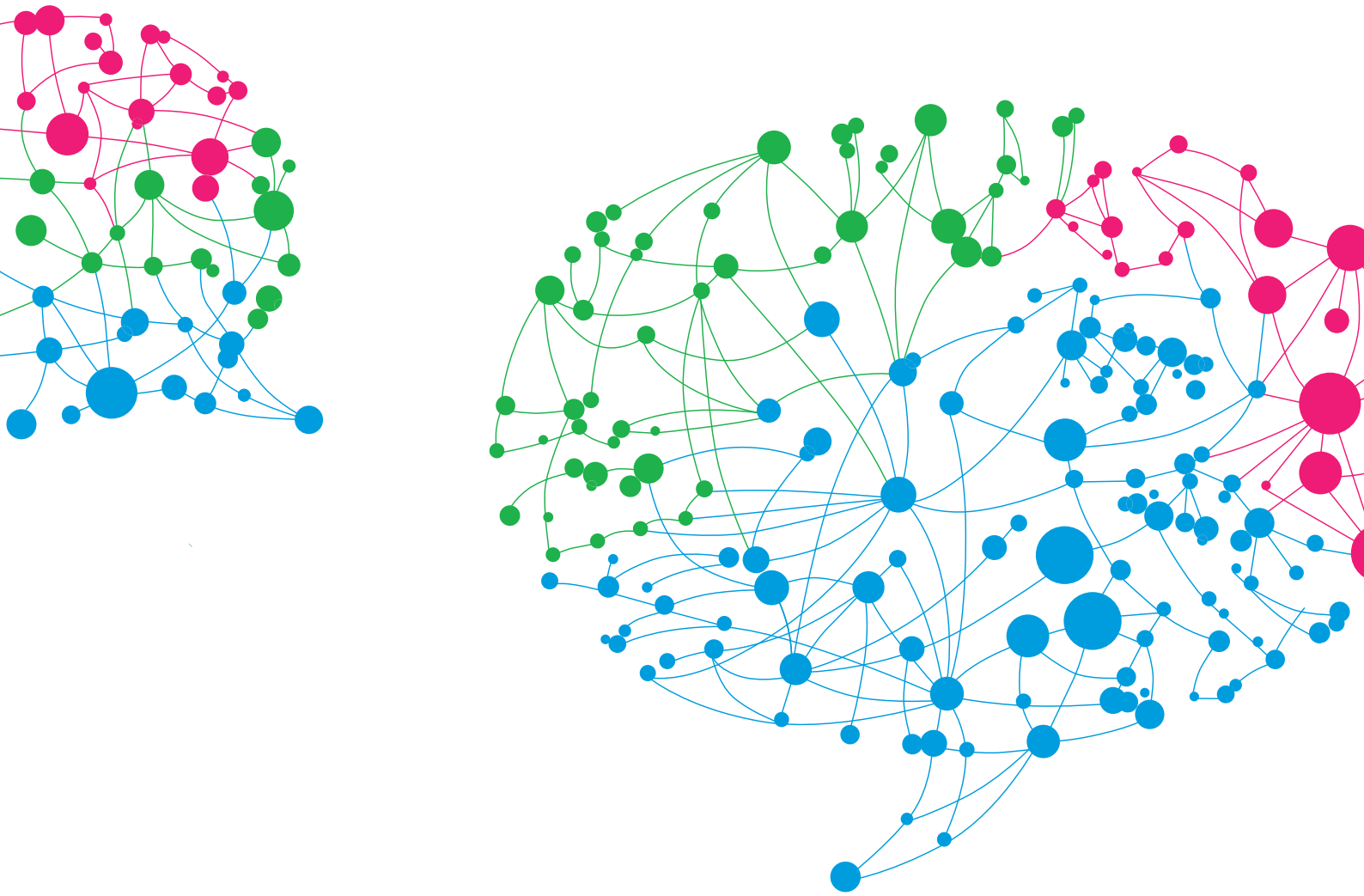
GOVERNING PLATFORMS

Are Algorithms a Threat to Democracy?
The Rise of Intermediaries: A Challenge for Public Discourse

Professor Dr. Birgit Stark and Daniel Stegmann, M.A.
with Melanie Magin, Assoc. Prof. & Dr. Pascal Jürgens

26 May 2020



Published as part of Governing Platforms, a research project by

in partnership with

with support by

Contents

Executive Summary

A healthy democracy needs informed citizens. People are expected to be aware of important issues and public affairs in order to provide feedback on the political system. Therefore, a diversity of viewpoints is considered a core democratic value and one of the core public values in media law and policy. With the increasing importance of intermediaries, the question arises whether algorithmic news curation is a threat to democracy.

Fears that algorithmic personalization leads to filter bubbles and echo chambers are likely to be overstated, but the risk of societal fragmentation and polarization remains

- The fear that large parts of the population are trapped in filter bubbles or echo chambers seems overstated. Empirical studies offer a much more nuanced view of how social media affects political polarization. Due to the fact that our information repertoires are still broadly dispersed, we adapt our worldview, allow opposing opinions, and are exposed to relevant societal issues. Several studies show that incidental exposure and network effects can even contribute to an expansion of the diversity of information.

- Nevertheless, there is evidence for possible polarization at the ends of the political spectrum. The current state of research permits the assumption that echo chambers may arise under certain circumstances; that is, facilitated by homogeneous networks, highly emotionalized and controversial topics, and strong political predispositions. In particular, social media logics can reinforce affective polarization because the features of social media platforms can lead to very stereotypical and negative evaluations of out-groups.

- Moreover, social media may indirectly contribute to polarization by facilitating a distorted picture of the climate of opinion. As a result, spiraling processes begin because the perception of the strength of one's own opinion camp compared to those of other camps is overstated. The entire process leads to an overrepresentation of radical viewpoints and arguments in the political discourse. At this point during the opinion formation process, individuals are more vulnerable to being influenced by "fake news" on Facebook or Twitter. Thus, strategic disinformation cannot only influence the media's agenda through specific agenda-setting effects but can also impact the public's climate of opinion.

Social media are vulnerable to facilitating a rapid dissemination of disinformation, but exposure seems to be limited

- There are orchestrated disinformation campaigns online, but data on the actual scope of and exposure to disinformation is scarce.

- The few available scientific studies suggest that the extent of the problem is likely to be overestimated since exposure to disinformation seems to be rather limited.

- Studies on the effects of disinformation of users show no persuasive effects but a confirmation bias: disinformation may therefore widen existing gaps between users with opposing worldviews because it is able to confirm and strengthen pre-existing attitudes and (mostly right-wing) worldviews. In this context political microtargeting poses a concern, as it can be used to disseminate content tailored to target groups particularly susceptible to disinformation.

- More research on the scope of, interaction with and individual and societal effects of disinformation is crucial to better assess the actual extent of the problem regarding disinformation.

Social media contribute to the dissemination of incivility and hate speech

- Incivility appears to be significantly widespread online and has real, negative effects on individual attitudes and the discourse climate.

- A potentially serious problem is the indirect effects of incivility on recipients and journalists: the credibility of journalistic content is reduced by incivility, including hate speech, in comment sections, which can have detrimental effects on trust in journalism as an institution of social cohesion in the long term.

- In addition, empirical evidence indicates that journalists react to incivility directed at them by shying away from controversial reporting or trying to hide controversies as a reaction to incivility. This is worrying because it hinders the free development of democratic discourse.

- A related problem is that especially women, who have been victims of hate speech, stop participating in discussions. This again backfires on the free development of public discourse on the macro-level, if whole groups of the population are cast out.

- Discourse moderation in comment sections that focuses on sociable replies on comments by journalists seems to be an effective tool in containing and preventing incivility including hate speech.

- Measures inhibiting freedom of expression have to be carefully applied and can only be used to combat illegal content such as hate speech.

Research agenda for platform governance

It can be stated that fears of filter bubbles and echo chambers seem overstated. Echo chambers and polarization seem to emerge only at the fringes of the political spectrum. Worrisome, however, are indications that social media may indirectly contribute to polarization by facilitating a distorted picture of the climate of opinion.

There are reasons for vigilance in the cases of disinformation and incivility. A serious assessment of the extent of the problem of disinformation is hampered by the scarcity of the available scientific data. The available studies suggest that an excessively alarmist political and societal debate is to be avoided, but the actual scope of the issues remains unclear. Incivility and hate speech are prevalent phenomena that should be tackled with evidence-based policy measures. That means that (further) regulation, bans or deletion of content, which entails legal problems, are not necessarily the best solution. From the

perspective of communication science, the main goal of any intervention should be to strengthen a reasonable, fruitful and free democratic discourse.

In this context, we emphasize that existing successful approaches (e.g., community management in the form of moderation that does not entail deleting content to contain and prevent incivility) should be extended and consistently applied. In addition, further scientific evidence is needed, in particular on disinformation, in order to investigate the extent of the phenomenon and its consequences for public discourse and society in more detail, so that evidence-based measures can be developed. From a communication science perspective, it is precisely at this point that regulation appears most useful, especially with respect to demystifying the inner workings of "black box" algorithms and providing relevant data for research purposes. Hence, without access to Google's or Facebook's internal data, it is hard to reach firm conclusions. Therefore, it is critical to continue monitoring the evolution of digital news markets and the ways in which users are exposed to news on social media platforms. In particular, structural changes in the news market require the attention of regulators and policy makers. Intermediaries establish economic power and create new dependencies.

1. Introduction

The media essentially contribute to the functioning of democracy since they set the agenda, provide background information and represent different viewpoints with regard to political issues. Based on this, citizens are supposed to form their own opinions and participate in democratic decision-making processes (Katz, 1996; Tewksbury & Rittenberg, 2008). Hence, a healthy democracy needs informed citizens. People are expected to be aware of important issues and public affairs to provide feedback to the political system.

1.1. Algorithmic Gatekeeping

In high-choice media environments, this function is no longer restricted to traditional mass media but also intermediaries, such as Facebook or Google (Jürgens & Stark, 2017; Webster, 2010). For several years, social media has been rapidly emerging as an important source of news. Their algorithm-based selection logics are subject to their own rules, which include both chances and risks for viewpoint diversity because they do not filter and select content according to editorial news values, but on economic guidelines and popularity (Nechushtai & Lewis, 2019). Such private Internet companies, which are by nature oriented toward profit, are therefore not necessarily obliged to foster the diversity of their content. Rather, their filtering and sorting is best suited to the users in order to best serve the interests of their advertising customers. The guiding principle of social media logic is the generation of attention. This is measured by the frequency of interaction between users and content in the form of clicks, reactions, sharing, and comments. The fact that these user reactions in turn influence the visibility of the respective contribution in other users' news feeds leads to a self-reinforcing cycle: the more interaction, the higher the visibility in the future (Napoli, 2019). This "platformization" changes the opportunities of access to the public and changes the logics of public communication and pressurizes professional journalism because of the outflow of advertising money to the platforms. News items are accessed less often than a bundled overall offer of individual media brands; therefore, every single post fights for attention in the news feed. Since user reactions are usually based on the very first impression, clickbaiting in the news feed is important for attracting attention (van Dijck & Poell, 2013; Welbers & Opgenhaffen, 2019).

1.2. Responsibilities of Intermediaries

In the meantime, a broad public debate has developed about the democratic implications of algorithm-based news or recommendation systems such as social media, search engines or news aggregators in science and media politics. Only a few years ago, intermediaries were ascribed an indirect influence on the processes of individual and public opinion formation, since they do not usually create or change content themselves, but rather control the degree or probability of finding content elsewhere. However,

this position has changed dramatically in recent years. Not only is the mediator role much more critically assessed due to possible risks of manipulation and abuse, but the debate about the services, functions and effects of algorithmic selection on the Internet has also intensified. Platforms are increasingly criticized for the lack of transparency of their criteria for the aggregation, selection, and presentation of content.

Increasingly, ethical questions are debated that address how much responsibility platforms can and should bear, to what extent the functions of algorithms influence public discourse sustainably and how algorithms change journalism itself. Meanwhile, algorithmic curation is progressing rapidly in other areas of society as well, since algorithms have long since ceased to be used only in the field of communication. Currently, the automated processing of data shapes a large part of our behaviors in digital, networked environments—the job offers we notice, what products we order, or how we navigate unknown places. Equipped with artificial intelligence, they record and interpret what is happening around them—this is how they learn to react to us and make automated decisions.

The range of concerns that have arisen around intermediaries has sparked not only a wide range of suggestions for courses of action, but also concrete legal regulations—at least in some countries. A more pro-active regulatory approach is not only found in Germany (NetzDG, the current state media treaty) or France, but now also at the European level (e.g., the adopted Audiovisual Media Directive, or the fines for Google, or the recommendations from the High-Level Expert Group on Fake News and Online Disinformation). The German NetzDG (Network Enforcement Act), implemented to counter the dissemination of hate speech, obliges platform operators to block and remove illegal content (Löber & Roßnagel, 2019). The latest reports show that users of Facebook, Twitter and YouTube reported almost 700,000 posts (Facebook, 2019; YouTube, n.d.; Twitter, 2019) that were removed by the platforms in the period from January to June 2019, which suggests that platform operators seem increasingly aware of their responsibilities. There are, however, legal concerns regarding unwarranted censorship, especially since this fundamental task of law enforcement is transferred to private companies (e.g., Eickelmann, Grashöfer & Westermann, 2017; cf., Cornils, 2020).

1.3. Structure of the Report

In our report, we will focus on major concerns that are all closely linked to the effects of changing individual and public opinion formation processes. From a democratic perspective, a fundamental question is how algorithmic curation influences the character and quality of our democracy. Therefore, we focus on 1) filter bubbles and echo chambers, 2) increasing fragmentation and polarization of the audience, 3) declining quality of news, and 4) the radicalization of public discourse (through disinformation/fake news and incivility/hate speech). The review of the literature is provided in separate chapters, each of which can be read individually but that cumulatively are intended to provide an overview of what is known—and unknown—about the relationship between social media, political polarization and fragmentation, incivility and disinformation.

The report is structured as follows. We start with a characterization of the differences between human and algorithmic curation of news. In this background chapter, readers will find important basic knowledge to better understand how algorithms filter, sort, and personalize news content. More specifically, we consider whether the fundamental characteristics of social media undermine the core assumption of diverse news. The third chapter is dedicated to the popular metaphors of filter bubbles and echo chambers. We try to define both phenomena clearly. To this end, we describe the often neglected

mechanisms that typically lead to their emergence and their potential danger to democracy. Moreover, we take a look at the importance of intermediaries, such as Facebook or Google, for news consumption: Do people inform themselves exclusively through algorithmically curated sources?

Based on a research overview, we then question whether filter bubbles and echo chambers actually exist according to the current state of knowledge, which conditions can favor their formation, and which research methods are suitable for their investigation. To illustrate some of the challenges involved in research on intermediaries' effects, we present a concise, realistic example of how such a study might be designed. Building on the issues raised in chapter three and four, we will discuss further risks for traditional media, in particular, whether the adaptation to social media logics leads to a decline of media quality or a softening of news. Subsequently, in the context of a narrative literature review, we discuss both possible causes of the perception of an uncivil language and debate effects on the users of social media. At first, we consider and define disinformation and fake news and outline how both phenomena are a veritable threat to a free public discourse from a theoretical point of view. We conclude this section by questioning whether disinformation negatively impacts public discourse as heavily as the public debate suggests. We then focus on the increasing incivility in online discourses, point toward negative effects on the users of social media, which subsequently affect public discourse in general, and discuss various countermeasures.

In light of the dangers reflected in the report, chapter six explores how and why intermediaries change the conditions for opinion formation and how their role differs from that of traditional mass media. Our concluding remarks identify key aspects in our understanding of these phenomena and the data that are needed to address them. We consider how to assess threats in terms of traditional policy ideas of media diversity.

Furthermore, we combine the findings from the systematic analysis of the state of research and our own empirical analyses (e.g., Geiß, Magin, Stark, & Jürgens, 2018; Jürgens & Stark, 2017; Stark, Magin, & Jürgens, 2017; Stark, Magin, & Jürgens, in press). The central media policy objectives are not part of our paper. For a detailed analytical framework that captures the diversity of both challenges and legal approaches in different European countries, see Cornils (2020).

2. Algorithmic Information Filtering

This section examines differences between human and algorithmic gatekeeping of news. It focuses on literature describing the effects of algorithmic gatekeeping or curation and their implications for news consumption. Readers will find important basic knowledge to better understand how algorithms filter, sort, and personalize news content. In offering personalized content intermediaries or news recommender systems are assumed to reduce news diversity. The underlying assumption is that personalized news sources prioritize like-minded content and opinions that conform to existing preferences.

2.1. Algorithmic Gatekeeping

The current debate is closely related to the rise of the so-called intermediaries—"brokers of information that position themselves between producers and consumers while altering the flow of information" (Jürgens & Stark 2017, p. 398). The terms intermediaries or information intermediaries have so far been used to describe various bundled platforms or services. Following Webster (2010), this definition includes search engines (e.g., Google), social media (e.g., Facebook, Twitter) and news aggregators (e.g., Reddit, Google News) under this heading. The current German state media treaty defines the term even more broadly and frames the concept under the term "media intermediaries": Media intermediaries are all telemedia that aggregate, select and present journalistic-editorial offers of third parties in a generally accessible manner without curating them into an overall offer (§ 2 Abs. 2 Nr. 13 b) (Medienstaatsvertrag, 2019, p. 11).[1] These include search engines, social media platforms, user-generated content portals or news aggregators. Central to this report are social networks and search engines, which are the most used platforms and the most prominent providers, with a particular focus on Facebook.

