained on AI-generated content
example of possible theoretical foundation	reinforcing spiral model (Slater, 2007, 2015)	gatekeeping (with audience feedback; Westley & MacLean, 1957; Shoemaker & Vos, 2009); dynamic-transactional approach (Früh & Schönbach, 1982)	algorithmic bias literature (e.g., Pagan et al., 2023)	algorithmic bias literature (e.g., Pagan et al., 2023)

Human Feedback Loops

By *human* feedback loops, I refer to all feedback loops that are not introduced by an automated actor. Similarly, Zuiderveen Borgesius et al. (2016) refer to “self-selected personalization” when focusing on human choices, and “pre-selected personalization” when an algorithm is involved. We can further distinguish between feedback loops *within* humans and *between* humans. Within-human feedback loops take place, in essence, within a person, with a prime example being the “selective exposure” concept of a vicious circle of ever-increasing polarization of an individual’s attitudes and ever-increasing selectivity in their media use (e.g., Stroud, 2008). The two elements of the loop thus reinforce one another. Relatedly, Bail (2021) observes a “feedback loop between identity, status seeking, and political polarization” (p. 122). Hence, I propose the following:

Definition 1: A within-human feedback loop occurs when human behavior affects human opinion, which, in turn, affects the behavior in question.

Between-human feedback loops also flourish in the digital society. The feedback loop of a politician's acting based on the output of a journalist, who then writes about the politician's action, and so on (e.g., Van Aelst et al., 2016), is not new. In a digital society, however, in which politicians and journalists follow one another closely on social media, this loop is accelerated. As the "inner circle [of political journalists on Twitter] comprises an elite of fellow journalists, news outlets, and politicians" (Bruns & Nuernbergk, 2019, p. 206), and as the frequency, velocity, and number of messages exchanged has increased, the opportunities for the creation of feedback loops in which politicians act based on what journalists do and vice versa have increased dramatically. I therefore propose the following:

Definition 2: A between-human feedback loop occurs when the behavior of a human or a group of humans affects the behavior of another human or group of humans and vice versa.

Naturally, within-human and between-human feedback loops can occur in combination in a more complex system, and the number of actors involved is, in any case, not restricted to two. Depending on how fine-grained an analysis aims to be, it is also possible to aggregate humans into groups and study, for instance, feedback loops between the categories of "journalists" and "politicians."

Algorithmic Feedback Loops

While one can easily conceive of feedback loops between two algorithmic actors (consider, for example, two chatbots talking to one another), research on the digital society usually implies that at least one party in the exchange is human. The typical example of an algorithmic feedback loop is the classic "filter bubble" account: A recommendation system learns an individual's preferences from their previous choices and offers more similar content, further strengthening those preferences, and so on. In the categorization scheme of Pagan et al. (2023), this would be termed an individual feedback loop within an automated decision-making system. A prominent example is the case of YouTube, which recommended increasingly extreme content (Ribeiro et al., 2020) and channels (Kaiser & Rauchfleisch, 2020), leading to the radicalization of some users. Whittaker et al. (2021) have confirmed this for YouTube but did not find similar feedback loops in operation on Reddit and Gab. Other examples are situated within an algorithmic system. For instance, Pagan et al. (2023) describe model feedback loops and feature feedback loops, in which the outcome of a system (such as a news recommender) influences the future development of the model itself or the future features of the model. I propose the following:

Definition 3: An algorithmic feedback loop occurs when a feedback loop exists between an algorithm and a human, another algorithm, or the same algorithm.

Without downplaying the danger of recommendation systems' promoting extremist content, we must be aware that it is invalid to conclude that algorithmic feedback loops necessarily lead to exposure to increasingly extreme (or even solely more-of-the-same) content over time for two reasons: (1) the technical design of typical recommendation systems and (2) the way in which humans typically interact with them.

First, the idea that contemporary news recommendation systems essentially present a user only with material that exactly matches their preferences is more of a caricature than an accurate picture. Overly similar recommendations lead to a decrease in user satisfaction and a lock-in from which the system cannot recover (see, e.g., Loecherbach & Trilling, 2020; Vrijenhoek et al., 2021). Contemporary systems, therefore, often introduce randomness or serendipity or limit the volume of personalized recommendations. In fact, news recommendation algorithms may not even aim solely at optimizing clicks, as news organizations may incorporate other (journalistic) goals and values in their systems (Bastian et al., 2021; Diakopoulos, 2019; Lu, Dumitrache, & Graus, 2020; Vrijenhoek et al., 2021). Kiddle et al. (2023), for instance, propose the concept of "navigable surprise" to balance the aim of recommending similar news content with the aim of enabling serendipitous discovery. Additionally, there exists only a certain number of articles on a given topic and with a given stance, which may explain why recommender systems on news sites create diverse outputs (Möller et al., 2018).

Second, even when a recommendation system promotes particular content, the user may not comply. Unless we consider a (hypothetical) platform where each and every piece of recommended content exactly matches the inferred user preference, the user is likely to correct the system by clicking one of the items that are not recommended. When people receive more personalized content (for instance, they consume content on sports and hence receive more recommendations on sports), they will, at some point, lose interest and click on something else, signaling to the system that it should reconsider offering them mainly sports content (Loecherbach et al., 2021). Therefore, the system may increase exposure to material matching the user's preferences but will learn over time when it is "overdoing it," because the click rate of the "preferred" items will decrease. We thus observe a negative feedback loop, just like the thermostat considered earlier: The system increases the energy used for heating (or the number of sports items recommended), but once the temperature (the amount of sports content) becomes too high, it will decrease the energy used until the temperature drops too low, from which point the energy will be increased again. The below thus follows:

Proposition 2: Algorithmic communicative feedback loops result in some equilibrium, which can prevent serving content that is too close to or too far from a user's preferences.