In many countries, social networks and search engines are among the main news sources for wide parts of the population (see chapter 4.1) and increasingly control the dynamics of the production, dissemination, and consumption of news (see chapter 5.1). There is still relatively little known about the way that the algorithmic systems of intermediaries function—they are often referred to as "black boxes"—which makes it difficult to accurately assess their societal influence (Napoli, 2019). It is evident though that intermediaries have a decisive influence on which political information reaches users by means of three general functions: (1) they algorithmically select content that is relevant for the user, while at the same time excluding other content classified as non-relevant (filtering). In this context, "relevant" always means relevant for the individual user, not for society as a whole. (2) They rank the selected content in such a way that the most relevant content for users is (supposed to be) at the top of the results list, such as a search results page, social media feed, or news recommendation app (sorting). (3) They customize

1 For a detailed description see Cornils (2020, pp. 15-16).

content to the interests and preferences of each user (personalization) (Jürgens & Stark, 2017).

Intermediaries act as "gatekeepers" by means of these functions. Gatekeeping "refers to the process by which news organizations make decisions as to which of the events taking place or issues being discussed receive coverage" (Napoli, 2019, p. 54), i.e., this task was traditionally performed by professional mass media (newspapers, television, radio). In the digital media environment, traditional gatekeeping is increasingly replaced by what Thorson and Wells (2016) refer to as algorithmic curating: in this sense, curating is understood as selecting, filtering and organizing, so the abundance of information available online can be received by the individual user. In their framework, key actors are thus not only professional journalists, strategic (political) communicators, individual media users (personal curators) and the social environment (social contacts), but also algorithms (Thorson & Wells, 2016, p. 314). Consequently, as mentioned above, intermediaries are increasingly taking over the gatekeeping tasks of professional journalism: they assess and weigh content, (co-)determine which content is visible and therefore noticeable to users, and influence the diversity of content that is consumed (Jürgens & Stark, 2017; Webster, 2010). The transfer of the gatekeeping process from news media to the algorithmic curating of intermediaries would be unproblematic if both actors applied the same criteria to decide which stories or issues were newsworthy and therefore reported or made available to the individual user.

Traditionally, journalistic gatekeeping has been a process characterized by complex institutional, individual, and organizational factors (Shoemaker & Reese, 2014). However, news values are at the core of these various levels of analysis as the criteria that determine newsworthiness and thereby—along with other important criteria such as professional assessments, the market, and the audience (Schudson, 2011)—guide journalistic decision-making (Napoli, 2019, p. 56). Commonly identified news values such as controversy, conflict, negativity, proximity or elite people (Galtung & Ruge, 1965) can, according to Shoemaker and Cohen (2005), be narrowed down to two formative dimensions: "deviance" and "social significance". The last dimension refers to public tasks undertaken by journalism in democratic societies (e.g., Burkart, 1995; Katz, 1996); from a democratic theory perspective, it is particularly relevant here that media inform citizens about issues, actors, and opinions relevant to society as a whole, so that citizens are able to make well-considered decisions (e.g., Coleman, 2012, pp. 35–36; van Aelst et al., 2017, p. 5). Thus, journalistic gatekeeping is, for the most part, oriented toward fulfilling this mandate for the public good.

In contrast, algorithmic curating functions are fundamentally different: first, the rise of intermediaries increasingly uncouples the traditionally vertically integrated production and distribution of news, in which news organizations were in charge of both the production and dissemination of news content to their audience. With the establishment of news aggregators, search engines and social media as intermediary institutions between news producers and recipients, concerns about a possible distortion of the free, principally unrestricted flow of news information have been raised since opportunities have arisen to suppress or filter certain news (Napoli, 2019, pp. 59–61).

Possible distortive effects must be assessed in the context of the economically oriented selection mechanisms of intermediaries. On the one hand, intermediaries are designed to serve the individual needs of their customers—the users—in the best possible way. On the other hand, due to their orientation toward click numbers, they favor popular offers that are preferred by the masses ("popularity bias"). These effects are self-reinforcing when users concentrate on frequently read or recommended articles and thus further emphasize them (Stark & Magin, 2019). Especially with regard to social media networks, the main imperative behind this is to "increase the number of

regular users and the time that they spend on the platforms—*to increase engagement*" (Napoli, 2019, p. 36) because for platforms, it is primarily important to keep as many users for as long as possible to—corresponding to their business model—sell advertising.

According to DeVito (2017) there are two major differences between algorithmic curating and the selection of news by journalists, which become particularly tangible when using Facebook as an example. Based on a content analysis, the author examined the few publications by Facebook that provided information about the operating principles of the algorithm (e.g., press releases, blogs or patents). Therein, nine relevant "editorial values" were identified: "Friend relationships, explicitly expressed user interests, prior user engagement, implicitly expressed user preferences, post age, platform priorities, page relationships, negatively expressed preferences, and content quality" (DeVito, 2017, p. 14).

In this context, of particular interest is their relative importance: the ranking criteria of Facebook show only slight overlaps with classical news factors: only the prioritization of new stories and local stories are related to the news factors "novelty" and "proximity" (Napoli, 2019, p. 63). However, the three most important characteristics mentioned are the quality of the relationships (affinity) as well as explicitly and implicitly stated user interests, the latter being identified by the user's previous behavior. The analysis of Facebook's patents in particular showed that, "friend relationships are a guiding value that mediates the application of all the other values" (DeVito, 2017, p. 14). This has been emphasized in an update of the news feed in 2018 when Facebook explicitly strengthened the prominence of posts assumed to stimulate discussions and other "meaningful interactions" such as shares and likes in user's networks (Mosseri, 2018). Ultimately, the information available on the selection principles of the Facebook algorithm is only an approximation of the actual selection, as the more than 200 selection criteria are well-guarded company secrets and constantly updated and adapted (De Vito, 2017).

The two main differences between journalistic and algorithmic curation can therefore be identified as follows: on the one hand, relevant editorial news values (such as controversy, negativity, and elite people) interact with each other. A single factor is never decisive—unlike on Facebook, where only popularity with the users and their personal network determines the content of the news feed. On the other hand, the basic direction is fundamentally different: while the news values that have traditionally guided journalistic gatekeeping emphasize social significance, the news values of intermediaries like Facebook focalize personal significance and are thereby primarily audience-oriented.

In conclusion, it can be stated that news selection by algorithms fundamentally differs from traditional news selection in journalism. The latter is oriented toward the audience and its (presumed) interests, but mainly toward the public tasks of journalism and its societal relevance. Online, new "editorial news values" are added that cannot be identified by the characteristics of the reported events, but are created and prioritized by algorithms and, above all, are intended to generate reach and interactivity. Thus, according to Loosen and Scholl (2017, p. 360), the orientation toward the audience leads to an algorithmic operationalization of news values from the user's point of view. Algorithmic curation is primarily oriented toward user preferences that reflect previous patterns of use and personal interests.

The extent to which such news curation can have important implications will be characterized in the next chapters. Since the debate about the negative consequences of algorithmic curating focuses on the impact of algorithmic personalization, we will first concentrate on personalization filters. The underlying assumption is that personalized news sources prioritize like-minded content and opinions that conform to existing preferences.

2.2. News Personalization

The personalization of content to the interests and preferences of each user is one major function by which intermediaries have a decisive influence on which political information reaches users (Jürgens & Stark, 2017). Mechanisms of personalization of news and information providers are usually based on explicit and implicit statements of users; for example, they differ in terms of the individual level of control users have on the processes of personalization (Bozdag, 2013). Thus, in general, two forms of personalization that are often combined with each other can be distinguished:

- **Explicit or user-driven personalization (customization)** describes the active disclosure of user-preferences in order to inform a provider about the content an individual user would like to receive in the future. Explicit personalization is therefore based on information that relates to personal interests (e.g., "liking" a Facebook-page, or "following" a Twitter account) and socio-demographic data consensually shared by users: for example, if users register on a platform and actively configure their profiles. This does not necessarily require an algorithm; classic newspaper subscriptions and customer files are also (simple) forms of personalization. However, algorithms enable much more precise personalization of content in real time.

- **Implicit or algorithmic personalization,** on the other hand, is unthinkable without algorithms. Based on users' previous interactions, systems predict what further content might also be of interest to users. This kind of personalization requires a technical system (e.g., a social network) to collect comprehensive data about its users; this includes behavioral data (e.g., location, links clicked on, resting time on certain pages, or comments written) and meta-information for previously viewed content (e.g., creation date, document type, or popularity cues such as the number of "likes" and "shares") (Bozdag, 2013; Schweiger, Weber, Prochazka, & Brückner, 2019, p. 8–9). In addition, some algorithms (e.g., those of Facebook, Spotify, or Netflix) also incorporate data from the users' personal online-networks based on similarity analyses (Bozdag, 2013). Such filters take for granted that users like what their friends like—based on the homophily of networks, i.e., the human tendency to enter into social relationships with people, with whom they have much in common (McPherson, Smith-Lovin, & Cook, 2001). The collected data enable the algorithm to recognize user preferences and thus select and display relevant information based on users' individual preferences—a dynamic communication situation that relies on reciprocal interactions (DeVito, 2017; Zuiderveen Borgesius et al., 2016).

As mentioned in the previous chapter, intermediaries rely on personalization technologies for economic reasons: advertising revenues are their main source of funding—and these increase the more precisely they can tailor advertising to their users ("micro-targeting", see also chapter 5.2). They also analyze user data to personalize their other content in order to make their offer as attractive as possible for the users and to tie them to it—since they are then able to collect even more data and accordingly increase their advertising revenues. The users prefer personalized offers because they are practical and convenient from their point of view and provide them with relevant content (Stark, Magin, & Jürgens, 2014).

These indisputable benefits, however, have their downsides: no user is aware of the content that is not displayed, and the systematic hiding of certain content can lead to a biased worldview, as described by the metaphors of the filter bubble and echo chamber. This not only applies to political content, but ideologies and extreme viewpoints of all kinds can become the nucleus of one-sided information

environments. In the political sphere, however, they can have particularly far-reaching consequences for democracies, which is why we will concentrate on them in the following.

Main Findings:

- Algorithms largely govern the selection, sorting, and presentation of information on the Internet.
- The logics that underpin algorithmic gatekeeping differ from the logics of human gatekeeping.
- Algorithms customize content to the interests and preferences of each user (personalization).
- Due to their business model—advertising—tech companies like Facebook or Google try to maximize the amount of time people spent on their platforms.
- Algorithms, like Facebook's news feed values, emphasize personal significance to increase audience engagement with particular types of content, whereas journalistic gatekeeping emphasizes social significance—oriented toward the public tasks of journalism.
- A deeper understanding of the contemporary gatekeeping process requires a detailed examination of the generally opaque algorithmic systems.

3. Concerns Related to Algorithmic Gatekeeping and Personalization

This section is dedicated to the popular metaphors of filter bubbles and echo chambers. We define both phenomena clearly. To this end, we describe the often neglected mechanisms that typically lead to their emergence and their potential danger to democracy: Fragmentation and polarization. Serious concerns have been raised about the common public agenda in society and the media's role in contributing to a social consensus. The move to high-choice media environments and personalized media repertoires have sparked fears over audience fragmentation. Moreover, polarization resulting from exposure to personalized news media is a longstanding concern. We define the risks and describe their implications for society as a whole.

3.1. Filter Bubbles and Echo Chambers

Intermediaries like social networks, search engines, or news aggregators are blamed for confronting their users only with content tailored to their personal preferences. In this context, the filter bubble is an omnipresent buzzword. Hardly any metaphor has become as popular as the filter bubble—that is, "a unique universe of information for each of us" (Pariser, 2011, p. 9)—a term first coined by Internet activist Eli Pariser. In his book, *The Filter Bubble: What the Internet Is Hiding from You*, Pariser refers to the personalization logics of Facebook and warns of the societal impacts of algorithmically limited perspectives. It is important to note that Pariser largely fails to provide a clear definition for the 'filter bubble' concept; it remains vague.

The metaphor of the echo chamber was first mentioned by legal expert Cass Sunstein in 2001. It describes the way that group dynamic processes of opinion formation proceed in personalized information environments, which consistently reflect an individual's opinion on him or herself, like an echo (Sunstein, 2001). Thus, the term "echo chambers" refers to the effects of the largely exclusive exposure to consonant opinions in an online environment with *like-minded users* (Stark et al., in press). Typically, echo chambers develop as follows: Since humans try to avoid cognitive dissonance (Festinger, 1957), they form networks with others who are similar to them on and offline ("homophily"; McPherson et al., 2001). These like-minded individuals constantly reassure themselves of their respective opinions, whereas contact with diverging opinions is increasingly marginalized. The results are parallel discourses uncoupled from the debate of the whole society and/or very small sub-audiences (Magin, Geiß, Jürgens, & Stark, 2019).

While both metaphors describe the dangers of personalized media environments, which transport individuals into a world in which new information echoes what we already know, have heard, or think, they are not equivalent. The decisive factors distinguishing filter bubbles from echo chambers are the focus on the individual user and the reference points of the respective metaphor: filter bubbles refer to the

distribution and usage of *information* and develop around *a single user* through algorithmic recommendations, in which the individual user may be largely uncoupled from relevant societal discussions. Echo chambers refer to communication situations where one is exposed only to opinions that agree with their own, thus one is never alone in an echo chamber.

A common concern linked to these effects of intermediaries relates to how diversity could be diminished by filter bubbles and echo chambers. According to Helberger (2018, p. 158), "their main goal is to channel audience attention and affect access to and the diverse choices people make. As such, they affect not so much the diversity of supply (social media platforms do not produce content), but rather the diversity of media content individual members of the audience are eventually exposed to (exposure diversity)." In particular, the lack of transparency fuels concerns of biases of algorithmic curation that might impair exposure diversity: media users are faced with a lack of control and transparency when interacting with these systems because algorithms largely remain black boxes to end users.

3.2. Fragmentation and Polarization of the Public Sphere

It is feared that this limitation of diversity regarding topics and issues may lead to increased fragmentation and polarization at the societal macro level (Bennet & Iyengar, 2008; Sunstein, 2018; Webster & Ksiazek, 2012). "The core argument is that societies are facing increasing political divides with respect to both media content and public beliefs" (van Aelst et al., 2017, p. 12). More specifically, the term "fragmentation" is used to describe the disintegration of society into smaller sub-units that no longer form a cohesive whole, induced by individualized media exposure (Lee, 2009; Webster, 2005). These sub-audiences barely interact in a "common meeting ground" (Katz, 1996); thus, the social consensus on which societal problems are the most urgent and require a solution is declining. This endangers the integration and stability of democratic societies (Tewksbury, 2005); "algorithmic news personalization could lead to a situation, in which the shared public sphere becomes increasingly disintegrated and breaks up into smaller issue publics" (Moeller & Helberger, 2018, p. 4).

Fragmentation of the public agenda prevents people from sharing a common experience and from understanding one another. In view of this, social media may dangerously reduce the common core of topics for public discourse in a democracy. Traditionally, fragmentation research focuses on a shared issue or agenda rather than on like-minded opinions, which is why the theory is more closely linked to the concept of filter bubbles than to the concept of echo chambers (Fletcher & Nielsen, 2017).

"Polarization" refers to the ideological division of a society into different (extreme) political camps. Polarization can be categorized into

- the division of citizens' attitudes ("attitude polarization"; Fiorina & Abrams, 2008; Prior, 2013) regarding party preferences ("party polarization") or certain issues ("issue polarization"), but also
- the divergence of their emotional attitudes to specific (social) groups ("affective polarization"; Iyengar, Sood, & Lelkes, 2012) and the citizens' perception of how polarized the society as a whole appears ("perceived polarization"; Lupu, 2015; Yang et al., 2016).

From a normative point of view, polarization is regarded as problematic because it makes compromises, which are existential for democracies, more difficult and in extreme cases impossible. Since polarization is focused on attitudes and the increasing extremity of opinion camps, it is associated with echo chambers and social media rather than filter bubbles and search engines.

The same also applies to the radicalization of public discourse, which is both the cause and effect of polarization. On social media, users receive a biased picture of the opinion climate, since neither the user community in general nor those users involved in certain discussions are representative of society as a whole. Users with extreme attitudes are over-represented in debates on social media, and algorithms prefer more radical statements that evoke reactions from other users. Thus, ambassadors of extreme viewpoints get the impression that such viewpoints represent the majority opinion and articulate them even more loudly—outside of social media as well. In contrast, the users representing moderate viewpoints (and thus the actual majority opinion) perceive themselves as belonging to a minority and fall silent—a classical "spiral of silence" process (Magin et al., 2019).

4. Empirical Findings Related to the Effect of Algorithmic Information Filtering

The extent to which these risks and dangers become reality depends largely on the relevance of intermediaries in the users' news repertoires. To assess the extent to which algorithmic curation is currently determining news consumption, we take a look at the Reuters Institute Digital News Report, which reveals insights about digital news consumption in different European countries (first subsection). The second part of this chapter shifts the focus from the literature on news consumption to the effects of algorithmic personalization. This subsection describes the recent literature on filter bubbles and echo chambers. Central empirical studies are differentiated along the focus of their research interest: the impact of algorithmic personalization on searching the Internet, its influence on social media, and the effects on opinion-formation processes. Finally, we illustrate how such a study might be designed to examine echo chambers.

4.1. Pathways to News: the Rising Prominence of Digital News

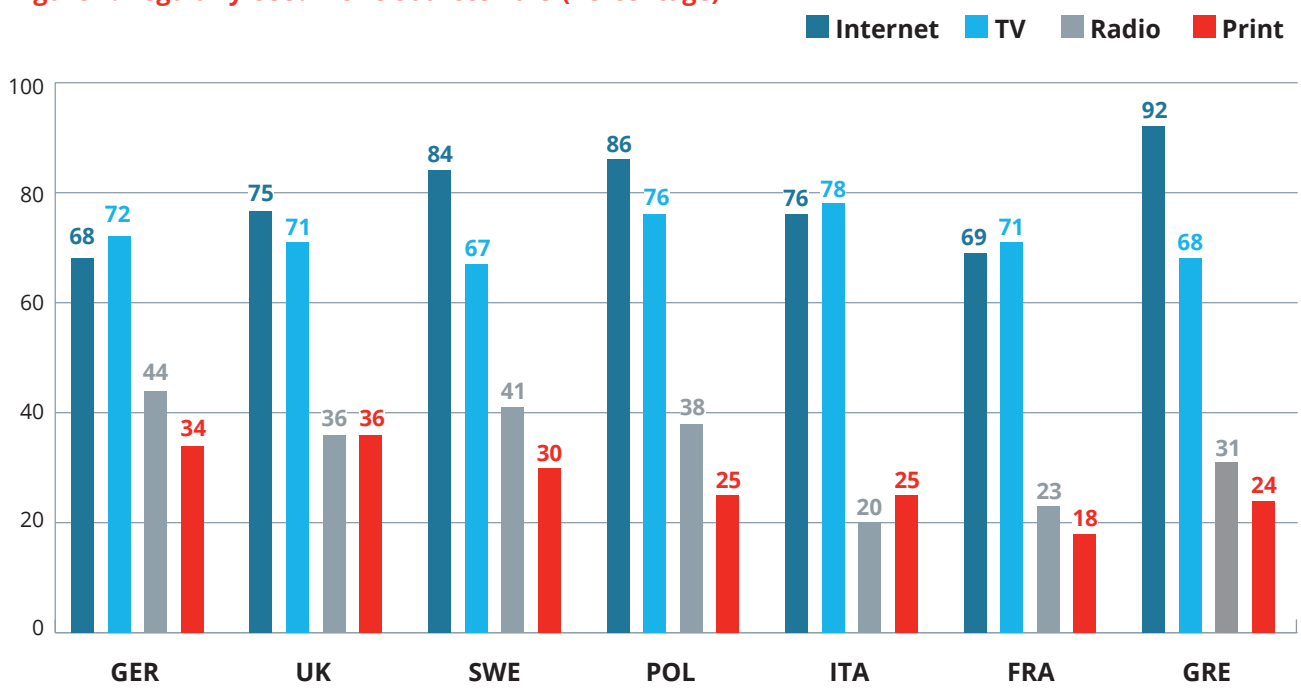
The Reuters Institute Digital News Report provides information on longer-term trends in media use and in particular examines the role of social media in the context of online news usage. National characteristics can also be assessed in comparison with other countries. The study is coordinated by the Reuters Institute for the Study of Journalism, based in Oxford (UK) and has been conducted since 2012 in more than 30 countries (Newman et al., 2019, p. 4). Different panel partners are responsible for the surveys in each country. For example, since 2013, the Leibniz Institute for Media Research | Hans Bredow Institute has been a cooperation partner for the German partial study. In early 2019, fieldwork was conducted by the YouGov survey institute, which drew random samples based on online access panels that were representative of Internet users in the participating countries aged 18 and over, i.e., structurally identical to the Internet population regarding the variables age, gender, region, and education, or weighted accordingly. Surveys in each country use different sample sizes.

As data on media use in 24 countries was collected in the European context (Newman et al., 2019, p. 4), a selection of countries is used here for clarity of presentation. The aim was to investigate different developments with regard to the role of algorithmic news channels and correspondingly different media structures, while at the same time considering different regions (Northern, Eastern, Southern and Western Europe) and including extreme examples in order to illustrate the heterogeneity of the prominence of digital news. Germany, for which more detailed analyses are available (cf. Hölig & Hasebrink, 2019), was used as a case study to more precisely illustrate the role of intermediaries in the media repertoire.

Regularly Used News Sources 2019

On a general level, results from the current Reuters Digital News Report confirm the growing importance of online news sources in European media repertoires. Figure 1 shows the share of the four "classic" media types (TV, Radio, Print, Internet) regarding news usage in the selected countries. It depicts how many respondents watched news on television, read any print medium, listened to radio news and consumed online news at least once a week.

Figure 1: Regularly Used News Sources 2019 (Percentage)



Source: Reuters Institute Digital News Survey 2019

Bases: GER: n = 2022; UK: n = 2023; SWE: n = 2007; POL: n = 2009; ITA: n = 2006; FRA: n = 2005; GRE: n = 2018

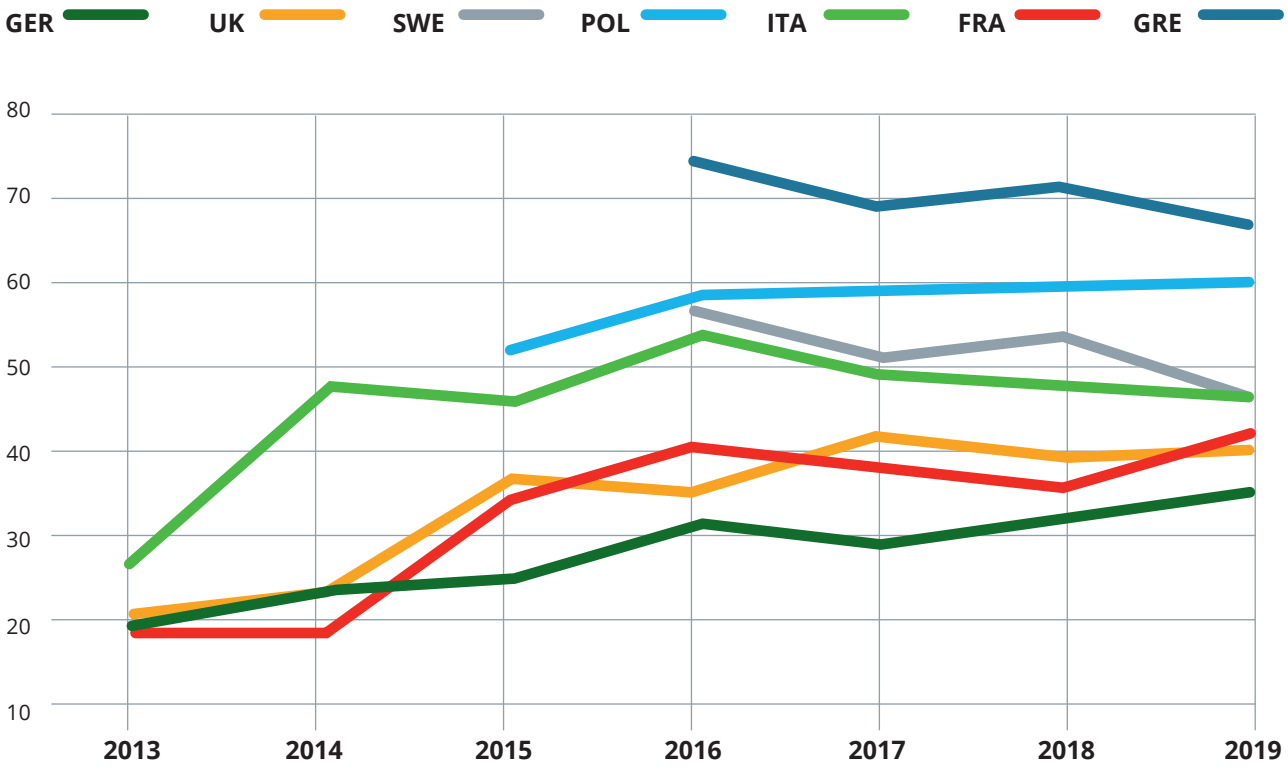
The data show that in all selected countries, television and the Internet are the most popular news sources, while radio and print news are consumed by a considerably smaller number of citizens. It also becomes evident that television, traditionally the most used news source, has been already overtaken by the Internet. The major implication of this development—which probably will continue in the future—is that online news sources, including various algorithm-based ones, already have a considerable impact on the distribution of information in European societies. In some countries, such as Greece, Sweden, and Poland, this impact is larger than in others (e.g., Germany, France, and Italy), but it is tangible everywhere.

The German case is a prime example here: although television remains the main source of news, the trend indicates a decline. Whereas in 2018 almost half (49%) of German online users stated that television was their main source of news, this figure had dropped to 45% by 2019. This decline was in favor of the Internet, which is the most important news source for 36% of respondents (2018: 32%). The difference is highest in the youngest age group (18 to 24 years), where more than two-thirds (69%) of respondents cited the Internet as the most important source of news, an increase of ten percentage points over the previous year (Hölig & Hasebrink, 2019, p. 20).

Social Media as a News Source Over Time

To quantify the importance of intermediaries, a more detailed look at the available data is necessary. In this context, the share of news users receiving their news from social media is crucial. Figure 2 depicts the development of the prevalence of social media as a news source over time between 2013 (the first wave of the Reuters Digital News Survey) and today. In most countries, a generally positive but not neces-sarily linear (e.g., POL, UK, GER, and FRA) trend can be deduced. A detailed look at the numbers for 2019 reveals that in many European countries—the coun-tries displayed here are exemplary—the percentage of the population that uses social media as a news source ranges from about 30 to 60%. A considerable amount, but—especially in comparison to Internet news usage in general and other sources—not an indication that intermediaries and thus algorithms are the dominant news source (Newman et al., 2019). There are, however, exceptions such as Poland or Greece, where social media is an above-average popular news source. In general, it can be assumed that the influence of social media in media repertoires will continue to grow since the majority of the younger generations focus their news usage on the Internet, including various social media platforms such as Facebook, Instagram, or WhatsApp (Newman et al., 2019, p. 55-59)

Figure 2: Social Media as a News Source Over Time (Percentage)



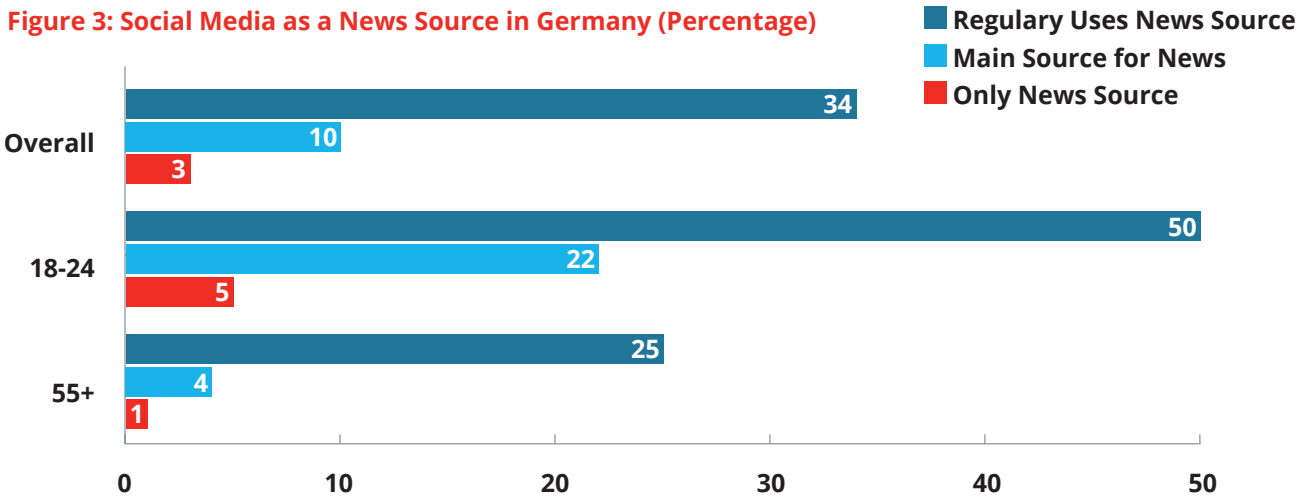
Source: Reuters Institute Digitale News Surveys 2013-2019. Bases varying

Social Media as a News Source in Germany

Another important aspect to quantifying the impact of algorithmic news is to examine the relevance of social media in news repertoires. The idea behind this is that the negative impact of algorithmic news selection on content diversity may be especially applicable when social media is used (almost) exclusively for news consumption and is thus not embedded in a broader mix of different news sources.

As this analysis is only available for Germany, this section focuses on the German partial study (Hölig & Hasebrink, 2019, pp. 21–22). Figure 3 depicts the number of German respondents who use social media as a news source in general, as their main news source, and as their only news source. The numbers show that a negligible number of users limit their news consumption to social media (3%). Even among the younger generation, which relies more on social media in general, the proportion remains stable at a low level of 5%.

Figure 3: Social Media as a News Source in Germany (Percentage)



Source: Reuters Institute Digital News Survey 2019 Base: n = 2022

In sum, the data show that social media is relevant for opinion formation. Every third online user has seen news on one of the platforms within an average week (34%). Among 18 to 24-year-olds, the figure is 50%. Social media dominates as the most important source of news on the Internet compared to other online sources in the age groups under 45 years (no figure). This is a result that only applied to the group of 18 to 24-year-olds in the previous year (Hölig & Hasebrink, 2019; p. 20).

Intermediaries as a News Source in Germany

The data from the Reuters Digital News Report presented so far provide valuable insight into the role of social media in opinion formation. However, it is not possible to give a more detailed picture of the importance of other intermediaries, such as search engines or news aggregators on the basis of the available data across countries. Therefore, we again turn to the case study of Germany and data from the "Medienvielfaltsmonitor"[2] (Media Diversity Monitor) of the Landesmedienanstalten (State Media Authorities) to elaborate on the importance

2 The Media Diversity Monitor continuously empirically examines and documents the development of the German media landscape. The study records the balance of power on the opinion market across different media types by, for example, calculating the daily ranges of individual media providing information (e.g., intermediaries, newspapers, or television) and determines the importance of media types for opinion formation (die Medienanstalten, 2019).

of intermediaries in the media repertoire of the German population.

The data show that intermediaries, such as Facebook, Google, or YouTube, have become firmly anchored in the media repertoire: a third of Germans obtains information from at least one intermediary on a daily basis. This proportion almost doubles in the younger age group (14 to 29-year-olds; 63%). For informational purposes, search engines (24% reach) are ahead of social networks (17%) and video portals like YouTube (8%). Here, the overall picture is changing in the younger age group as well, where the daily reach of social networks (42%) is almost level with search engines (44%). At 22%, video portals are also firmly anchored in this age group.

In summary, almost two-thirds of younger people obtain information from at least one intermediary on an average day. In the middle age group (30–49 years) this share is 40% and in the over-50 age group it is only 17% (Zimmer & Kunow, 2019, pp. 46–50). When the data are broken down by platform, it is not surprising that usage is also concentrated on Google, Facebook, and YouTube—both among the overall population and among younger people (Zimmer & Kunow, 2019, pp. 50–51).

Based on the data presented in the previous sections, an interim conclusion on the role of intermediaries in opinion formation is as follows: Intermediaries play an increasingly important role in European media repertoires, qualifying them as a potentially veritable factor for processes of opinion formation. The dominance of the position of intermediaries is thereby dependent on both the age group, whereby intermediaries are more prominent and relevant for young users, and to a certain extent on the respective country. The data from the Reuters Digital News Report show a generally favorable trend for the anchoring of online news sources in the European media repertoires, but there are still considerable differences across the selected countries. For our case study—Germany—it also becomes evident that intermediaries are still mostly embedded in a relatively broad mix of various media outlets, suggesting that in democratic countries with a diverse information environment, traditional media can act as a corrective for possible negative effects of intermediaries on opinion formation.

Main Findings:

- News consumption is changing: news sources on the internet are gaining in importance. In particular, social media plays a vital role among young people; therefore algorithmic curation is increasingly influencing news consumption.
- In some countries, such as Greece, Sweden, and Poland, this impact is larger than in others (e.g., Germany, France, and Italy), but it is tangible everywhere.
- The importance of intermediaries as news sources will continue to grow in the future.
- In Germany, it becomes evident that intermediaries are still mostly embedded in a relatively broad mix of various media outlets. Or to put it in another way, intermediaries are seldom the only source of information on political topics.
- To deepen our understanding of the effects of intermediaries, we should investigate patterns of media use across multiple media outlets by taking a media repertoire approach.

4.2. Filter Bubbles and Echo Chambers – Overview of the Academic Body of Work on the Effects of Algorithmic Gatekeeping and Personalization

Chapter 3 focused on widespread theoretical assumptions in the context of personalized news sources. But how realistic are they? Are the effects on public discourse really as significant as common sense seems to suggest? This question is addressed in the following summary of the state of research. Research in this area is virtually exploding, which is why it is impossible to give a complete overview. Instead, central empirical studies are differentiated along the focus of their research interest: the impact of algorithmic personalization on searching the Internet, its influence on social media, and the effects on opinion-formation processes. This approach allows the identification of key findings and the evaluation of the research designs used.