This proposition leads directly to relevant research questions and hypotheses. For instance, we may ask the following: While the position occupied on the continuum will vary, it is unlikely that a well-functioning system will settle at either of the extremes – but how must systems be designed to find the “sweet spot”? Loecherbach et al. (2021) describe such an ideal point in the form of a personal “diversity corridor” (p. 283) in which algorithmic recommendations for a user should fall so that they are not so distant from the user's preferences as to be irrelevant nor so close as to constitute only more-of-the-same (see also Kiddle et al., 2023). Researchers may aim to define and/or estimate such an optimal level of personalization according to some normative goal (see Vrijenhoek et al., 2022).

The Interplay of Human and Algorithmic Loops

We have seen that algorithmic feedback loops and human behavior are not independent, and in that sense, real-world phenomena will not always fit neatly into the columns of Table 1. Humans and algorithmic systems often interact. A simple example is described by Diakopoulos (2019), in which three (groups of) actors interact: a news bot that automatically tweets when government agencies edit Wikipedia pages, media that report on this, and the government, which reacts by editing less. This leads to a feedback loop in which the news bot tweets less frequently and the issue hence becomes less newsworthy for the media.

We can also think, however, of situations in which human and algorithmic loops are more difficult to disentangle. A particularly relevant example is so-called *collaborative filtering*. With this technique, the recommendations that an individual receives are not only based on the content that that individual has previously viewed and/or some other personal characteristics but on the preferences of other individuals, whom the user most likely does not even know. This can lead to an interesting but unintended effect:

[F]or users with multiple interests that are not jointly common in the general population, the recommendation system cannot jointly recommend all the items in which they are interested; the system may instead only recommend them items corresponding to one interest. We note that this problem may especially affect cultural minorities. (Guo et al., 2021, p. 1)

We can easily see, then, that the number of actors exerting an influence on the composition of an individual's news diet can grow extremely large; and again, these effects work in two directions, as the changed composition again influences users' choices, which again have an effect on the recommendation engine. For a given individual, these effects may be tiny, but the complex nature of the system may cause substantial patterns to emerge at the system level.

As an at least equally complex example, let us consider the recommendation of accounts for a user to connect with on social media. We speak of between-human feedback loops when people follow one another on social media and act based on their interactions there. In this case, however, the very connection may be algorithmically suggested to the user based on their similar behavior. A similar question concerns the downstream effects of algorithmic suggestions for content to post, be that a suggestion of what to type or which link to share.

At some point, we must admit that – quite frustratingly – it becomes impossible to disentangle the exact role of humans and algorithms. Healy (2017) points out that all theory must abstract from the complex world because too much nuance renders a theory useless. At the same time, however, it is crucial to acknowledge the interdependence of humans and algorithms to avoid simplistic ideas about the functioning of algorithms, which could lead us to draw invalid conclusions. The new field of computational social science has embraced the older notion of “emergence” (see Conte et al., 2012) and highlights the challenges and opportunities of studying emergent behavior on the aggregate level, which results from the interaction of complex individual-level systems. Relatedly, González-Bailón (2017) describes how communication in the digital society can have substantial and impactful unintended consequences that can follow – in a non-deterministic way – from very minor differences that are amplified by feedback processes. The below thus follows:

Proposition 3: In digital communication environments with multiple feedback loops, it is difficult or even impossible to infer the emergent behavior of the system as a whole from the workings of its parts alone.

2.3 The Value of a Feedback Loop Perspective for Contemporary Problems

The Spread of Mis- and Disinformation

Tsfati et al. (2020) observe an interesting paradox: Only a minority of the population visits “fake news websites,” yet many are aware of specific so-called fake news stories, probably due to the “role played by mainstream news media in the dissemination of disinformation” (p. 170). Media reporting about rumors circulating on social media are quoted on social media (Andrews et al., 2016),

further accelerating those rumors' dissemination. Apparently, then, misinformation is not confined to some "bubble" in which it circulates but rather spreads as supporters of extreme ideologies "actively exploit the very absence of echo chambers and filter bubbles [...] to ensure that these messages travel far beyond the partisan in-group of the already converted" (Bruns, 2019, p. 108).

A feedback-loop-based approach can help us model this: An item of mis- or disinformation is picked up by an individual (critically or not), which generates further attention, leading to further dissemination, and so on. Benkler et al. (2018) offer the model of a "propaganda feedback loop" between public, politicians, and the media, which – based on the tendency of all actors to strive for identity-confirmation – leads to "a steady flow of bias-confirming stories that create a shared narrative of the state of the world; a steady flow of audiences, viewers, or clicks for the outlets; and a steady flow of voters highly resilient to arguments made by outsiders on outlets that are outside the network" (p. 80). Nevertheless, the authors find evidence in left-wing and non-partisan media in the US for negative, self-correcting feedback loops that "dampen and contain partisan statements that are demonstrably false" (p. 75), while on the right, they identify positive feedback loops with "susceptibility to information cascades, rumor and conspiracy theory, and drift toward more extreme versions of itself" (p. 73). I propose, therefore, the following:

Proposition 4: The spread of mis- and disinformation can at least partly be explained by the operation of a feedback loop in which attention drawn to it, even if meant to criticize it, creates more attention.

Radicalization of Fringe Groups

Throughout this article, we have observed another paradox: On the one hand, there is a lack of evidence for omnipresent "bubbles"; on the other hand, radicalized homophilic groups undeniably exist. Möller (2021) refers to these as "fringe bubbles," a fringe (rather than widespread) phenomenon that also occurs at the fringes of society. Such bubbles can be occupied by extremists with radical ideas and create a "spiral of noise" that suggests – incorrectly – to the bubble's inhabitants that they are in the majority, thereby causing them to become more vocal. This effect could then increase support for their position – a feedback loop. Focusing instead on distorted perceptions of the other side's radicalism, Bail (2021) describes a similar phenomenon as "feedback loops of extremism" (p. 67).

Kaiser and Rauchfleisch (2019) argue that fringe groups such as far-right extremists use outward-facing extreme communication to gain media attention. The message involved will often be too extreme to convince any out-group members, but its very extremism makes it newsworthy. This can reinforce in-group identity, fuel the process of radicalization, and cause even more extreme messages, a "vicious cycle of extremism" (p. 241) or "vicious cycle of coun-

terpublics” (p. 249). In the terminology of the present work, this constitutes a feedback loop between fringe group members and journalists. I therefore propose the following:

Proposition 5: Fringe groups have an incentive to express radical positions and exaggerate the level of support for them, which creates a feedback loop that leads to further radicalization.

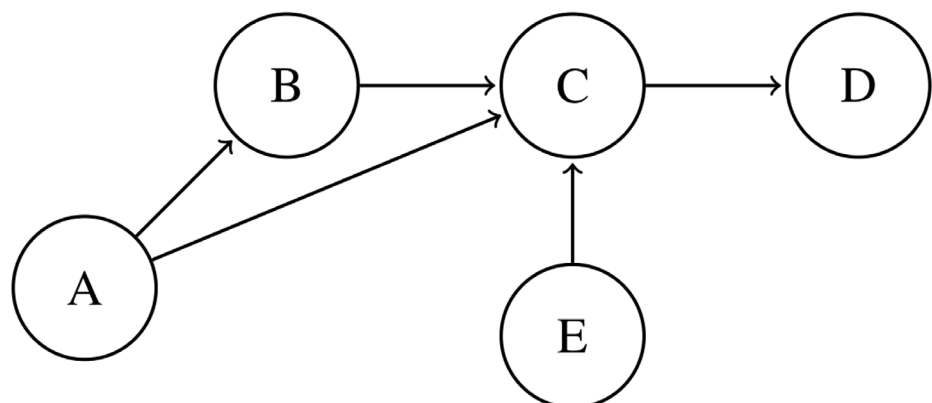
Studying Feedback Loops Empirically

I hope to have shown that a feedback loop perspective can be beneficial from a theoretical and conceptual point of view. Readers may, nonetheless, wonder how this can be translated into feasible empirical studies. Let us assume that we have drawn a model (mentally or with a pen and paper) of feedback loops to study a communication phenomenon. How can we test this model? In causal modeling, so-called directed acyclic graphs (DAGs) allow the systematic assessment of the role of different variables in a process (Figure 3). DAGs have strict rules as to which arrows are allowed in making causal claims. Most notably, DAGs are indeed acyclic: Feedback loops are explicitly forbidden.

Consequently, we cannot use a causal inference framework. Even the less strict assumptions inherent in survey research (as manifested in the language of “predictor” and “outcome” or “independent” and “dependent” variables) may not be adequate. Nonetheless, we do, of course, wish to investigate the role (or, for that matter, the “effect”) of the elements included in our model.

I wish to highlight three approaches that may help. In particular, I will show that it is not necessary to have, say, a comprehensive dataset pertaining to some social media platform, nor the most advanced computing resources, to take a feedback loop approach.

Figure 3: A directed acyclic graph (DAG). If we were to insert an additional arrow from E to B, the graph would no longer be acyclic: The B influences C, but would then, in turn, be influenced by the output of C (via E)—a feedback loop.



Simulation

The first suggestion relates to uncovering non-linear relationships and gauging whether positive or negative feedback loops prevail and under which circumstances an equilibrium is reached. If we consider the study of feedback loops in the context of social networking sites, it is often not possible to collect all of the data that would be needed to conduct an observational study – particularly if one seeks to measure not only behavior but also opinions. Simulation approaches, which are especially suited to the study of complex systems (Conte et al., 2012), can be a solution. Waldherr and Wettstein (2019) see agent-based modeling (ABM) as an opportunity to formalize theories and test which mechanisms explain the observed data. For instance, this would allow the study of which levels of attitude-congruent and attitude-incongruent exposure would be needed, based on different assumptions as to how people change their minds, to actually lead to outcomes such as polarization or fragmentation. Using this approach, Geschke et al. (2019) investigate how filtering processes on different levels could lead to filter bubbles, and Tambuscio et al. (2016) study the spread of misinformation. ABM can also be used to extend previous theories: Song and Boomgaarden (2017) find that their models align with Slater’s reinforcing spiral model, while at the same time adding the suggestion that the processes involved are amplified or dampened by interpersonal discussion networks. Beyond ABM, Möller et al. (2018) use a simulation approach by comparing the output of multiple news recommender systems, Bountouridis et al. (2019) present the development of software for such simulations, and Perra and Rocha (2019) simulate opinion dynamics affected by algorithmic filtering.

Simulation allows us to understand effects and dynamics based on a “what if” approach: How does diversity of news exposure at time t_1 look when a news recommendation algorithm at t_0 is personalizing slightly more aggressively? Crucially, there is no assumption of linear, or even monotonic, effects – it is possible, for instance, that overly aggressive personalization may reverse initial effects.

Simulations can inspire us to rethink theoretical assumptions. Above, for instance, I quoted Bruns (2019), who suggests that “the very *absence* of echo chambers” (p. 108) enables the spread of problematic messages. Törnberg (2018), however, shows, based on simulations, that in a scenario of “complex contagion,” a message can first be reinforced locally (through, in Törnberg’s words, an “echo chamber”), before, having achieved sufficient spread within the group, it can reach a wider audience. Some might object that simulations suffer from low ecological validity. This need not, however, be the case, as models can be compared to empirical data (as in Mønsted et al., 2017) – which, as Waldherr and Wettstein (2019) argue, they should be. The inclusion of empirical data can be as straightforward as setting model parameters (such as the likelihood of opinion changes) based on evidence from other studies. Other forms, however, are also possible: Törnberg et al. (2023), for instance, use a Large Language Model to create agents that mimic the behavior of respondents to an existing survey. Importantly, the use of simulation approach-

es – while still relatively uncommon in communication science – does not require huge resources: Appropriate software is readily available, as is training (Waldherr & Wettstein, 2019).