The studies conducted in different countries focus on Facebook, Twitter, search engines like Google, and news aggregators like Google News. In particular, they attempt to determine the degree of personalization more closely (e.g., through simulation, manipulation, and observation or experiments in real or artificial settings) (cf. Ørmen, 2018 for this systematization) and thus transfer the known effects of selective exposure to technical personalization processes ("preselected personalized communication"; Zuiderveen Borgesius et al., 2016). They test the effect of both implicit and explicit personalization on the content and source or viewpoint diversity. Many research designs show the same methodological problem: the difficulty in simulating personalization effects according to real usage behavior. The general functioning of such methods is very simple, but in practice often becomes complex and difficult to predict due to the large number of variables involved and the unpredictable, dynamic nature of the data used. Another fundamental problem emerges from the profound lack of robust definitions for these terms. Therefore, empirical studies exploring the existence and impact of filter bubbles (and echo chambers) have generally been forced to apply their own definitions, which reduces their comparability. According to Bruns (2019b, p. 3) "this terminological confusion – about the exact definitions of either term in itself, and about their interrelationship with each other – has significantly hindered our ability to test them through rigorous research." Moreover, it is important to note, that these results stem from studies that were conducted in different countries and utilized a variety of methods (Bruns, 2019b).

Personalization Effects When Searching the Internet

Empirical studies taking the filter bubble metaphor as their starting point, such as Pariser (2011), originally focus mainly on possible personalization effects when searching for information or news (especially via the market leader Google Search as well as Google News). For example, they measure the degree of overlap of search results, similarity of hits between different users or the diversity of sources, issues or viewpoints received through Google News or Google Search compared with other information sources. Under artificial settings, Dylko and colleagues (Dylko, Dolgov, Hoffman, Eckhart, Molina, & Aaziz, 2017) observed filter bubbles in experiments with students. However, studies on the effects of implicit personalization in real settings find little empirical evidence for the threat scenario. Only explicit personalization fosters exposure to content that thematically fits personal preferences. Overall, the findings show that the much-cited effects of the filter bubble are significantly smaller than previously assumed or even completely absent (e.g., Dubois & Blank, 2018; Haim, Graefe, & Brosius, 2018; Krafft, Gamer, & Zweig, 2018; Nechushtai & Lewis, 2019).

For example, in Germany, the so-called "Datenspendeprojekt" ("data donation project") before the 2017 federal elections, analyzed a list of almost 280,000

Google search results on prominent politicians. More than 4,000 volunteers "donated" their results for certain search terms via a plug-in. The analysis showed that seven to eight out of an average of nine results did not differ, while the differences were mainly due to regionalization tendencies. The authors interpreted these findings as a clear indication that personalization by search engines plays a smaller role than previously assumed or that search engines personalize their results much less than presumed (Krafft et al., 2018; Puschmann, 2018).

A study by Haim, Graefe, and Brosius (2018) also shows only minor personalization effects. Based on two exploratory studies measuring the impact of both implicit and explicit personalization on the diversity of content and sources on Google News, the authors conclude that the relevance of algorithmically generated filter bubbles is overrated. The design used to measure implicit impact potentials worked with four virtual user accounts, whose profiles were "trained" within a week using various search terms and purchasing processes. At the end of the training week, the researchers saved the search hits on various topics. The ranking and exclusive content only marginally differed between the four test accounts and an "untrained" control profile. A simulation study by Jürgens, Stark, and Magin (2014) comes to a similar conclusion.

Thus, following the works of several authors, Google and Google News seem to aim less at providing specifically personalized content. Rather, a general bias is apparent because the algorithm overrepresents certain media outlets—whose content is probably particularly well search engine optimized and accessible without payment walls—compared to less well adapted but higher-reach providers (Bruns, 2019a; Courtoius, Slechten & Coenen, 2018; Haim et al., 2018). For the US, Nechushtai and Lewis (2019) state that the news agenda of Google News reproduces that of traditional media, since users are mainly referred to four or five mainstream news sources. According to a study by Magin, Steiner, and Stark (2019), more than 60 percent of the results of five different search engines on current political issues also come from professional journalistic media in Germany.

Hence, the idea that we all live in our own filter bubbles has to be considerably qualified. How permeable or closed filter bubbles really are, is decisively dependent on the individual tendency toward selective exposure of users, their personality traits, and their conscious selection behavior. Therefore, it is not the general question about the existence of filter bubbles that is relevant, but rather the indication that their walls are "porous" (Zuiderveen Borgesius et al., 2016). From a methodological point of view, however, it is still true that the acquisition of data is very complex and better insights into the "black box" of algorithms remain essential for scientists.

Personalization Effects on Social Media

This requirement is particularly applicable to studies that investigate the effects of implicit personalization on Facebook. Methodologically, they are particularly difficult to implement because Facebook strictly regulates access to possible data interfaces and has become even more restrictive in recent months due to many scandals (e.g., the Cambridge Analytica data scandal). In most cases, investigations in this field are therefore experimental or work with tracking data. However, in the social media context, these studies differentiate even less between filter bubbles and echo chambers and frequently also dispense with an explicit theoretical conceptualization of the metaphors (Bruns, 2019a, p. 5). When interpreting the results, the references to the overall societal consequences—especially fragmentation and polarization—are strongly accentuated.

Many studies are limited to Twitter data because they are particularly easily accessible to researchers. This focus on a single platform with a very specific elite user group, in which higher educated people and in

particular journalists or politicians are clearly over-represented, means that they present only a small part of reality. Through network analysis, inferences can be made about the distribution of news characteristics as well as about the characteristics of interactions and relationships. A series of these studies show segregated, ideologically consonant networks that are only slightly interconnected (e.g. Conover et al., 2011; Gruzd & Roy, 2014; Himelboim, McCreery & Smith, 2013). However, this is often limited to certain topics, political milieus or politically active actors (e.g., Barberá 2015; Williams, McMurray, Kurz, & Lambert, 2015). In addition, some studies (e.g., Garimella, De Francisi, Morales, Gionis, & Mathioudakis, 2018; Smith & Graham 2019) specifically select emotional, morally charged issues from specific hashtags (e.g., #obamacare, #guncontrol, #abortion), so that the documented polarizing effects are not surprising.

A very early study by Bakshy, Messing, & Adamic (2015) that was created in cooperation with Facebook received much attention. The researchers analyzed the news feed of about ten million Facebook users in the U.S. who had stated their political viewpoints (liberal vs. conservative) in their profiles. According to the results, it is not so much the Facebook algorithm that causes less contact with political information from the opposing camp. Rather, the users themselves tend to choose the information that corresponds to their own political position from their diverse news feed, thereby relativizing the influence of algorithm-based selection decisions. It is in particular the composition of the personal network that limits the perceived diversity of opinion. Nevertheless, the algorithm has a documented, albeit weaker, effect. Since both effects add up, the algorithm contributes to political filtering but its influence is less strong than expected.

Similarly, Flaxman, Goel, and Rao (2016) investigated the web browsing behavior of 50,000 Internet users in the U.S. They used a data set from Microsoft collected from users of an Internet Explorer toolbar. Using automated content analysis, the authors analyzed different access ways to online political news (i.e., search engines, social media, news aggregators, and direct access to web sites). The findings showed that users of politically left- or right-wing media are less frequently exposed to conservative content and vice versa. However, the level of exposure to opposing opinions via social media and search engines was higher than for direct visits to websites or news aggregators.

There is also evidence in follow-up studies that the use of social media tends to increase exposure to opposing viewpoints because it extends the range of content that is used. Contrary to the theoretical assumptions of the metaphors, such studies draw the conclusion that users do indeed come into contact with more diverse or contrasting sources and content (e.g., Beam, Hutchens, & Hmielowski, 2018; Beam & Kosicki, 2014; Dubois & Blank, 2018; Flaxman et al., 2016; Moeller, Trilling, Helberger, Irion, & de Vreese, 2016).

This result can be explained by opposing mechanisms, such as “incidental exposure” (Fletcher & Nielsen, 2018), network effects (Bechmann & Nielbo, 2018; Bode, 2016; Messing & Westwood, 2012; Scharkow, Mangold, Stier & Breuer, 2020) and the influence of the algorithms themselves, which—as described above—preselect more diverse content than expected (Moeller, Trilling, Helberger, & van Es, 2018; Nechushtai & Lewis, 2019).

Social Media and Polarization of Opinion

There are contradictory findings regarding the relationship between the use of personalized news channels and the polarization of opinions. For example, several studies confirm polarized usage tendencies for the USA depending on political predispositions and attitudes—but in some cases only for certain topics (Hagen, Wieland, & in der Au, 2017; Nelson & Webster, 2017; summarizing see Bright, 2018; Schweiger et al., 2019). However, these findings cannot simply

be transferred to other countries since the competitive political system of the USA per se causes stronger polarization than the more consensus-oriented political systems in many European countries.

Various (experimental) studies—also using German data—analyze the assumed effects of the use of personalized news sources, especially with regard to processes of opinion formation in the online public sphere. They aim to identify psychological processes, in particular personality traits that either inhibit or foster an attitude-congruent selection of news content. Based on the spiral of silence-theory, both the perceived climate of opinion on relevant issues and the articulation of opinions and its influence on dependent variables, such as political participation or polarization of opinion, are examined (Gleich, 2019; Magin et al., 2019; Schweiger et al., 2019; Weber, Mangold, Hofer, & Koch, 2019).

In addition, various survey studies compare, particularly with regard to the concept of media repertoires, the effects of different news sources and intervening variables, which could promote or inhibit the formation of echo chambers (Auxier & Vitak, 2019; Dubois & Blank, 2018; Magin et al., 2019; Yang et al., 2016). In a broader sense, these analyses include sociodemographic characteristics (e.g., age, education, and gender), general media usage patterns and media-related attitudes (e.g., selective exposure tendency and skepticism toward personalization), politics-related characteristics (e.g., political interest, political orientation, and self-efficacy perception) and psychological personality traits (e.g., Big-Five) (Schweiger et al., 2019). However, these studies do not provide a uniform picture of whether the increased use of personalized news services leads to increasing polarization. Yang et al. (2016), for example, were only able to demonstrate correlations between online news use and a polarization of personal attitudes for certain countries in a cross-country comparison study. In nine out of ten examined countries, the subjectively perceived polarization increased with the intensity of online news usage. This effect can be interpreted in the sense that social media platforms make extreme positions visible but do not necessarily generate or reinforce them. Dubois and Blank (2018) also re-evaluate the relevance of echo chambers based on a representative survey in Great Britain: according to them, the risk of entering echo chambers is minimized for people with a diverse news repertoire and high political interest. Furthermore, they criticize measurements that focus on one news platform and emphasize how important it is to examine the use of different sources of information as a whole.

To summarize, both the highly fragmented field of research and the lack of a uniformly applied polarization concept make it almost impossible to specify an overall effect of personalized news sources on polarization of opinion. In addition, different effects manifest themselves with regard to personality traits (e.g., political orientation, interest in politics, and strength of partisanship) for certain population groups (Barberá, in press).

4.3. Concluding Remarks: The Overestimated Effects of Algorithmic Personalization

Our overview of the current state of research shows that the actual scope of filter bubbles and echo chambers is widely overestimated. From a theoretical point of view, both phenomena appear plausible, especially since they are often underpinned by vivid descriptions of individual cases in the public discourse. But why are these phenomena so difficult to confirm by empirical studies? Considering the implicit premises of both concepts, the answer to this question becomes more tangible: in their ideal-typical form, filter bubbles presume that users are not interested in diverse news and recommendations and that the algorithms neither recognize nor satisfy their needs in terms of their interests. At the same time, users may only use algorithmically personalized

information sources and not be aware of their effects (Bodó, Helberger, Eskens, & Moeller, 2019; Moeller et al., 2018; Stark et al., 2017). To create echo chambers, users must be involved online in very homogeneous networks, in which all members share their opinions. However, social media are particularly suited to maintaining volatile or sporadic contacts with friends from various contexts (e.g., former schoolmates, neighbors, or holiday buddies). Such "weak ties" promote the diversity of personal networks more than limiting it (Barberá, 2015; Granovetter, 1973).

Moreover, the potential negative consequences of both phenomena for democracy are based on the premise that people intensively use algorithmically personalized news sources for political information, but this only applies to a minority (see chapter 4.1). Notably on social media, many users rarely or never come into contact with political information (Stark et al., 2017) and do not choose their network of friends according to political worldviews (Duggan & Smith, 2016). All these mentioned premises should apply to large parts of the population. However, this scenario is difficult to imagine.

4.4. Example of Research Requirements: How to Examine Echo Chambers?

To illustrate some of the challenges involved in research on intermediaries' effects, we present a concise, realistic example of how such a study might be designed. Let us assume—for the purpose of this thought experiment—that the question at hand concerns echo chambers. If users get their news from and exchange viewpoints within groups which have a single, clearly defined political leaning, this puts them at risk of obtaining an incomplete, biased or even factually incorrect view of the world. In our hypothetical example, we would therefore like to know whether such tightly knitted and narrow-minded communities exist on Facebook, how many users are part of an echo chamber, and whether they lead to negative consequences for individuals and society at large.

Research question: In formal terms, the goal of this study would be to measure the ideological markup of individual information environments, specifically their homogeneity. Users are considered to be part of an echo chamber if they participate in a group whose expressed or implicit viewpoints show very high consistency. Viewpoints may be defined as either partisan ideology (such as measured in a survey), news consumption (selecting strongly partisan sources such as Breitbart), or expressed viewpoints (supporting strongly partisan positions such as xenophobia). Consistency is measured by the overall average prevalence of viewpoints that run counter to the majority position.

Data requirements: The concept of echo chambers utilized here is based on data about groups, not single individuals. Determining the internal ideological homogeneity ideally requires a full sample of the group's activity—otherwise it would be easy to miss decisive non-conforming messages. Communication within the group may take place in a number of distinct channels: (1) As comments on publicly available pages (i.e., the pages of news outlets, parties, politicians, NGOs etc.) (2) as posts and comments within a publicly accessible group (i.e., a NGO's Facebook group), (3) as posts and comments within a private (invite only) group, (4) as messages within a Facebook-owned messaging service (Facebook Messenger or WhatsApp), or some other, less common or hybrid form. The corresponding type of data from Facebook's API would be: (1) All comments from a designated set of pages, including those that were deleted by Facebook itself, the page administrators, or users themselves (these deletions would reveal crucial insights about the reason for a group's homogeneity). (2) All posts and comments from designated publicly available groups, including the deletions as mentioned in (1). (3) All posts and comments from designated private groups, including the deletions

mentioned in (1). (4) all messages exchanged in a messenger group.

Practical data availability: The four distinct channels where echo chambers might form are accessible to researchers to different degrees. Within the framework of Facebook's current rules on data collection, (1) has limited feasibility: The 600 most recent posts and comments from public pages are available (access is granted after a review from Facebook), but deletions are not. (2) Accessing group content through data collection software requires approval by a group administrator. This prevents research on groups who reject scientific inquiries into their composition. (3) Private groups are inaccessible to data collection. (4) The content of messenger systems is inaccessible to data collection; furthermore, WhatsApp messages are end-to-end encrypted, meaning even Facebook has no access to their content.

Ethical considerations: Most ethical and legal settings (such as the GDPR) prevent collection of individuals' protected data without explicit and prior consent. A study as outlined above could either (1) only resort to publicly available (non-private) data, mostly from channel 1 (public posts), or (2) attempt to obtain prior consent from subjects (plausible for channels 2-3 (groups) and channel 4 (messengers)), or work with Facebook to receive pre-anonymized or aggregate data.

Feasibility of study: Given the constraints outlined above, only one of the four relevant channels could be studied without dedicated help from Facebook: The presence of echo chambers in comments on publicly available pages. Even in this case, the study would be limited by the lack of access to historic and deleted comments. While comments on political or news posts are clearly important, the concept of echo chambers aligns much more closely with semi-private or private exchanges, which for now remain outside of researchers' reach.

As this example has shown, the path towards a conclusive and accurate assessment of echo chambers on Facebook is blocked by multiple competing interests, legal requirements, the highly limited nature of Facebook's API, and technological measures for privacy protection.

5. Beyond the Filter Bubble: Challenges for the Traditional Media and the Public Discourse

Having analyzed the state of research on filter bubbles and echo chambers, this section will now discuss other risks of the digital transformation of the news ecosystem. We argue that it is extremely important to move beyond the filter bubble discourse and discuss challenges for traditional media in a broader sense. In particular, whether the adaptation to social media logics leads to a decline of media quality or a softening of news. In this context, worries about a dissemination of disinformation and fake news have increasingly been voiced in the public debate, often in combination with warnings about an increasingly disrespectful discourse culture around controversial political issues.

Accordingly, in the context of a narrative literature review, we discuss possible causes of the perception of an uncivil, low-quality discourse in social media. At first, we consider and define disinformation and fake news and outline how both phenomena are a veritable threat to a free public discourse from a theoretical point of view. We conclude this section by questioning whether disinformation negatively impacts public discourse as heavily as the public debate suggests. We then focus on the increasing incivility in online discourses, point toward negative effects on the users of social media, which subsequently affect public discourse in general and discuss various countermeasures. The implications arising from this critical summary for opinion-formation processes will subsequently be presented in chapter 6.