Algorithm Audits and Agent-Based Testing

While simulation approaches excel in formalizing theories and studying emergent behavior, other researchers may be interested in understanding specific real-life platforms. For instance, Van Hoof et al. (2024) aim to find out how repeatedly searching for specifically framed political information influences search results at later points in time, which could provide valuable insights into the supposed human–algorithmic feedback loop. However, the inner workings of (in this case) Google are unknown, so simulation studies are of limited use, at least as an initial step. Instead, researchers can perform so-called *algorithm audits* to systematically understand the effects of systems that include feedback loops by observing what happens over time. Of course, once this has been done, the resulting insights could later be reused to model agents in an offline simulation.

In its most simple form, the algorithm audit can be performed manually, with humans simply instructed to use a specific system and record the output: For instance, to see whether YouTube recommendations indeed become more extreme over time, we could ask individuals to watch YouTube videos, click on the recommendations according to a scheme devised for this purpose, and then see whether, depending on the scheme, the recommendations change. Such experiments can, however, easily be automated. For example, by combining a crowd-based approach (individuals taking screenshots) with a sock puppet approach (computer programs acting as humans), Bandy and Diakopoulos (2020) were able to gain an understanding of the curation mechanisms of Apple News.

Haim (2020) refers to such sock puppet approaches as “agent-based testing,” which he contrasts with “agent-based modeling,” as the former “is not meant to model behavioral interactions but rather to test various inputs for their respective software outcome within algorithmically curated media environments” (p. 901). As such approaches mimic human browsing behavior while at the same time allowing for systematic variation of that behavior, they can help uncover the outcomes of feedback-loop-based systems: If a system serves content based on previous clicks, then systematic variation of clicking patterns can help uncover the underlying process.

This even holds true if the next click depends on the output of the system. For researchers who wish to employ such approaches, tools are freely available. For instance, Haim (2020) developed an open-source toolkit for agent-based testing, available at <https://github.com/MarHai/ScrapeBot/>.

From Large N and Small T to Small N and Large T

We have discussed two approaches that allow us, respectively, to formalize models and study emergent behavior (simulations) and to better understand the behavior of specific algorithmic systems in a feedback-loop scenario (algorithm audits). Both are excellent for answering “what if” questions. Some questions, however, require human participants to occupy a more central role: What if, indeed, we wish to observe human behavior in the real world, without making assumptions about how humans behave in social interactions or their interactions with an algorithmic system?

This is possible, although to capture dynamics over time, one must move somewhat beyond classic experimental designs. Even extensive panel surveys consist of only a handful of waves, but to disentangle the close interaction between, for instance, a news recommender system and individuals’ media use, we need many more measurements over time (T). Studying feedback loops requires increasing T , even at the expense of the number of participants (N). The experience sampling method (ESM), data donations, and online field experiments can support this.

ESM studies involve sending participants regular prompts (e.g., multiple times a day) on their cell phones, asking one or a few very short questions or asking for a screenshot (Ohme et al., 2021). While ESM has not yet been applied to study feedback loops specifically, it has been used to understand, for instance, exposure to news on Facebook (Karnowski et al., 2017). Multiple commercial services are available that can provide the infrastructure for ESM studies. However, due to the burden imposed on participants, ESM studies are typically not run over more than a few days or weeks: Thus, only feedback loops expected to occur over a very short period can be studied in this way. Additionally, we know from data donation studies (see below) that behavior such as using search engines to look for news (van Hoof et al., 2023) happens quite rarely, with the result that ESM studies simply do not include enough observations per person to study feedback loops between, for example, search engines, political content, and attitude.

In data donation studies, participants are asked to donate data, such as the results of queries made using their devices or their browser histories. Typically, these data are available on the scale of months, if not years. Such studies have resulted in insights that other methods would not allow. For instance, one analysis of Google search results donated by 4,000 people shows how small the personalization of these results is (Krafft et al., 2019; Puschmann, 2019). The European GDPR law gives every citizen the right to request the data an organization holds about them in a machine-readable format. Asking users to donate these data (or a cleaned and anonymized subset of them) offers new opportunities for researchers (Araujo et al., 2022; Breuer et al., 2020). Even for a moderate N , the extremely high T available in many cases may allow

the reconstruction of typical feedback loops over time. As with agent-based testing, software is available to this end. For instance, Araujo et al. (2022) published the source code of the data donation tool they developed, OSD2F, available at <https://github.com/uvacw/osd2f/>; another option is PORT, developed by Boeschoten et al. (2023) and available at <https://github.com/eyra/port>. It must be noted, however, that while such tools are very user-friendly for participants, installing and configuring them requires technical expertise.

Field experiments can achieve a similar goal. For instance, Loecherbach and Trilling (2020) developed a news app that displays real-time news from real journalistic sources but enables the researcher to implement different recommendation systems. This allows the researcher, for example, to understand negative feedback loops, such as participants' no longer clicking on recommended items if they have been offered too much similar content previously (Loecherbach et al., 2021). However, as news exposure is highly habitual, it can be challenging to find participants willing to use the researcher's news app instead of their own app for a considerable period, even if they are satisfied with the experimental app and are appropriately compensated. In another study on the relationship between diversity of news exposure and a recommender system, Heitz et al. (2023) built a mobile app, available through an app store, to make participation easier and more natural for participants.

Field experiments can also be embedded in existing platforms, for instance by deploying a news bot on social media and studying how users interact with it (e.g., Gómez-Zarà & Diakopoulos, 2020); alternatively, they can be deployed using a chatbot, for instance using the toolkit developed by Araujo (2020), which is available for free at <https://github.com/uvacw/cart>. It must be noted that setting up field experiments with recommender systems is not a trivial task, but researchers do not have to start from scratch. Loecherbach and Trilling's (2020) app, for example, can be downloaded and adapted for other experiments from <https://github.com/ccs-amsterdam/3bij3/>.

Following this discussion of some of the approaches that researchers interested in applying a feedback-loop perspective to their empirical research may consider, Table 2 offers a simplified summary that may help guide their decisions.

Table 2: Comparison of different approaches to studying feedback loops.

	Simulation	Agent-Based Testing	Experience Sampling	Data Donation	Field Experiment
Purpose	Formalizing theories and studying emergent behavior	Understanding behavior of an algorithmic system	Studying short-term feedback loops	Obtaining digital trace data to study long-term dynamics in individuals' media use	Testing human-algorithmic feedback loops
Example	Simulating opinion change dynamics	Understanding how Google reacts to search queries	Understanding relationship between media use and other behaviors	Studying how YouTube use influences subsequent YouTube use	Testing how users interact with different recommender systems
Problem Addressed	Allows manipulation of variables that cannot be (ethically, feasibly, etc.) manipulated with real humans	Allows study of systems outside of researchers' control	Much higher temporal resolution than panel surveys	Data can be collected retrospectively and have very high ecological validity; maximum temporal resolution	Allows full experimental control
Drawbacks	Many assumptions necessary	Assumptions on behavior of the agents necessary	Limited to short-term processes	No experimental intervention possible; only behavioral – no opinion data or similar available over time	Hard to recruit participants for long-term studies; platform may still be perceived as artificial
Easy to Implement?	Yes; multiple software packages and tutorials available	Software exists, but setup can be non-trivial; once set up, relatively easy	Yes; commercial services available	Software exists, but setup is non-trivial	Depending on planned design, own development work necessary; building on existing code possible

3 Conclusion

This article began with two observations. First, despite the prevalence of the metaphors of “echo chambers” and “filter bubbles” in both public discourse and academic research, they do not offer an accurate description of how communication in today’s digital society operates. Second, and to some degree in contrast, there is some truth in these metaphors: Increased opportunities for selective exposure, as well as the increased use of algorithmic systems, can indeed lead to reinforcing effects that are undesirable from a normative point of view.

I proposed to reconcile these two observations by conceptualizing the processes involved as feedback loops. Feedback loops allow us to develop a testable model that encompasses both echo chambers and filter bubbles and thus renders these metaphors moot.

One may object that while my observations may be true, my argument for the study of feedback loops may result – to again follow Healy (2017) – in theoretical models that are too nuanced to produce testable predictions of practical relevance. However, this is not the case. First, an awareness of the broader context can help in avoiding pitfalls and inappropriate shortcuts (such as wrongly assuming that a reinforcing relationship between a recommender system and users’ opinions necessarily leads to a filter bubble). Second, even a complex model can generate testable predictions. For instance, it is perfectly possible to predict whether, given the feedback loops that we hypothesize, changing a given parameter of a system will have an effect or not and if so, in which direction. Third, understanding the characteristics of feedback loops, such as their typical non-linearity, can actually help us make better predictions. Fourth, as we have seen also in the discussion of how Figure 1 might be extended, we can apply a feedback-loop perspective at various levels of nuance and still decide to simplify reality.

One may also object that the feedback-loop concept is so broad that it ultimately encompasses almost all forms of communication. To some extent, this is true, but there remain plenty of examples where the feedback component of a process is absent or negligible for practical purposes. For instance, the current rapid advancements in the field of generative AI affect communication (consider, for instance, the debate concerning the role of deepfakes), but the reverse is not necessarily true. Similarly, in scenarios with restricted choice (for example, there may not be multiple articles that discuss the same topic but take a different stance), one may find media effects of exposure on attitudes, but not the reverse.

In this article, I have offered suggestions of empirical avenues that researchers might consider taking to explicitly apply a feedback-loop perspective in their work. Which of these to choose will differ depending on what types of feedback loops one wishes to study, and some are easier to implement than others. Nonetheless, we may need to acknowledge that communication in the digital society is messy, and so are (necessarily) many research designs. Unless one is willing to sacrifice a great degree of ecological validity, it is often impossible to achieve clean research designs that fully isolate causes and effects. We may take the example of news sharing, which depends partly on an article's features, partly on user preferences, and partly on the simple fact of previous shares. Fully isolating the individual effects may be impossible, but by combining a study on the shareworthiness of news items with an algorithm audit to determine how far a given feature of an item influences the system's outcome, we can make reasonable headway. If it turns out that in a given system, neither the number of shares nor the presence of the feature has a substantial influence on whether the item is shown to others, then one has good reason to argue that the observed effect is mainly attributable to user preferences. One could also use the parameters identified in these studies as inputs for a simulation study to further explore the workings of the system.

These suggestions, I believe, can help foster studies that allow us to develop the theoretical frameworks we need to understand communication in the digital society. These will have to incorporate feedback loops and will present us with new challenges when we think of media "effects." Nonetheless, as I hope to have shown, this is a necessary step to understand a media system in which humans and algorithms interact in networks and relationships that are increasingly difficult to disentangle. Working toward a research agenda that aims at developing and testing hypotheses based on the propositions and empirical avenues I pose in this article may be a first step.

References

- Andrews, C., Fichet, E., Ding, Y., Spiro, E. S., & Starbird, K. (2016). Keeping up with the tweet-dashians: The impact of 'official' accounts on online rumoring. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 27, 452–465. <https://doi.org/10.1145/2818048.2819986>
- Araujo, T. (2020). Conversational Agent Research Toolkit: An alternative for creating and managing chatbots for experimental research. *Computational Communication Research*, 2(1), 35–51. <https://doi.org/10.5117/CCR2020.1.002.ARAU>