5.1. Softening the News

Against the backdrop of a changing media landscape a major concern is that economic constraints and the increasing competition for audience attention will threaten the quality of political news in the media (van Aelst et al., 2017). This mainly affects traditional mass media, which are no longer able to maintain their high quality due to economic pressures caused by structural changes in the news market.[3] Traditional media companies are facing unprecedented competition from information intermediaries: they no longer have to compete only with other media companies or non-journalistic providers, but also with tech giants that offer users services similar to their own. The few "Internet giants" account for a large part of advertising revenues, without a substantial transfer of these funds back to news journalism. This competitive situation creates completely new conditions for the financing of quality journalism. To survive, media content must be displayed on the platforms and ranked as highly as possible. Therefore, media are becoming increasingly dependent on the platforms.

Softening the news might be a strategy to adjust to these rules. These trends, often labeled with buzzwords such as sensationalism, tabloidization, infotainment, or soft news, are not new. However, the debate has flared up in recent years, particularly in

3 This concern is based on the widely shared assumption that high-quality political news is crucial for public knowledge and a prerequisite for a healthy democracy (Napoli, 1999).

the context of social media and in terms of the often so-called social media logics.[4] It is argued that the competition for audience attention and advertising revenues has a negative impact on the production and presentation of news (Karlsson, 2016). Concrete fears relate to the assumption that journalists will use soft news strategies to adapt to the social media logics—and present their news, for example, in a more sensationalist or emotional style—in particular, against the background that a growing number of younger users make social media their main news source (Newman et al., 2019). The social media logics favor user engagement, shareworthiness, and virality—including using clickbait headlines in their attempts to gather clicks and to gain attention. Such curiosity headlines (Blom & Hansen 2015; Kuiken, Schuth, Spitters & Marx, 2017) seek to encourage audience engagement by leaving important information unstated. These headlines contrast with summary headlines that give an overview of the main facts of news stories.

Accordingly, the debate about "softening the news" has a new, reinvigorated relevance (Lischka, 2018; Martens, Aguiar, Gomez-Herrera & Mueller-Langer, 2018; Moeller et al., 2018; Otto, Glogger, & Boukes, 2017; Welbers & Opgenhaffen, 2019). The concept relates mainly to the topic of a story and its style or presentation mode (Reinemann, Stanyer, & Scherr, 2016). Therefore, "softening" usually describes specific characteristics of news item features but can also refer to changes in the journalistic system on the macro level (Otto et al., 2017). The term is often associated with concepts such as tabloidization and popularization, which suggest a trend over time. A more recent framework model of softening of political communication highlights four prominent concepts (sensationalism, soft news, infotainment, and tabloidization) on different levels of investigation (i.e., item, genre, or type of media). These concepts represent the particular forms of softening on each of these levels (Otto et al., 2017).

However, no overall trend can be detected for the declining reporting quality in political communication. Insofar as comparative or longitudinal studies are available, they document fluctuating trends—even across different media outlets (van Aelst et al., 2017). Research evidence about longitudinal changes in the supply of hard versus soft news is also mixed, as Reinemann et al. (2016) show, the amount of soft news varies significantly across countries. One of the few studies that examined the popularization of news (combining sensationalization, scandalization, emotionalization, common people narratives, and privatization of public figures) over a longer period (1960s to 2010s) found an increase over time, mainly in the US and UK media, but no convergence in the popularization of political news (Umbricht & Esser, 2016). Another important long-term study from Switzerland (The Yearbook Quality of the Media) also documents the increasing relevance of soft news.[5] Although the study attests to the examined media a good performance by a wide range of professional journalistic standards, there is a decline regarding "relevance" and "diversity," which means that not all quality dimensions are equally affected. Soft news is gaining in importance; in particular, the explanatory, background political reporting has decreased significantly over the years. The loss of resources is noticeable here because investigation and in-depth reporting are known to be particularly resource-intensive (Vogler, Eisenegger, Schneider, Hauser &

4 Concretely, this term refers to the replacement of "mass media logic" by "social media logic" (van Dijck & Poell, 2013 or Klinger & Svensson, 2015) and the influence of this "social media logic" on journalism. In this context, the urgent question arises as to whether "quality media" can meet the high standards set for them. The concept of network media logic or social media logic regards how social media platforms change political communication. It discusses the different ways of producing content, distributing information and using media as well as the consequences for traditional media.

5 The quality of reporting is measured using a content analysis of the daily output of 64 media outlets in Switzerland in all three language regions. By including newspapers, online news sites, radio and television newscasts or news magazines all relevant types of news media were considered.

Udris, 2019). Comparably, van Aelst et al. (2017) came to the conclusion that one must be more concerned about shrinking resources for journalism and increasing quality differences between media outlets, which can result in different knowledge gaps among users than for a universal trend toward soft and entertaining news (van Aelst et al., 2017, p. 10).

In addition, global tech companies are increasingly forcing news content as a separate business segment, thereby creating new competition in the field of journalism. Since 2015, Google and Facebook have launched a series of initiatives to "gain a foothold in information journalism" (Gisler 2016). These include Google's "Digital News Initiative", Facebook's "Instant Articles", "Twitter Moments", "Apple News" and "Snapchat Discover". There is a common idea behind these projects, namely to offer established media houses new distribution channels or financial project support in order to make journalistic content more easily findable and consumable on the internet and to develop new digital monetization strategies. However, the rules for this are dictated by Facebook or Google, and the conditions are set in their favor.

In 2017, Emily Bell and Taylor Owen had already given a very negative assessment of the change processes in the news ecosystem. In their study, they examined the influence of platforms such as Facebook, Google, Snapchat and Twitter on journalism and the rapid takeover of the roles of traditional publishers by these companies (Bell & Owen, 2017). The authors also stressed that the structure and economic functioning of platforms favor the distribution of low-quality content: "Journalism with high civic value—journalism that investigates power, or reaches underserved and local communities—is discriminated against by a system that favors scale and shareability" (Bell & Owen, 2017, p. 10). One of the main dilemmas of publishers was described as follows: "Should they continue the costly business of maintaining their own publishing infrastructure, with smaller audiences but complete control over revenue, brand, and audience data? Or, should they cede control over user data and advertising in exchange for the significant audience growth offered by Facebook or other platforms?"

The ongoing, multi-year study by the Tow Center for Digital Journalism at Columbia Journalism School into the relationship between large-scale technology companies and journalism, shows in the latest phase, that platforms have become more explicitly editorial in their own practices and structures (e.g., in the case of Apple News, which publishes its own "exclusive" content from a newsroom, staffed with writers and editors). Whereas news organizations are showing signs of pushing back with strategies that help them retain some autonomy and control over these processes, e.g., bringing audiences back to their own sites. The belief that large technology platforms could lead to significant advertising revenue for publishers has been debunked. However, it is clear that journalistic outlets remain affected by constant change and uncertain revenues (Rashidian, Tsiveriotis, & Brown, 2019).

Many empirical questions are still unanswered, in particular there is little empirical evidence regarding the question if the influence of "social media logic" is contributing to the above-mentioned trends. Empirical results provide indications about different levels (e.g., subjective language, clickbaiting, or changing editorial values on social media). Various recent studies show that emotional and surprising story elements are becoming more prominent on Facebook because editors assess the news values of emotion and surprise as being more important. Thus, news editing for news feeds has become more geared toward user engagement and is more algorithm driven. Posted news favor entertaining topics, other news, such as international politics and economic news, are neglected (Lischka, 2018, see also chapter 2.1). Moreover, subjective language in the news might be an essential part of the social media logics in news reporting. A comparison between the classic headlines, leads of news items to the social status

messages, and status messages with respect to the amount of subjectivity and the polarity (i.e., positive vs. negative), confirm this assumption. Hence, based on content analysis data, a shift toward a more subjective and positive style of news reporting could be confirmed (Welbers & Opgenhaffen, 2019). A study by Blom & Hansen (2015) also provides initial evidence that tabloidization leads to an increase in the use of forward-referencing in Danish online news headlines. The authors examine clickbaiting in the form of forward referencing in online news headlines and find stronger tendencies in commercial and tabloid media for using such forward referring headlines. These results show that platforms continue to shape both the style and content of traditional media on Facebook.

In a nutshell: the growing orientation of journalists toward both audience demands and algorithms affects journalistic news selection by fostering a self-reinforcing feedback loop since algorithm-based decision-making is necessarily oriented toward the audience and the "mainstream." The mutual observation among the users themselves further strengthens the orientation toward entertainment—off hard news toward more humor and emotion. That is, media logics adapt to social media logics to some degree.

5.2. Disinformation and Fake News

This section introduces and analyzes the issue of disinformation and fake news. It is shown that fake news is only one facet of the broader matter of disinformation campaigns that stand for a fundamental shift in political and public attitudes to what journalism and news represent and how facts and information may be received in high-choice media environments. Disinformation has harmful effects on public discourse by inhibiting the free individual and collective formation of opinion and political will. Intermediaries, and in particular social media, are susceptible to fast and wide-spread dissemination of disinformation due to their attention-oriented logic that can be exploited by the various producers of disinformation. We show that empirical evidence on the scope and effects of disinformation is scarce so that further research is needed to illuminate the issue further.

5.2.1. Theoretical Perspectives on Disinformation

Defining Disinformation and Fake News

Fake news has become an integral part of the political and public debate, most recently in the campaign for the 2016 U.S. presidential election (Martens et al., 2018, p. 8). At the center of the debate is its allegedly wide and fast dissemination through social media (Zimmermann & Kohring, 2018, p. 526). The existing literature shows many, heterogeneous definitions of fake news (Egelhofer & Lecheler, 2019; Zimmermann & Kohring, 2018) from which two dimensions of the term emerge: the *fake news genre* describes "the deliberate creation of pseudojournalistic disinformation", while the *fake news label,* points to "the political instrumentalization of the term to delegitimize news media" (Egelhofer & Lecheler, 2019, p. 97). Despite its continuing popularity, the term has come under considerable scientific critique: the German communication scientists Philipp Müller and Nora Denner (2019, p. 6), for example, point out that the fake news genre has increasingly been used ambiguously in social and academic discourse, with reference to not only verifiably false information presented in a news-like way, but also inaccurate or decontextualized information, slanted news, or has been applied in non-news contexts, such as historical information or scientific studies.

While some authors therefore reject the term fake news altogether (e.g., Wardle & Derakshan, 2018) and others plead to restrict its usage to the fake news label (Zimmermann & Kohring, 2018, p. 527), it can be argued that it remains relevant, if only because it is so prominently placed in the public debate (Müller

& Denner, 2019, p. 8). However, it is necessary to distinguish between the various distinct phenomena that are often summarized under the catchphrase fake news. Following a systematization proposed by Wardle and Derakshan (2018, p. 20–22), this report therefore recommends differentiating between *misinformation,* that is, when false or misleading information is unintentionally shared*, disinformation*, the conscious and deliberate dissemination of false and misleading information with a harmful intent and *malinformation*, a term referring to the sharing of genuine information to cause harm (e.g., leaking private or secret information). In this categorization, fake news is classified as a subcategory of *disinformation*, which can be defined "as verifiably false or misleading information that is created, presented, and disseminated for economic gain or to intentionally deceive the public, and may cause public harm." (EC, 2018b, p. 1)—a definition excluding and distinguishing disinformation from 'poor' or 'bad' journalism, i.e., flaws in journalistic practice, (political) satire, parody or native advertising (Egelhofer & Lecheler, 2019, p. 101). Additionally, disinformation can be distinguished from illegal content, such as hate speech or racism, although it should be noted that this distinction cannot be held as clearly as desired in all cases (Martens et al., 2018, p. 10).

From a societal perspective, understanding disinformation as individual, deliberately disseminated falsehoods falls short (Kohring & Zimmermann, 2019, p. 19). Some authors rather identify it as part of a systematical attempt to destabilize democratic institutions and processes (Bennett & Livingston, 2018; Kohring & Zimmermann, 2019). Disinformation is typically intentionally created in order to affect citizens and make an impact on public opinion and society and therefore clings to existing identities and ideologies (Moeller & Hameleers, 2019, p. 7). The result is narratives, i.e., emotionally and ideologically charged stories comprised of misleading information that embodies a certain worldview and attempts to create an alternative reality. Society is thus confronted with a comprehensive *information disorder* (Bennett & Livingston, 2018; Kohring & Zimmermann, 2019; Wardle & Derakshan, 2018). This *information disorder* is said to having emerged from the plethora of communication channels available online, causing a disrupted public sphere in which disinformation narratives are able to circulate freely—not least because journalistic gatekeepers have become less important as an institution of quality control on the Internet (cf. chapter 2; Napoli, 2019)—and a loss of trust in democratic institutions such as the media and politics of parts of Western populations (Bennett & Livingston, 2018; Egelhofer & Lecheler, 2019; Zimmermann & Kohring, 2020). Therefore, the following sections focus on the broader concept of disinformation, starting with an exploration of its negative effects on democracy and public discourse from a theoretical perspective.

Harmful Effects of Disinformation on Democracy and Public Discourse

The societal debate about the negative effects of disinformation rests on the assumption that it poses a threat to the collective (political) self-determination of society (Zimmermann & Kohring, 2018, p. 526). Potentially harmful effects can therefore be distinguished at the individual micro-level and the societal macro-level. At the individual level disinformation undermines legitimate processes of opinion and will formation (Zimmermann & Kohring, 2018, p. 526) by threatening to violate basic political rights, namely the free and self-determined formation of opinion and its passive counterpart, the freedom of information (Bayer et al., 2019, pp. 76-77; Jaursch, 2019, p. 8). While freedom of opinion ensures that citizens can form their own free and informed opinions based on all available information, intact freedom of information is necessary for the formation of opinion to be realized. If freedom of information is hindered and citizens therefore lack essential, trustworthy information, the formation of opinion and thus their political decision-making is hindered. A concrete example of such a distorted formation of opinion and will is the assumption that disinformation entices individuals to

make concrete political decisions such as voting on the basis of false information (Jungherr, 2019, p. 28).

At a societal level, the violation of the above-mentioned political rights threatens democratic discourse (Jaursch, 2019, p. 8) with harmful consequences for democracy as a whole. Democracies require a fruitful public discourse characterized by a diversity of trustworthy and therefore correct information to function properly. If the prevalence of disinformation reaches a level that distorts the public discourse by essentially replacing and suppressing truthful information, the foundation of democracy becomes unstable (Bayer et al., 2019, pp. 77–78; Jaursch, 2019, p. 8). In other words, some authors fear that disinformation fundamentally questions factuality as the foundation of the democratic processes (Lewandowsky, Ecker, & Cook, 2017; Zimmermann & Kohring, 2018). Lewandowsky et al. speak of a *post-truth era*, "in which a large share of the populace is living in an epistemic space that has abandoned conventional criteria of evidence, internal consistency, and fact-seeking" (Lewandowsky et al., 2017, p. 360).

In addition to these indirect effects, disinformation may also have direct negative effects on democracy: disinformation campaigns, for example, may threaten the integrity of elections (EC, 2018a, p. 5), incite polarization on conversely debated issues (e.g., migration) and undermine trust in democratic processes, and the "common meeting ground" of shared facts, issues, and values (Jaursch, 2019, p. 8). In this context, it is worthwhile mentioning that disinformation might also undermine the integrity and trustworthiness of national media systems (EC, 2018a, p. 11), which ideally provide the necessary "common meeting ground" (Katz, 1996).

The question now arises as to what extent these fears are actually justified. The following sections therefore focus on the creation, dissemination and scope of disinformation in order to assess their impact on society.

5.2.2. Empirical Findings: Creation, Dissemination, Scope and Effects

Producers of Disinformation

Academic research about the creation and dissemination of disinformation is still in its early stages (Müller & Denner, 2019, p. 8). In addition, the available literature on the producers, which mostly consists of journalistic reports (Bayer et al., 2019, p. 36), is highly atomized so that the true scope of the role of different actors is impossible to evaluate adequately (Tucker et al., 2018, p. 29). It is, however, certain that there are targeted disinformation campaigns (Wardle & Derakshan, 2018; for detailed information about disinformation campaigns targeting EU member states see Bayer et al., 2019, pp. 36–50; cf. Tucker et al., 2018) and that foreign governments such as Russia, Iran and China play a part in the dissemination of disinformation (Tucker et al., 2018; Wardle & Derakshan, 2018).

A study commissioned by the European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs for the EU points out that a precise estimation of the proportion of disinformation targeting EU members states that originates from foreign or domestic actors cannot be given (Bayer et al., 2019, pp. 28–29). However, it seems that Russia plays an important role as one of the most active originators and disseminators of disinformation, since Russian activities have been particularly prominent in recent democratic processes in the EU and overseas (see Bayer et al., 2019 for details).