- Araujo, T., Ausloos, J., van Atteveldt, W., Loecherbach, F., Moeller, J., Ohme, J., Trilling, D., van de Velde, B., de Vreese, C., & Welbers, K. (2022). OSD2F: An open-source data donation framework. *Computational Communication Research*, 4(2), 372–387. <https://doi.org/10.5117/CCR2022.2.001.ARAU>
- Ashby, W. R. (1956). *An introduction to cybernetics*. Chapman & Hall.
- Bail, C. (2021). *Breaking the social media prism: How to make our platforms less polarizing*. Princeton University Press.
- Bandy, J., & Diakopoulos, N. (2020). Auditing news curation systems: A case study examining algorithmic and editorial logic in Apple News. In *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1), 36–47. AAAI. <https://doi.org/10.1609/icwsm.v14i1.7277>
- Bastian, M., Helberger, N., & Makhortykh, M. (2021). Safeguarding the journalistic DNA: Attitudes towards the role of professional values in algorithmic news recommender designs. *Digital Journalism*, 9(6), 835–863. <https://doi.org/10.1080/21670811.2021.1912622>
- Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.
- Boczkowski, P. J., & Mitchelstein, E. (2013). *The news gap: When the information preferences of the media and the public diverge*. MIT.
- Bodó, B., Helberger, N., Eskens, S., & Möller, J. (2019). Interested in diversity: The role of user attitudes, algorithmic feedback loops, and policy in news personalization. *Digital Journalism*, 7(2), 206–229. <https://doi.org/10.1080/21670811.2018.1521292>
- Boeschoten, L., De Schipper, N. C., Mendrik, A. M., Van der Veen, E., Struminskaya, B., Janssen, H., & Araujo, T. (2023). Port: A software tool for digital data donation. *Journal of Open Source Software*, 8(90), Article 5596. <https://doi.org/10.21105/joss.05596>
- Bountouridis, D., Harambam, J., Makhortykh, M., Marrero, M., Tintarev, N., & Hauff, C. (2019). SIREN. In *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19* (pp. 150–159). ACM. <https://doi.org/10.1145/3287560.3287583>
- Breuer, J., Bishop, L., & Kinder-Kurlanda, K. (2020). The practical and ethical challenges in acquiring and sharing digital trace data: Negotiating public-private partnerships. *New Media & Society*, 22(11), 2058–2080. <https://doi.org/10.1177/1461444820924622>
- Bruns, A. (2019). *Are filter bubbles real?* Polity.

- Bruns, A., & Nuernbergk, C. (2019). Political journalists and their social media audiences: New power relations. *Media and Communication*, 7(1), 198–212. <https://doi.org/10.17645/mac.v7i1.1759>
- Chong, D., & Druckman, J. N. (2010). Dynamic public opinion: Communication effects over time. *American Political Science Review*, 104(4), 663–680. <https://doi.org/10.1017/S0003055410000493>
- Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., Loreto, V., Moat, S., Nadal, J. P., Sanchez, A., Nowak, A., Flache, A., San Miguel, M., & Helbing, D. (2012). Manifesto of computational social science. *European Physical Journal: Special Topics*, 214(1), 325–346. <https://doi.org/10.1140/epjst/e2012-01697-8>
- Dahlgren, P. M. (2021). A critical review of filter bubbles and a comparison with selective exposure. *Nordicom Review*, 42(1), 15–33. <https://doi.org/10.2478/nor-2021-0002>
- Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.
- Früh, W., & Schönbach, K. (1982). Der dynamisch-transaktionale Ansatz. Ein neues Paradigma der Medienwirkungen. *Publizistik*, 27, 74–88.
- Gallagher, R. J., Doroshenko, L., Shugars, S., Lazer, D., & Foucault Welles, B. (2021). Sustained online amplification of COVID-19 elites in the United States. *Social Media + Society*, 7(2). <https://doi.org/10.1177/20563051211024957>
- 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 Social Psychology*, 58(1), 129–149. <https://doi.org/10.1111/bjso.12286>
- Gómez-Zarà, D., & Diakopoulos, N. (2020). Characterizing communication patterns between audiences and newsbots. *Digital Journalism*, 8(9), 1093–1113. <https://doi.org/10.1080/21670811.2020.1816485>
- González-Bailón, S. (2017). *Decoding the social world: Data science and the unintended consequences of communication*. MIT.
- Guo, W., Krauth, K., Jordan, M. I., & Garg, N. (2021). The stereotyping problem in collaboratively filtered recommender systems. In *EAAMO '21: Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Article 6) <https://doi.org/10.1145/3465416.3483298>
- Hagar, N., & Diakopoulos, N. (2019). Optimizing content with A/B headline testing: Changing newsroom practices. *Media and Communication*, 7(1), 117–126. <https://doi.org/10.17645/mac.v7i1.1801>

- Haim, M. (2020). Agent-based testing: An automated approach toward artificial reactions to human behavior. *Journalism Studies*, 21(7), 895–911. <https://doi.org/10.1080/1461670X.2019.1702892>
- Healy, K. (2017). Fuck nuance. *Sociological Theory*, 35(2), 118–127. <https://doi.org/10.1177/0735275117709046>
- Heitz, L., Lischka, J. A., Birrer, A., Paudel, B., Tolmeijer, S., Laugwitz, L., & Bernstein, A. (2022). Benefits of diverse news recommendations for democracy: A user study. *Digital Journalism*, 10(10), 1710–1730. <https://doi.org/10.1080/21670811.2021.2021804>
- Hewett, K., Rand, W., Rust, R. T., & van Heerde, H. J. (2016). Brand buzz in the echoverse. *Journal of Marketing*, 80(3), 1–24. <https://doi.org/10.1509/jm.15.0033>
- Jiang, R., Chiappa, S., Lattimore, T., György, A., & Kohli, P. (2019). Degenerate feedback loops in recommender systems. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19)* (pp. 383–390). ACM. <https://doi.org/10.1145/3306618.3314288>
- Jungherr, A., Rivero, G., & Gayo-Avello, D. (2020). *Retooling politics: How digital media are shaping democracy*. Cambridge.
- Jungherr, A., & Schroeder, R. (2021). Disinformation and the structural transformations of the public arena: Addressing the actual challenges to democracy. *Social Media + Society*, 7(1), Article 205630512198892. <https://doi.org/10.1177/2056305121988928>
- Kaiser, J., & Rauchfleisch, A. (2019). Integrating concepts of counterpublics into generalised public sphere frameworks: Contemporary transformations in radical forms. *Javnost*, 26(3), 241–257. <https://doi.org/10.1080/13183222.2018.1558676>
- 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. <https://doi.org/10.1177/2056305120969914>
- Kalimeris, D. (2021). Preference amplification in recommender systems. In *27th ACM SIGKDD conference on knowledge discovery and data mining (KDD '21)* (pp. 805–815). ACM. <https://doi.org/10.1145/3447548.3467298>
- Karnowski, V., Kümpel, A. S., Leonhard, L., & Leiner, D. J. (2017). From incidental news exposure to news engagement. How perceptions of the news post and news usage patterns influence engagement with news articles encountered on Facebook. *Computers in Human Behavior*, 76, 42–50. <https://doi.org/10.1016/j.chb.2017.06.041>

- Kiddle, R. T., Welbers, K., Kroon, A., & Trilling, D. (2023). Enabling serendipitous news discovery experiences by designing for navigable surprise. In Vrijenhoek, S., Michiels, L., Kruse, J. Starke, A., Viader Guerro, J., & Tintarev, N. (Eds), *Proceedings of NORMalize 2023: The First Workshop on the Normative Design and Evaluation of Recommender Systems* (Short Paper 2). <https://ceur-ws.org/Vol-3639/short2.pdf>
- Krafft, T. D., Gamer, M., & Zweig, K. A. (2019). What did you see? A study to measure personalization in Google's search engine. *EPJ Data Science*, 8(1), Article 38. <https://doi.org/10.1140/epjds/s13688-019-0217-5>
- Loeberbach, F., & Trilling, D. (2020). 3bij3: Developing a framework for researching recommender systems and their effects. *Computational Communication Research*, 2, 53–79. <https://doi.org/10.5117/CCR2020.1.003.LOEC>
- Loeberbach, F., Welbers, K., Moeller, J., Trilling, D., & Van Atteveldt, W. (2021). Is this a click towards diversity? Explaining when and why news users make diverse choices. In *13th ACM Web Science Conference* (pp. 382–390). ACM. <https://doi.org/10.1145/3447535.3462506>
- Lu, F., Dumitrache, A., & Graus, D. (2020). Beyond optimizing for clicks: Incorporating editorial values in news recommendation. In: *UMAP '20: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization 9* (pp. 145–153) <https://doi.org/10.1145/3340631.3394864>
- Möller, J. (2021). Filter bubbles and digital echo chambers. In H. Tumber & S. Waisbord (Eds.), *Routledge companion to media, disinformation, and populism* (pp. 92–100). Routledge.
- Möller, J., Trilling, D., Helberger, N., & van Es, B. (2018). Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society*, 21(7), 959–977. <https://doi.org/10.1080/1369118X.2018.1444076>
- Mønsted, B., Sapieżyński, P., Ferrara, E., & Lehmann, S. (2017). Evidence of complex contagion of information in social media: An experiment using Twitter bots. *PLoS ONE*, 12(9), 1–12. <https://doi.org/10.1371/journal.pone.0184148>
- Ohme, J., Araujo, T., de Vreese, C. H., & Piotrowski, J. T. (2021). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function. *Mobile Media & Communication*, 9, 293–313. <https://doi.org/10.1177/2050157920959106>

- Ng, L. H. X., Cruickshank, I., & Carley, K. M. (2023). Coordinating narratives framework for cross-platform analysis in the 2021 US Capitol riots. *Computational & Mathematical Organization Theory*, 29, 470–486. <https://doi.org/10.1007/s10588-022-09371-2>
- Pagan, N., Baumann, J., Elokda, E., De Pasquale, G., Bolognani, S., & Hanák, A. (2023). A classification of feedback loops and their relation to biases in automated decision-making systems. In *EAAMO '23: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Article 7). <https://doi.org/10.1145/3617694.3623227>
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin.
- Perra, N., & Rocha, L. E. C. (2019). Modelling opinion dynamics in the age of algorithmic personalisation. *Scientific Reports*, 9, Article 7261. <https://doi.org/10.1038/s41598-019-43830-2>
- Puschmann, C. (2019). Beyond the bubble: Assessing the diversity of political search results. *Digital Journalism*, 7(6), 824–843. <https://doi.org/10.1080/21670811.2018.1539626>
- Rau, J. P., & Stier, S. (2019). Die Echokammer-Hypothese: Fragmentierung der Öffentlichkeit und politische Polarisierung durch digitale Medien? *Zeitschrift für vergleichende Politikwissenschaft*, 13, 99–417. <https://doi.org/10.1007/s12286-019-00429-1>
- Reveillac, M., & Morselli, D. (2022). Populism in an identity framework: A feedback model. *Communication Theory*, 32(1), 88–115. <https://doi.org/10.1093/ct/qtab003>
- Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira, W. (2019). Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. ACM. <https://doi.org/10.1145/3351095.3372879>
- Schoenbach, K. (2007). 'The own in the foreign': Reliable surprise – an important function of the media? *Media, Culture & Society*, 29(2), 344–353. <https://doi.org/10.1177/0163443707074269>
- Schönbach, K. (2017). Media effects: Dynamics and transactions. In Rössler, P., Hoffner, C.A., & Van Zoonen, L. (Eds.), *The international encyclopedia of media effects*. Wiley. <https://doi.org/10.1002/9781118783764.wbieme0026>
- Shehata, A., Andersson, D., Glogger, I., Hopmann, D. N., Andersen, K., Kruike-meier, S., & Johansson, J. (2021). Conceptualizing long-term media effects on societal beliefs. *Annals of the International Communication Association*, 45(1), 75–93. <https://doi.org/10.1080/23808985.2021.1921610>