There is also evidence that in several countries the far right plays a decisive role in disseminating disinformation online: for example, in the US a large share of disinformation derives from alt-right, far-right and hyper-partisan websites (Bennett & Livingston, 2018; Tucker et al., 2018, pp. 26-27). This trend is, albeit on a much weaker scale, mirrored with regard to the right-wing populist party AfD in Germany (Hegelich & Thieltges, 2019). In general, it can be stated that

disinformation is neither limited in time nor in place, yet the probability of an increase is highest before significant democratic events (elections, referenda). Moreover, the virulence is especially at its strongest during high-profile events and/or crises (Bayer et al., 2019, pp. 29–30).

Spreading of Disinformation Through Intermediaries

The dissemination is often assumed to be strengthened by intermediaries, with social media suspected of being particularly suitable for this purpose. Online, disinformation can be published without quality-control or fact-checks and may theoretically reach an unlimited audience (Jungherr, 2019, p. 24), since the traditional vertical integration of the public sphere is broken open: in the offline era, both the production and distribution of news were in the hands of traditional mass media, a fact that implied that news items could only be made available to a wider public through journalistic gatekeepers (Napoli, 2019). In the Internet age, bypassing the classic gatekeepers is easier than ever. While this is true for all kinds of content, several characteristics make social media particularly vulnerable to disinformation.

Social media base their business model on ad revenue and, therefore, work according to an economically oriented, attention-driven logic (see also chapter 2.1; Jaursch, 2019, pp. 9–11; Jungherr, 2019, p. 24-25; Tucker et al., 2018, p. 37): at the core of most social networks is microtargeting, defined as "the data-based, targeted addressing of individual people or groups" (Jaursch, 2019, p. 10). The primary revenue source of social media platforms is selling personalized ads whereby the user's attention acts as the sole driving force behind this system. In other words, the main objective of platforms is to maximize the attention of individual users—so that they spend as much time as possible on the respective platform—to display and thus sell as much advertising as possible. In this context, strong negative emotions, such as anger or resentment, appear to be most effective at binding the user's attention (Wollebæk, Karlsen, Steen-Johnsen, & Enjolras, 2019). In the context of disinformation political microtargeting is of special interest: Political direct marketing, i.e., personalized messages to individual voters using predictive modeling techniques and political behavioral advertising, i.e., using collected information about users' online behavior to display individually targeted ads (Zuiderveen Borgesius et al., 2018, p. 83) can both be used to strategically disseminate disinformation to particularly susceptible groups of users (Jungherr, 2019, p. 25).

In addition, their optimization algorithms also make social media vulnerable to disinformation (Tucker et al., 2018, p. 37-38). As has been pointed out in chapter 2.1 social media algorithms favor content that is predicted to create high levels of engagement in the form of likes, comments and shares. Thus, disinformation that is packaged in emotional stories with sensational headlines has a higher chance of diffusing through social media. This is due to the emergence of a new communication style on social media that favors short, simplistic and emotional messages (Bayer et al., 2019, p. 56; see also chapter 5.1). It seems that disinformation around controversial topics, such as migration, children and abuse, war and peace, or conspiracy theories with easily accessible statements is able to fulfill these criteria and thus spread widely and quickly through social media (Müller & Denner, 2019, p. 8). Furthermore, once it reaches a certain popularity threshold it can spread exponentially quicker since popularity cues are one important reason why users engage with messages (Moeller & Hameleers, 2019, pp. 10-11). Actors using social bots or human troll factories also take advantage of the fact that posts receiving higher levels of interaction, in general, are ranked as more relevant, and therefore more likely to be shown to a larger audience. By using fake accounts that automatically like and share posts ("click farms"), social bots and human trolls are used to manipulate this mechanism by generating artificial reach (Müller & Denner, 2019, p. 9; Wardle & Derakshan, 2018, p. 49).

Thus, the interactive and participative component of social media is of paramount importance for the dissemination of disinformation. In this context, a connection to the above-mentioned debates about filter bubbles and echo chambers can be established: As pointed out in chapter 2.1, the individual user's network plays a decisive role in what content is visible to him or her (DeVito, 2017). This means that disinformation spread fastest in homogeneous networks with like-minded or identical world views (del Vicario et al., 2016; Guess, Nagler, & Tucker, 2019; Zollo, 2019), showing that interactions by individual users are a key factor for the fast and wide spread of disinformation (Müller & Denner, 2019, p. 8). In this context, algorithmically driven selective exposure may consolidate and strengthen an already existing inclination toward disinformation (Wardle & Derakshan, 2018, p. 51).

Evaluation of the Scope and Effects of Disinformation

As mentioned previously, the reality of the threats described in the previous paragraphs is dependent on both the scope of disinformation, and the level of interaction with it by users. Compared to the extensive theoretical work on defining disinformation and mapping out potential threats to society, (peer-reviewed) empirical studies are surprisingly scarce. Overall, there is little reliable information on the extent and consequences of disinformation, so that further research is needed to better assess the extent of the issue.

Scope of Disinformation

A conclusive assessment of the extent of disinformation, for example as a proportion of the total information available online, is impossible since the total amount of the ever-growing information online simply cannot be assessed empirically (Müller & Denner, 2019, p. 10). Most of the few studies attempting to measure the scope of disinformation stem from the US and show that disinformation only reaches very small parts of the population, so that a widespread impact on society is improbable (Allcott & Gentzkow, 2017; Guess et al., 2018; Grinberg et al., 2019; Nelson & Taneja, 2019). For example, the results of Grinberg et al (2019) and Guess et al. (2018) point towards a concentrated spread of disinformation: in the 2016 US presidential election only 1% of Twitter users generated 80% of exposure to sources disseminating disinformation and only 0,1% generated 80% of shares of such sources. This tightly packed group of users mostly consisted of Republicans, who were older than the average user and highly politically interested (Grinberg et al., 2019). Guess and colleagues (2018) largely confirm this notion by showing that disinformation websites comprised only 6% of the overall news diet of Americans during the election and that 20% of users with the most conservative news repertoire generated almost two-thirds of the traffic to disinformation websites. Overall, disinformation is mostly disseminated through social media and in particular Facebook (Allcott & Gentzkow, 2017; Guess et al., 2018) although the amount of disinformation on Facebook has gone down since the 2016 election (Allcott, Gentzkow, & Yu, 2018; Guess et al., 2018).

The data situation for Europe is even thinner: from the perspective of users, disinformation seems a relevant factor in their online news encounters. Surveys suggest that users encounter (Eurobarometer, 2018) and share (Chadwick & Vaccari, 2019) false information on a regular basis: In an EU-wide survey in February 2018, for example, 37% of respondents stated that they come into contact with false information on a daily basis, and almost a further third (31%) that they do so at least once a week. The perceived extent of the spread seems to be more or less constant across all member states, as in each country at least half of the respondents reported being exposed to false information at least weekly (Eurobarometer, 2018). However, it is premature to draw conclusions about the real extent of disinformation in Europe from these figures: with the notoriously flawed self-report methodology used in these type of studies (e.g., Prior, 2009) "it is likely that the actual frequency

of false information is overstated" (Müller & Denner, 2019, p. 10).

A further problem concerns the extent to which respondents are able to correctly assess the credibility of information. The Eurobarometer-survey, for example, asked how often respondents came across news or information they "believed to misrepresent reality or is even false" (Eurobarometer, 2018: Q2). The responsibility to assess whether a news item is correct or whether it is to be considered disinformation was thus completely transferred to the respondents. From a "normal" citizen's point of view, estimating the credibility of sources and their content is a difficult task at best—it is not without reason that one core task of professional journalists is to provide and classify information about events that ordinary citizens are not able to obtain themselves (Coleman, Anthony, & Morrison, 2009, p. 4). The difficulty of distinguishing trustworthy sources and information from disinformation is further increased online, as there is an unmanageable number of available sources of information, which, in many cases, can also show a professional journalistic presentation, regardless of their actual credibility and trustworthiness (Martens et al., 2018, p. 43).

Tracking data—obtained by Fletcher and colleagues (Fletcher et al., 2018) in France and Italy—are therefore able to paint a more realistic picture. It seems that there is a very small core of users who are regularly exposed to disinformation (1–3%), while the overwhelming majority does not encounter disinformation at all. The reach of the selected disinformation websites is thus ten to a hundred times lower in comparison to established news sites but few exceptional disinformation pages are able to generate higher interactions on Facebook than its mainstream counterparts. These results are in line with findings suggesting that disinformation spreads faster on social media than true information (Vosoughi, Roy, & Aral, 2018).

Effects of Disinformation

Effects at the individual level are connected to two questions: how probable is it that disinformation reaches its target audience? And are the persuasive effects of disinformation feasible (Jungherr, 2019, p. 28)? The few available studies regarding the first question show that users of disinformation most likely already share the worldviews presented in disinformation (Guess et al., 2018; Grinberg et al., 2019; Hameleers, Powell, van der Meer & Bos, 2020; Zollo, 2019). When confronted with disinformation it is accepted as credible and true (e.g. Arendt, Heim & Beck, 2019; Hameleers et al., 2020) and may influence people's beliefs and (subsequently) voting-decisions (e.g., Li, 2020; Zimmermann & Kohring, 2020). Most effects point toward a confirmation bias, i.e., disinformation taps into pre-existing attitudes that are confirmed and (moderately) strengthened. Studies show that people with right-wing or conservative attitudes are more likely to believe in disinformation (Arendt et al., 2019; Swire, Berinsky, Lewandowsky, & Ecker, 2017) and can be persuaded to change their voting intention towards right-wing parties and/or politicians (Barrera, Guriev, Henry, & Zhuravskaya, 2018; Zimmermann & Kohring, 2020). At the societal level, a study by Huang (2017) shows that disinformation may reduce trust in political institutions indicating that the actual effects of disinformation may rather lie in eroding political support than in changing beliefs about specific facts.

5.2.3. Concluding Remarks: How to Combat Negative Effects

Disinformation has been identified as a veritable and tangible threat to democratic societies in recent years (Bennett & Livingston, 2018; Lewandowsky, Ecker, & Cook, 2017; Wardle & Derakshan, 2018). Despite the huge attention the issue has received in public, political and scholarly debate, empirical evidence on the actual scope and impact of disinformation on societies is scarce. The very limited available data suggests

that fears of a large scope of disinformation seem exaggerated: exposure to disinformation seems to be confined to a small, politically right-wing sub-population (e.g., Guess et al., 2018; Zimmermann & Kohring, 2020). This group appears to be susceptible to being influenced by disinformation (Arendt et al., 2019), likely because disinformation narratives dock at their preexisting attitudes (confirmation bias). Persuasive effects, however, have not been confirmed, so that disinformation is rather able to widen existing gaps between users with opposing worldviews (Jungherr, 2019; Moeller & Hameleers, 2019).

Authors like Jungherr (2019, p. 28) are therefore puzzled about crisis scenarios painted in conceptual articles (Bennett & Livingston, 2018) and reports commissioned by political institutions (e.g., Wardle & Derakshan, 2018) warning of the widespread effects of disinformation but lacking actual empirical evidence. This is, however the sticking point: there is a distinct lack of empirical knowledge regarding the scope and interaction with disinformation, especially in Europe, which impairs a serious evaluation of the threat disinformation poses for democratic discourse and societies. Accordingly, most researchers agree that more research on the scope, content, exposure and effects of disinformation is needed to seriously evaluate the impact of disinformation on society (Jungherr, 2019; Li, 2020; Moeller & Hameleers, 2019; Müller & Denner, 2019; Tucker et al., 2018). In this context, an important step would be to open up the black box of algorithms for researchers and to secure the access to data so that existing research gaps can be closed—so that evidence-based legal decisions regarding the handling of disinformation can be made (Jaursch, 2019, p. 26; Müller & Denner, 2019, p. 24; cf. chapter 7).

Summarizing, the above-mentioned evidence suggests that in online environments, the groups most vulnerable to negative effects of disinformation, are like-minded users who share it amongst themselves. Political microtargeting on social media, in particular, can be used to manipulate these vulnerable groups, which can trigger or reinforce processes of social division. The latter aspect might be most relevant to the debate: several authors have pointed to the fact that the debate about fake news is misjudged, and that actual manipulation is not so much about right or wrong information, but about weakening, dispersing, and polarizing society, whereby the manipulation of information is only one strategy that is applied (Bayer et al., 2019; Jaursch, 2019; Kohring & Zimmermann, 2019; Moeller & Hameleers, 2019). Behind the phenomenon of disinformation in the digital space lie questions about the data power of some corporations, the attention-driven system of many social networks, weakened journalistic gatekeepers, and the lack of rules for online political communication. Therefore, searching for solutions must cover several legal and political areas.

Measures to Combat Disinformation

Most authors recommend a holistic strategy to combat possible negative effects of disinformation (Lewandowsky et al., 2017; Li, 2020; Moeller & Hameleers, 2019). Thereby, directly countering narratives or single disinformation storylines through *debunking* is only one facet. While a recent meta-study has shown that debunking may correct belief in false information, political disinformation is harder to correct than other topics (Walter & Murphy, 2018). Research on the effects of correcting political disinformation, therefore, show mixed results (e.g., Hameleers & van der Meer, 2020) and debunking comes with several additional drawbacks: target audiences are often not reached by corrective information (Guess et al., 2018; Zollo, 2019) and even if they are, believers in disinformation may reject fact-checkers because they distrust them or perceive their information as attacks on their beliefs—which may cause reactance and strengthen belief in false content (Moeller & Hameleers, 2019, p. 13). Additionally, even if fact-checking succeeds in correcting people's false beliefs, their positive attitudes towards, and voting-intentions for politicians spreading disinformation

remain (Swire et al., 2017; Thorson, 2016). Finally, debunking cannot verify every piece of disinformation online and is most often applied after the original disinformation has already been published, so that the damage may have already been done (Moellers & Hameleers, 2019, p. 13).

Due to these drawbacks, authors recommend a shift from purely reactive to proactive measures, aiming at strengthening users in digital communication environments (Jaursch, 2019; Lewandowsky et al., 2017; Moeller & Hameleers, 2019). This shift is also advisable from a communication-ethical and legal perspective: since the free exchange of ideas and information is fundamental in democratic societies, all forms of censorship have to be avoided when countering disinformation. From the perspective of communication science, this is the reason why concrete legal and legislative regulatory measures, such as content deletion (e.g., Germany's NetzDG), can be criticized. Apart from the legal problems that such measures may entail (e.g., overblocking, potential restriction of free speech; Müller & Denner, 2019, p. 20), the mandatory deletion of suspected disinformation can be criticized since it is plausible to assume that those whose worldviews coincide with disinformation content will, when it is deleted on a large scale by the dominant platforms, look for alternative ways to obtain such information and shift their media use to alternative niche offers (Müller & Denner, 2019, p. 23)—a potential development that may facilitate fragmentation and polarization.

Rather than focusing on banning disinformation or even developing tools to change people's worldviews, measurements should therefore focus on positive goals of securing the integrity of information (Moeller & Hameleers, 2019, p. 14). This includes ensuring access to high quality information across all strata and strengthening users on social media. In particular, it is recommended to educate users in information and media literacy. This can stimulate more critical skills that enable them to challenge their own biases and may nudge them to argue the trustworthiness of sources and messages (Moeller & Hameleers, 2019, p. 13-14).

Main Findings:

- Disinformation poses a threat to democratic societies since it inhibits the free formation of opinion and political will.
- Due to their attention-driven logics, social media are particularly vulnerable to being misused as a conductor for the fast and wide-spread dissemination of disinformation.
- Scientific data on the scope of disinformation are scarce. The available data suggests that the reach of disinformation is very small compared to established media sources.
- Furthermore, actual exposure to disinformation is confined to small groups of (right-wing) users, who are, however, susceptible to being influenced by it.
- Disinformation may widen existing gaps between users with different worldviews. Persuasive effects cannot be confirmed.
- More research is paramount to better assess the extent of the issue.

5.3. Incivility and Hate Speech

The following chapter examines incivility as manifestations of an alarming online discourse culture. Readers will find information on the distinction between incivility and hate speech and its scope and dissemination through intermediaries. Studies will show that incivility and hate speech have negative consequences on public discourse. Concluding, countermeasures to combat incivility and hate speech are presented, bearing in mind that measures restricting freedom of expression may only be applied against illegal hate speech content.

5.3.1. Defining Incivility and Hate Speech

Incivility in public debates has received increasing attention in the academic and public debate in recent years. An especially extreme aspect of the radicalization of public discourse is the spread of hate speech in online communities. Similar to the phenomenon of disinformation, there are no universally accepted scientific definitions of incivility and hate speech (Coe, Kenski, & Rains, 2014; Erjavec & Kovačič, 2012; Kümpel & Rieger, 2019; Siapera, Moreo, & Zhou, 2018). In both cases, the current use of the term applies to rather heterogeneous acts of speech ranging from offensive and/or derogatory speech to unlawful criminal acts (Siapera et al., 2018).

Following Kümpel & Rieger (2019, p. 9), we understand incivility as a form of communication that transgresses norms of interpersonal and/or deliberative communication and can be observed in (partially) public discussions online (e.g., on social media, commentary sections on websites, video platforms, and online communities). Here a clear distinction is drawn between incivility, which can have negative consequences for democracy, and ordinary negativity as part of a functioning deliberative discourse aiming at criticizing political ideas and policies, questioning the qualifications of political opponents, or expressing disagreement with ideas or behaviors of group representatives (Su et al., 2018, p. 3680).

Following the understanding of Kümpel and Rieger (2019, pp. 9–10), hate speech is identified as one of several sub-phenomena of incivility, such as "trolling" and "flaming" (use of aggressive and hostile language against individuals; Kümpel & Rieger, 2019, p. 10).[6] There is an ongoing debate about the concept and its legal and ethical implications. Consequently, broader and narrower definitions of hate speech can be identified, from which some common elements emerge:

> *"Hate speech refers to an expression that is abusive, insulting, intimidating, harassing, and/or incites to violence, hatred, or discrimination. It is directed against people on the basis of their race, ethnic origin, religion, gender, age, physical condition, disability, sexual orientation, political conviction, and so forth" (Erjavec & Kovačič, 2012, p. 900).*

Broader definitions thus include any expression of disdain and animosity toward groups and individuals and any stereotypical disparaging remarks (Siapera et al., 2018, p. 12). At the European level, illegal or unlawful hate speech is defined as "all conduct publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin" (EC, 2016). Much like the Council of Europe's Committee of Ministers' Recommendation 97(20) (Council of Europe, 1997), the definition distinguishes between expressions that are, albeit offensive and insulting, fully protected by the right to freedom of expression, and acts of speech

6 Note that establishing clear-cut distinctions between the different sub-phenomena of incivility is very difficult, which is why some authors demand not to focus on searching for precise definitions but rather examine overarching problems associated with incivility on social media (Kümpel & Rieger, 2019, p.10). This approach is followed in this report, whereby the special legal status of hate speech is taken into account with regard to concrete recommendations.

that do not fall under this protection. In general, definitions that define unlawful hate speech emphasize the element of intention or incitement.

Hate speech is therefore classified as unlawful if it is directly and immediately potentially harmful to individuals or the social order. Consequently, there are great differences, often in accordance with the respective political and constitutional traditions or local language cultures, as to which forms of language are still permitted and which are prohibited. In the US, for example, on the basis of the First Amendment, disparaging or even racist outbursts may be covered by freedom of speech, as long as they do not directly or explicitly threaten with violence ('fighting words'). On the contrary, in most European countries, comparable content is prohibited solely on the basis of its offensive content, regardless of whether it contains threats of violence. However, despite attempts at harmonization by the EU, there are still differences between countries, as to whether or not, for example, denying the Holocaust, trivializing fascism, or blasphemy are considered hate speech (Siapera et al., 2018, pp. 12–13).

5.3.2. Empirical Findings: Scope, Dissemination and Effects on Public Discourse

Scope of Incivility and Hate Speech and Its Dissemination Through Intermediaries

In 2016 the European Commission implemented a code of conduct on countering illegal hate speech online, which obliges platform operators to create structures that allow notifications of hate speech to be reviewed as quickly as possible (by the majority within 24 hours) in accordance with European and/or national legislation and deleted, if necessary (EC, 2016). The latest round of evaluation, from February 2019, indicates that exposure to hate speech seems to remain a prevalent phenomenon on social media (Jourová, 2019), while Johnson et al. (2019) are able to identify an "online hate ecology" consisting of interconnected, extremely right-wing hate clusters spreading across the globe. What are the reasons for this? In their report for the Konrad-Adenauer Foundation, German communication scientists Anna Sophie Kümpel and Diana Rieger (2019, pp. 13–16 for the detailed analysis) address this question in-depth.

Kümpel and Rieger (2019, pp. 13–14) first focus on reasons that many observers *perceive* a negative change in language and debate culture and a rise in incivility as well as hate speech on social media. A major factor is the increased visibility and public accessibility of debates and discourses online (e.g., Coe et al., 2014, p. 658). This is especially prevalent on social media, where users have the opportunity to become active participants in the discussion. Studies suggest that negative and controversial news, disruptive events such as terrorist attacks, and acute social changes in particular increase interest and motivation to participate in online discourses (e.g., Kümpel & Rieger, 2019, p. 13; Müller & Denner, 2019, p. 8; Oksanen et al., 2018; Ziegele, Breiner, & Quiring, 2014). In the context of incivility, affective/emotional responses and strong emotions, such as anger, seem to play a significant role that can trigger an increase in hate speech.

In the wake of Islamist-motivated terrorist attacks, for example, Muslim communities often serve as scapegoats and have to expect increased hate speech (Kümpel & Rieger, 2019, p. 13). Closely related to this are the same dissemination mechanisms already discussed in the section on disinformation. Highly emotional content is more likely to be disseminated and interacted with on social media. That, in turn, increases the visibility of heated, uncivil debates since the algorithms of social media platforms tend to place content that is highly engaged in the form of likes, comments and shares more prominently in the feeds of other users (DeVito, 2017; Kümpel & Rieger, 2019, p. 14). At this point, a dual role of the platform providers can be critically noted: on the one

hand, they take action to prohibit incivility and delete illegal hate speech, while on the other hand they provide the ideal infrastructure for its distribution (e.g., Ben-David & Matamoros-Fernandéz, 2016; Kümpel & Rieger, 2019, p. 14).

In a second step, Kümpel and Rieger (2019, pp. 14–15) identify changing communication conditions and (group) norms that may force negative communication behaviors. A development which, in turn, suggests that there is indeed an increase in online incivility. In particular, the role of anonymity has long been discussed, which can contribute to disinhibition and thus an increased inclination to incivility (e.g., Brown, 2018). Moreover, social psychologists assume that anonymity in the sense of deindividuation can raise awareness of belonging to social groups and thus increase orientation toward (perceived) group norms (e.g., Reicher, Spears, & Postmes, 1995). A negative spillover effect can, therefore, be assumed for uncivil online discourses, whereby new participants adapt to the disparaging tone of the debate (e.g., Chen & Lu, 2017; Hsueh, Yogeeswaran, & Malinen, 2015). However, the same applies conversely to civil discourses, in which positive adjustment can be expected. However, empirical findings regarding the relation between anonymity and incivility generally show mixed results (e.g., Rösner & Krämer, 2016; Rowe, 2015).

A final important cause for the change in debate culture is motivation and personality traits (Kümpel & Rieger, 2019, pp. 15–16). Individual motivations, like the reduction of pent-up emotions such as anger in relation to a controversial and/or emotionally charged issue, or the heated reaction to an attack on one's own political view or social identity can result in rather unintentional incivility. However, racist incitement in the form of hate speech, in particular, is also deliberately used as a means of distinction from outgroups and the strengthening of the cohesion of the respective in-group (e.g., Glaser, Dixit, & Green, 2002; Hutchens, Cicchirillo, & Hmielowski, 2015). In the context of trolling and hate speech, fun and enjoyment are also popular motives for engaging in uncivil behavior (e.g., Erjavec & Kovačič, 2012). In addition, personality traits such as an inclination to verbal aggression (e.g., Cicchirillo, Hmielowski, & Hutchens, 2015) or sadism (e.g., Craker & March, 2016) also facilitate uncivil commenting.

In summary, it can therefore be stated that changed communication conditions, increased visibility and easier participation in discourses, and the technical possibilities for the further dissemination of content on social media, as well as the actors themselves with their personality traits and motives, have led to an increased perceptibility of incivility in online discourses (Kümpel & Rieger, 2019, p. 16). There are, at the same time, first empirical indications that incivility has actually increased (Su et al., 2018; Kümpel & Rieger, 2019, pp. 3–4 for a discussion of German findings).

Effects on Public Discourse

The negative impact of the proliferation of online incivility and hate speech is twofold; direct and indirect effects can be distinguished. A major direct effect has already been previously mentioned: Incivility and online hate speech negatively affect the communication behavior of discussion participants whereby the individual exposure to uncivil comments increases the likelihood to engage in incivility (Chen & Lu, 2017) and new participants joining an uncivil discourse are more likely to use incivility themselves (Hsueh et al., 2015; Kümpel & Rieger, 2019, p. 22). The result is a spiral of incivility, radicalizing the discourse culture online. Another worrying effect is that female victims of incivility, in particular, react by withdrawing from online discourses, which can cause problems on a societal level (Sobieraj, 2018; Kümpel & Rieger, 2019, p. 23). A qualitative study with German Facebook users confirms these results and furthermore shows that discussions on Facebook are generally perceived as uncivil, disrespectful, and characterized by a low level of discourse (Stark et al., 2017). This, and the fear

of being targeted with hate speech, which is particularly pronounced by women who, in extreme cases, have to fear is a reason to not only withdraw from but even not join in discussions on Facebook at all.

In addition, experimental studies have shown indirect negative effects of the presence of uncivil user comments on online news articles (Kümpel & Rieger, 2019, p. 21): they lower the perceived quality of articles in lesser known brands (e.g., Prochazka, Weber, & Schweiger, 2018), make reporting on political actors and neutral blog posts appear more distorted (e.g., Anderson et al., 2018), reduce credibility and the assessment of the importance or relevance of topics addressed in articles (e.g., Waddell, 2018), and may polarize opinions, perceptions, and viewpoints (e.g., Anderson, Brossard, Scheufele, Xenos, & Ladwig, 2014; Kümpel & Rieger, 2019, p. 21). This may be exacerbated by the tendency of users with moderate opinions to abstain from discussions (Stark et al., 2017), so that online those with more radical opinions are over-represented. These findings can, in the long term, lead to a loss of trust in the media and media disenchantment, which further endangers public discourse (Kümpel & Rieger, 2019, p. 21), since it undermines trust in the institution that ideally provides the "common meeting ground" that is necessary for the functioning of democracies (Katz, 1996). In this context, the frequent incivility directed against journalists can inhibit public discourse further, as journalists may be intimidated by incivility and in particular hate speech directed towards them and subsequently avoid reporting on controversial topics or try to report on topics in a less controversial manner (Binns, 2017; Nilsson & Örnebring, 2016; Post & Kepplinger, 2019; Kümpel & Rieger, 2019, p. 22–23).

5.3.3. Countermeasures against Incivility and Hate Speech

The question of how to appropriately and effectively counter incivility and hate speech, is controversial: repressive and preventive strategies can be distinguished (Kümpel & Rieger, 2019, pp. 25–28). Repressive measures, such as those anchored in the German NetzDG (Network Enforcement Act), include the deletion of illegal content such as hate speech, or the blocking of individual users, while preventive measures focus on community management and moderation of discourses. Different types of measures seem appropriate for different kinds of incivility which is mostly tied to the specific legal status of hate speech.

Repressive Strategies

Repressive measures must be applied with great caution. The OHCHR Rabat Plan of Action[7] recommends a high threshold for violations of the freedom of expression in the context of hate speech. It is therefore clear that the deletion of content or the blocking of individual users can only be used for illegal hate speech (cf. Cornils, 2020). But, from a legal perspective, there are concerns that the NetzDG and similar international laws could restrict freedom of expression through the obligation of preventive deletion, or that fundamental governmental tasks such as law enforcement are transferred to private companies and their self-regulation (e.g., Eickelmann, Grashöfer, & Westermann, 2017, pp. 182–183; Müller & Denner, 2019, p. 22; Siapera et al., 2018, pp. 14–17). From the point of view of communication science, overblocking could lead to a restriction of public discourse just as much as illegal hate speech itself. Finally, it should be noted that the perpetrators of racist hate speech, in particular, have learned to encode their language in such a way that they neither violate community guidelines nor produce criminally relevant content (Kümpel & Rieger, 2019, p. 26; Siapera et al., 2018, pp. 27–29).

7 https://www.ohchr.org/EN/Issues/FreedomReligion/Pages/RabatPlanOfAction.aspx

At this point, repressive measures aimed at deleting potentially criminal content prove futile and become ineffective. Despite these drawbacks, the detection and deletion of illegal hate speech should remain a cautiously applied tool among others.[8]

Preventive Strategies

Preventive strategies not only focus on content moderation, but also on softer forms that aim at establishing and upholding a polite and respectful discourse climate. These forms of community management are able to curb incivility—but only if it is objective and constructive. Sarcasm, or satirical replies, are not helpful in reducing an uncivil discourse culture (Ziegele & Jost, 2016; see also Kümpel & Rieger, 2019, pp. 25-26). However, these forms of community management have to be applied carefully, since uncivil but not illegal content is not subject to legal regulation (cf. Cornils, 2020). Forms of interactive interaction, i.e., public responses by journalists to the comments of their readers, that focus on sociable replies, aiming at creating an informal and pleasant discussion atmosphere, appear to be able to reduce incivility particularly well (Ziegele, Jost, Bormann & Heinbach, 2018).

8 A detailed legal analysis on this matter is provided by Cornils (2020).

Main Findings:

- Incivility and hate speech are relevant areas of concern in online discourses.
- Especially worrisome are studies showing that the credibility of journalistic content is reduced by incivility, which can have detrimental effects on trust in journalism as an institution of social cohesion in the long term.
- A related problem is that especially women, who have been victims of hate speech, and also users with moderate opinions, stop participating, or do not join in discussions. This effectively hinders the free development of public discourse on the macro-level, since whole groups of the population opt-out.
- Constructive, sociable replies by journalists on comments can contain and prevent incivility.
- Measures inhibiting freedom of expression, such as content deletion, have to be carefully applied and can only be used to combat illegal content such as hate speech.

6. Opinion Formation in the Digital Age

The previous chapters have clearly shown that platforms are able to influence public discourse. They can alter the way that users form and express opinions on topics of public interest. What does this mean for individual and public opinion-formation processes? In what way does the perceived climate of opinion influence online opinion expression? To what extent can we observe a polarization and narrowing of opinion formation? What is the role of algorithms in this context? These and related questions will be discussed in this chapter, in which we will elaborate on the implications of the findings regarding the identified concerns on opinion-formation. In order to assess the risks, it is important to understand how opinions are formed and influenced in the current digital age.

The formation of public opinion has been conceptualized as a multilayered process involving not only interpersonal communication among individuals in closed spaces, but also mass communication over media outlets with a huge audience. In general, intermediaries may function similarly to traditional media in terms of their agenda-setting function and knowledge mediation roles regarding both information and attitudes. They can also provide an impression of the climate of opinion as well as how attitudes and sentiments are spread among the population. Therefore, potential effects can be assigned to the different levels of the individual opinion formation process: agenda setting, as well as the processes of the dissemination and acquisition of knowledge and attitudes (Stark et al., 2017).

Agenda Setting through Social Media: A Fragmented Audience?

We want to approach the question of the influence of intermediaries on the formation of opinion through the concerns addressed in the previous chapters. Our starting point is the initial question of whether, and to what extent, intermediaries influence the public agenda. This question is closely related to the debate about filter bubbles and consequently linked with the fragmentation thesis. As pointed out in chapter 4.2, the effects of the filter bubble seem to be significantly smaller than expected or even absent (e.g., Dubois & Blank, 2018; Haim et al., 2018; Krafft et al., 2018; Nechushtai & Lewis, 2019). There is news personalization but on a much smaller scale than assumed. In this context, the question therefore is no longer if filter bubbles exist but how "porous" they are (Zuiderveen Borgesius et al., 2016). Based on the current state of research, it seems that filter bubbles are generally permeable enough as to not form hermetically sealed, personalized information environments for every single user.

This conclusion has substantial implications for opinion formation. The emergence of filter bubbles is linked to the fear that they exclude central social issues and alternative perspectives, thus limiting the access of their 'inhabitants' to relevant information (Stark, 2013) and thereby negatively affecting agenda setting. Intermediaries do not seem to produce this undesired effect on their own: studies indicate that intermediaries like social media (Geiß et al., 2018) or

news aggregators (Nechushtai & Lewis, 2019) merely reproduce the agenda of traditional mass media.

A fragmentation of the agenda only takes place in the long-tail (Stark et al., 2017). However, several studies show that incidental exposure (Fletcher & Nielsen, 2018), network effects due to weak ties (e.g., Bechmann & Nielbo, 2018) and the algorithms themselves, which polarize less than expected (e.g., Moeller et al., 2018), can even contribute to an expansion of the diversity of information and opinions. These findings indicate that, under the right circumstances, intermediaries may have positive effects on opinion formation by expanding the user's individual thematic, informational, and attitudinal horizon.

Social Media and Political Polarization: Distorted Perception of the Climate of Opinion

However, in the context of attitude polarization induced by intermediaries, empirical findings show that the active selection of users interferes at this point. Hence, it is hard to evaluate the extent to which automated selection decisions contribute to a polarization of attitudes and opinions online. The studies regarding the echo chamber metaphor presented in chapter 4.2 indicate that several conditions have to be fulfilled under which algorithmic filtering can cause polarization. An important factor is predispositions and political attitudes: users with strong predispositions and attitudes are more likely to fall into the parallel discourses of echo chambers, which is especially true for those already at the edges of the political spectrum (e.g., Bright, 2018).

Another, highly relevant point is the perception of polarized opinions and the climate of opinion toward certain topics (Stark et al., 2017). The inherent logics of social media platforms favor highly emotional content that generates user engagement, shareworthiness and virality (De Vito, 2017; cf. chapter 2.1). Since controversially debated issues, in particular, generate user engagement, these issues are more likely to be highly ranked by algorithms and thereby more likely to be visible to a larger audience on social media. As Hagen and colleagues (Hagen, Wieland, & in der Au, 2017) show for migration, it is precisely these topics around which the polarizing effects of algorithmic news selection can be confirmed. This indicates that algorithms rather make extreme positions visible, but do not necessarily generate or reinforce them directly, since polarizing effects are not an automatic consequence of algorithmic news selection. The visibility and prevalence of polarization-inducing topics might be facilitated by media if they adapt to the logics of social media by softening the news (cf. chapter 5.1) and thereby providing exactly the kind of sensational, highly emotional news that incites controversial debates. How do these findings align with the argument that the climate of opinion is an important factor for the polarizing effects of algorithmic news selection by intermediaries?