- Shoemaker, P. J., & Reese, S. D. (2014). *Mediating the message in the 21st century* (3rd ed.). Routledge. <https://doi.org/10.4324/9780203930434>
- Shoemaker, P. J., & Vos, T. (2009). *Gatekeeping theory*. Taylor & Francis.
- Slater, M. (2007). Reinforcing spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. *Communication Theory*, 17, 281–303. <https://doi.org/10.1111/j.1468-2885.2007.00296.x>
- Slater, M. D. (2015). Reinforcing spirals model: Conceptualizing the relationship between media content exposure and the development and maintenance of attitudes. *Media Psychology*, 18(3), 370–395. <https://doi.org/10.1080/15213269.2014.897236>
- Song, H., & Boomgaarden, H. G. (2017). Dynamic spirals put to test: An agent-based model of reinforcing spirals between selective exposure, interpersonal networks, and attitude polarization. *Journal of Communication*, 67(2), 256–281. <https://doi.org/10.1111/jcom.12288>
- Spinelli, L., & Crovella, M. (2017). Closed-loop opinion formation. In *Proceedings of the 2017 ACM on Web Science Conference (WebSci '17)* (pp. 73–82). ACM. <https://doi.org/10.1145/3091478.3091483>
- Starbird, K., Arif, A., Wilson, T., Koevering, K. V., Yefimova, K., & Scarnecchia, D. P. (2018). Ecosystem or echo-system? Exploring content sharing across alternative media domains. In *12th International AAAI Conference on Web and Social Media (ICWSM)*. AAAI. <https://doi.org/10.1609/icwsm.v12i1.15009>
- Stroud, N. J. (2008). Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior*, 30(3), 341–366. <https://doi.org/10.1007/s11109-007-9050-9>
- Sunstein, C. R. (2001). *Republic.com*. Princeton University Press.
- Tambuscio, M., Oliveira, D. F. M., Ciampaglia, G. L., & Ruffo, G. (2016). Network segregation in a model of misinformation and fact checking. *Journal of Computational Social Science*, 1, 261–275. <https://doi.org/10.1007/s42001-018-0018-9>
- Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS ONE*, 13(9), Article e0203958. <https://doi.org/10.1371/journal.pone.0203958>
- Törnberg, P., Valeeva, D., Uitermark, J., & Bail, C. (2023). Simulating social media using Large Language Models to evaluate alternative news feed algorithms. arXiv. <https://doi.org/10.48550/arXiv.2310.05984>

- Trilling, D., Kulshrestha, J., De Vreese, C., Halagiera, D., Jakubowski, J., Möller, J., Puschmann, C., Stępińska, A., Stier, S., & Vaccari, C. (2022). Is sharing just a function of viewing? The sharing of political and non-political news on Facebook. *Journal of Quantitative Description: Digital Media*, 2, Article 16. <https://doi.org/10.51685/jqd.2022.016>
- Tsfati, Y., Boomgaarden, H. G., Strömbäck, J., Vliegenthart, R., Damstra, A., & Lindgren, E. (2020). Causes and consequences of mainstream media dissemination of fake news: Literature review and synthesis. *Annals of the International Communication Association*, 44(2), 157–173. <https://doi.org/10.1080/23808985.2020.1759443>
- Vaccari, C., & Valeriani, A. (2021). *Outside the bubble: Social media and political participation in Western democracies*. Oxford University Press.
- Van Aelst, P., Van Santen, R., Melenhorst, L., & Helfer, L. (2016). From newspaper to parliament and back? A study of media attention as source for and result of the Dutch question hour. *World Political Science*, 12(2), 261–282. <https://doi.org/10.1515/wps-2016-0011>
- van Hoof, M., Trilling, D., Meppelink, C., Moeller, J., & Loecherbach, F. (2023). Googling politics? The computational identification of political and news-related searches from web browser histories. OSF. <https://doi.org/10.31235/osf.io/5h9qr>
- van Hoof, M., Trilling, D., Moeller, J., & Meppelink, C. (2024). It matters how you google it? Using agent-based testing to assess the impact of user choices and algorithmic curation on political Google search results. OSF. <https://osf.io/preprints/socarxiv/hf6sc>
- Vasterman, P. L. (2005). Media-hype: Self-reinforcing news waves, journalistic standards and the construction of social problems. *European Journal of Communication*, 20(4), 508–530. <https://doi.org/10.1177/0267323105058254>
- Vrijenhoek, S., Bénédic, G., Gutierrez Granada, M., Odijk, D., & De Rijke, M. (2022). RADio – Rank-aware divergence metrics to measure normative diversity in news recommendations. In *Sixteenth ACM Conference on Recommender Systems*, 208–219. <https://doi.org/10.1145/3523227.3546780>
- Vrijenhoek, S., Kaya, M., Metoui, N., Möller, J., Odijk, D., & Helberger, N. (2021). Recommenders with a mission: Assessing diversity in news recommendations. In *CHIIR '21: Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (pp.173–183). ACM. <https://doi.org/10.1145/3406522.3446019>

- Waldherr, A., Geise, S., & Katzenbach, C. (2019). Because technology matters: Theorizing interdependencies in computational communication science with actor-network theory. *International Journal of Communication*, *13*, 3955–3975.
- Waldherr, A., & Wettstein, M. (2019). Bridging the gaps: Using agent-based modeling to reconcile data and theory in computational communication science. *International Journal of Communication*, *13*, 3976–3999.
- Westley, B. H., & MacLean, M. S. (1957). A conceptual model for communications research. *Journalism Quarterly*, *34*(1), 31–38.
<https://doi.org/10.1177/107769905703400103>
- Whittaker, J., Looney, S., Reed, A., & Votta, F. (2021). Recommender systems and the amplification of extremist content. *Internet Policy Review*, *10*(2).
<https://doi.org/10.14763/2021.2.1565>
- Whitten-Woodring, J., Kleinberg, M. S., Thawngmung, A., & Thitsar, M. T. (2020). Poison if you don't know how to use it: Facebook, democracy, and human rights in Myanmar. *The International Journal of Press/Politics*, *25*(3), 407–425. <https://doi.org/10.1177/1940161220919666>
- Zuiderveen Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., de Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, *5*(1). <https://doi.org/10.14763/2016.1.401>

Acknowledgements

This work is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 947695). I would like to thank all team members and collaborators of recent years for the many insightful discussions that made this article possible.

Date received: July 2023

Date accepted: April 2024