A study by Stark et al. (2017) examined this question in closer detail. The increased accessibility of public debates on social media means, in the case of controversial topics, that polarized and thus more radical opinions and positions are more visible online than in the offline world. Following the theory of the spiral of silence (Noelle-Neumann, 2001), fear of isolation determines the willingness to articulate one's opinion publicly. The new possibilities to connect over the Internet, combined with the algorithmically induced increased visibility of controversial debates, makes the politically extreme—at all ends of the spectrum—aware of the fact that there are more likeminded people "out there" than they may have thought before. This may reduce their fear of isolation and increase their willingness to articulate themselves even outside of their echo chambers, in which the politically extreme are usually gathered (e.g., Bright, 2018; del Vicario et al., 2016), and thus lead to an overrepresentation of radical viewpoints and arguments in the political discourse.

On a societal level, this, in turn, may distort the public's picture of the actual climate of opinion, influence

the willingness of different opinion camps to speak out, and lead to spiraling processes because the perception of the strength of one's own opinion camp compared to those of other camps is overstated. If Facebook plays an important role as a source of information on a topic, users may feel that it is more likely that their opinion is held by the majority of users and may subsequently be more willing to articulate their views on a topic (Stark et al., 2017, p. 149). To summarize, intermediaries and, in particular, social networks, may indirectly contribute to polarization by facilitating a distorted picture of the climate of opinion.

In this way, social media (in particular) can reinforce affective polarization (sometimes also called "psychological polarization"; Settle, 2018). Pablo Barberá describes this, with reference to Settle, as follows: "She convincingly argues that core features of social media platforms, such as the fusion of social and political content, the ubiquity of social feedback, the ability to easily infer others users' political identity or the incentives to produce "clickbait-y" and inflammatory content to catch people's attention, have a direct impact on the aggregate level of psychological polarization" (Barberá, in press, p. 11). The characteristics of social media logics promote exactly the psychological processes behind it: the reinforcement of social and political identities. Combined with users' biased information processing, this can lead to very stereotypical and negative evaluations of the out-group.

Discussion: Opinion Formation

Due to the complexity of the relationships presented here, a final evaluation of the impact of intermediaries on processes of opinion formation is a difficult task at best. Research from the perspective of filter bubbles and fragmentation suggests that the overall impact of intermediaries on opinion formation seems rather limited. Filter bubbles are less of a problem than theoretically suggested, and intermediaries do not appear to fragment societal agendas. This assessment is strengthened by the finding that, despite the growing importance of online news sources, people usually integrate intermediaries in a relatively broad mix of news sources (cf. chapter 4.1).

However, research focusing on polarization through intermediaries provides a more nuanced picture. The current state of research makes the fear of the proliferation, and harmful influence of echo chambers, appear exaggerated, but there are conditions under which polarizing effects can be traced. Homogeneous networks, highly emotionalized and controversial topics, strong political predispositions, and, especially, a distorted perception of the climate of opinion, may facilitate polarization and the formation of echo chambers, which cause opinion formation as the foundation of consensus-finding to become problematic. Although some of these negative effects are mitigated by interpersonal communication about political issues (Stark et al., 2017), some experimental studies (e.g., Karlsen, Steen-Johnsen, Wollebæk, & Enjolras, 2017) suggest that exposure to, and discussion with, opposing opinions may reinforce polarized opinions—especially for those with already strong political attitudes.

Moreover, manipulation by disinformation can—although the limited available data suggests it to be of a rather limited scope—strengthen the problem of polarization. Microtargeting enables actors to strategically incite and escalate debates by precisely disseminating false information to susceptible groups and thereby not only influencing the agenda through specific agenda-setting effects but also distorting the public's climate of opinion. The visible incivility in many online debates is a clear indicator that polarized debates in comment sections, and on social media, are a veritable concern that threatens public discourse, and thereby negatively affects opinion formation.

Main Findings:

- With regard to agenda-setting effects, intermediaries function in a similar manner to traditional mass media. Studies indicate that intermediaries merely reproduce the agenda of traditional mass media. Hence, a fragmentation of the agenda only takes place in the long-tail.
- The current state of research permits the assumption that echo chambers may arise under certain circumstances; that is, facilitated by homogeneous networks, highly emotionalized and controversial topics, and strong political predispositions.
- Due to their core features, social media can directly reinforce affective polarization and may indirectly contribute to polarization by facilitating a distorted picture of the climate of opinion.
- Microtargeting enables actors to strategically incite and escalate debates by precisely disseminating false information to susceptible groups and thereby distorting the public's climate of opinion.
- The trend toward softening the news, with its focus on emotional topics and correspondingly controversial and uncivil discourses, can also trigger negative discourse spirals.
- Since algorithms are continuously being developed, it is possible that problems caused by news personalization could become more threatening in the future than is currently the case.

7. Concluding Remarks: Reconceptualizing the Threat

Platform Transparency Deficits

Information intermediaries are increasingly important actors in high-choice media environments. They change the structures and processes of how communication in digitalized societies proceeds—with potentially profound consequences for the functioning and stability of our democracies. In contrast to existing media organizations, intermediaries wield far broader power because they internalize markets: Users of intermediaries choose their sources within an environment whose logic is set by the platform itself. The open market—whether the selection of newspapers at a newsstand or the list of channels available on TV—is typically regulated to prevent anti-competitive behavior, and it guarantees some degree of transparency and a level playing field. Market participants can evaluate the behavior of competitors simply by "walking over and having a look". Internal markets for news, however, are opaque to individual and institutional observers. Within them, users of intermediaries are presented with a personalized pre-selection of content, but neither other users nor content producers can easily identify what those are. This implies two transparency deficits:

- (1) Individual users only see their own recommendations. They have no way of knowing what information was hidden from them, and
- (2) they cannot observe what information was presented to other users.

Outside actors (such as competitors, content suppliers, media authorities, and researchers) suffer from these limitations: They have no way of observing the treatment and behavior of individuals or groups of users. Information intermediaries, therefore, create new potential impacts (through personalization), along with detailed measurements thereof (creating what Webster, 2010 termed "user information regimes")—but hide both within a proprietary product. As intermediaries' importance to public opinion formation and political processes grows, societies will need to encourage effective transparency in order to safeguard a level playing field in information dissemination. As identified above, two elements are crucial for such transparency to work:

- (1) Individual users should be empowered with regard to the recommendations presented to them—they should be able to download everything that was presented, along with a sensible description of the processes that produced this exact set. Doing so would help individuals understand whether they were the subject of a biased selection and enable them to seek legal recourse in that case. Such a model is unproblematic from the perspective of user privacy, as users could only receive information which they can view anyway, yet it would facilitate the act of "whistleblowing" in the case of perceived wrongdoing.
- (2) An arguably more difficult goal would be the creation of some form of transparency that encompasses not individual users but the overall

impact of intermediaries. Such a measure is necessary for monitoring the societal influence that arises from the use of platforms, including endogenous algorithmic effects but also encompassing external factors, such as manipulation attempts from malevolent third parties (such as hackers and foreign government agencies). Transparency on this level could, for example, take the form of providing privileged parties (state offices, researchers or trusted NGOs) with accurate aggregate information on the demographic makeup of user populations, the prevalence of news usage and other key insights. It might also be worth considering making less detailed data publicly available, but such a decision would need to be weighed against the potential negative effect of facilitating targeted manipulation.

In general, transparency should help users make informed choices. But it is clear that "digital transparency—whether it is enacted through technological solutions or more classical administrative and organizational means—will not on its own provide an easy solution to the challenges posed by the growing role of platforms in political and public life" (Gorwa & Ash, 2020, p. 20). The ongoing debates on the implementation of the transparency rules in the German Media State Treaty clearly support this statement (Dogruel, Stark, Facciorusso, & Liesem, 2020).

Key Data Needs

The combination of novel impacts, and a lack of transparency, create three distinct classes of threat to societies. Protecting against each threat requires access to specific data—most crucially the individual personalized results presented to users, the actions taken by intermediaries to change the flow of information, and the overall impact on the whole user population—that is currently unavailable (Tucker et al., 2018).

- (1) As illustrated by existing anti-trust cases against technology companies, it is possible that (for whatever reason) intermediaries fail to adhere to internal or external guidelines, resulting in detrimental treatment of users, advertisers, content providers, or other actors. In contrast to past well-documented legal disputes (e.g., in competition regulation), affected parties will have difficulties monitoring for, detecting, and collecting evidence of unfair treatment because (due to personalization and highly individual behavior) they do not have access to the recommendations that intermediaries produce for their users. Assessing (and proving) unfair bias in intermediaries would require access to a representative set of recommendations, so that differences in consumption could be clearly attributed. Consider, for example, a hypothetical search engine that systematically alters access to political cross-cutting information—displaying only conservative results to conservatives and only liberal results to liberals. Creating a legal case against such a platform would require access to a representative set of users (anecdotal evidence could always be discounted as spurious. Or chance findings). For each of those users, researchers should identify the political leaning, and then record the search results they obtained. Such data collection is typically unfeasible in practice for two reasons: (a) Platforms offer no way of accurately recording the output of an intermediary for a single user. Neither third parties, nor the users themselves, have access to technical interfaces that would show a comprehensive dataset of personalized recommendations, such as the personal news feed on Facebook. Even though users can access the feed visually in their browser, considerable effort would be required to extract it in an automated fashion (i.e., through web scraping). Furthermore, there is currently no company that provides such data, and researchers' capabilities to obtain them independently are increasingly limited by the locked nature of proprietary smartphones (Jürgens, Stark, & Magin, 2019). The second factor that encumbers research is (b) the unavailability of a suitable sampling

strategy. Intermediaries are, of course, cognizant of their entire population of users along with key socio-demographic data (which is required i.e., for selling personalized advertisement spaces). Researchers, on the other hand, usually have no way of creating random samples from a platform's population. Since the available information on the socio-demographic makeup of the userbase is typically drawn from (moderately sized, i.e., N in the thousands) surveys, attempts to create samples which are representative, with regard to the intermediary's national users, offer somewhat limited precision (Jürgens et al., 2019).

- (2) Outside actors (individual, institutional and state-sponsored) frequently attempt to manipulate intermediaries in order to gain influence over citizens—e.g., through misinformation, disinformation, and the manipulation of public opinion perception (Lazer et al., 2017; Magin et al., 2019). Although intermediaries spend commendable resources on the containment and removal of such attempts, two risks remain which are outside the reach of the companies themselves: (a) Without access to large-scale data containing purported manipulation attempts, external watchdogs cannot perform independent audits in order to identify overlooked external influences. (b) Problematic content is also routinely deleted, so that external watchdogs cannot scrutinize and understand those attacks. An exception is Twitter, which regularly publishes datasets on malevolent campaigns).[9] Intermediaries should provide trustworthy actors with a way to perform their own, independent attempts at large-scale detection of manipulation, including data that were already removed by in-house systems. The simplest strategy that would enable such attempts is simply making the raw public flow of information accessible through a technical interface (an API), as Twitter has done: It not only offers a true random sample of tweets, but includes information about which of them are removed later on. Furthermore, the company offers a growing list of datasets containing all content produced by bot networks and state-sponsored manipulation campaigns. Platforms with more restrictive privacy contexts such as Facebook (where much of the flow of information is not visible to the broader usership or public) could still allow automated analyses, for example by offering to run researchers' models without providing access to the data itself.

- (3) In addition to institutional actors, harmful dynamics (such as radicalizing echo chambers, Stark et al., in press) may develop within intermediaries, even in the absence of external influence. Such dynamics need not have a clearly identifiable culprit; they could equally arise from the interaction of multiple individuals that leads to a mutual reinforcement of harmful tendencies. Detecting such structural phenomena is contingent on a complete picture of users' interaction networks. Furthermore, singular snapshots do not provide much insight; instead, the development must be traced over time in order to assess the true impact as well as causes. Researchers should, therefore, gain access to representative, longitudinal, and detailed data on those (semi-public) parts of intermediaries that pertain to public debates. This includes first and foremost the discussions surrounding media, politics, and social issues.

New Models for Partnerships

Some attempts have been made to increase transparency; while they attempt to address the issues outlined above, they have not yet achieved any significant success. Following an initiative from King and Persily (2019), a consortium of scientists cooperated with Facebook in order to create an institution

9 https://blog.twitter.com/official/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html.

(Social Science One)[10] and a process that would allow scholarly usage of a limited collection of pre-defined datasets. The project received much criticism from scientists, who warned that it would decrease transparency in research practices, lead to a dependence on Facebook that would encumber critical research, and create divisions between privileged "insiders" with access to data and the rest of the field (Bruns, 2019c). So far, Facebook also failed to facilitate the agreed upon access, frustrating even the existing scientific partners (Statement from the European Advisory Committee to Social Science One). Despite its pragmatic appeal, the cooperative model underpinning Social Science One has a fatal conceptual flaw: Even though it provides some access to some data, that access is pre-defined and limits researchers to a specific approach in tackling the above-mentioned threats. Participating teams are prevented from finding problems, answers, and solutions that intermediaries themselves did not identify. A cooperation between an intermediary and scientific partners can only succeed in generating trust if researchers are given the freedom to seek and find potential negative effects. Where such inquiries are prohibited ex ante, through pre-defined datasets or topical questions, both sides suffer from a lack of credibility.

There is also a deeper issue to the proposed cooperative model: Just as independent scholarly work from different institutes is required for long-term trustworthy, rigorous, and reliable scientific insights, independence is a defining feature of work on intermediaries. Only when external observers are free to implement an autonomous process for data collection, analysis, and interpretation can they serve as the much-needed check and balance that society requires (and demands). Researchers' ability to do so ultimately hinges on two key ingredients, both mentioned above: The capacity to obtain or create high-quality representative samples, and the availability of tools that record digital content, recommendations and user behavior within intermediaries. While the first is certainly possible (if perhaps expensive), the second remains under strong pressure from the progressive "lock-down" of platforms and mobile devices (Jürgens et al., 2019).

Diversity as Policy Goal

From a normative point of view, diversity is the key term: "Media pluralism and diversity of media content are essential for the functioning of a democratic society," as the Council of Europe (2007) put it, because functioning democracies require all members of society to be able to participate in public debate—and to be able to fully participate in this democratic debate, citizens need to encounter a diversity of high-quality information and opinions. Or, as the Council of Europe (2007) adequately phrased it, the right to freedom of expression "will be fully satisfied only if each person is given the possibility to form his or her own opinion from diverse sources of information." The close link between media diversity and democratic participation may also explain the scope of the public debate about the rise of platforms and their growing influence in the information landscape. The advances of at least some of these platforms into the business of distributing and aggregating media content have fundamentally changed our news ecosystem. In this context, a very important aspect is the newly created economic dependencies. The better part of the advertising revenue flows to the platform providers, Google and Facebook. This raises the normative question: to what extent should diversity matter in the context of social media platforms?

Communications researchers (e.g., Moeller, Helberger, & Makhortykh, 2019) emphasize that to ensure the preservation of the news ecosystem that has come under considerable strain, the public needs to receive the curated news supply provided by traditional mass media. The length of time that traditional mass media can still assume this function is uncertain. If, at some point, this is no longer the case, the

10 For more information: https://socialscience.one/.

underlying conditions of the news ecosystem and consequently opinion-formation will fundamentally change. If traditional mass media disappears, high-quality news products will no longer be available, and the impact of low-quality information on processes of opinion formation will rise. If only softened and highly personalized content will be distributed, the news ecosystem will change dramatically with potentially negative consequences for democratic societies.

Against this background, the social dynamics of media diversity become apparent (Helberger, 2018, p. 156). Thus, "media diversity on social media platforms must be understood as a cooperative effort of the social media platform, media organizations, and users. The way users search for, engage with, like, shape their network, and so forth has important implications for the diversity of content, ideas and encounters that they are exposed to. Similarly, the way the traditional media collaborate with information intermediaries to distribute content and reach viewers impacts structural diversity" (Helberger, 2018, p. 171). Put differently, future diversity policies must therefore go beyond the traditional framework and generate a new conception of media diversity, which addresses the different actors (platforms, users, and media organizations) together. These future policies must, first and foremost, ensure that diversity reaches users.

A potential way of increasing exposure diversity could be to employ a design that focuses on serendipity and/or on diversity as a principle (see considerations on "diversity by design": Helberger, 2011). Such a design would, for example, focus less on search engine ranking and would encourage users to have a closer look at different teasers and click on more results. Besides, users should have the opportunity to choose between, or weight, different filtering and sorting criteria. Such changes could also create more diversity in Facebook's news feed, e.g., Facebook could implement the ability for users to adopt a different point of view, exposing them to a feed with totally new perspectives (Ash, Gorwa, &, Metaxa, 2019).

As the debate about the impact of algorithmic news recommenders on democracy is still an ongoing process, diversity-sensitive design as part of a possible solution should be taken into account. For such solutions to work, it should be clear that different perspectives on the democratic role of news recommenders imply different design principles for recommendation systems (Helberger, 2019), i.e., an explicit normative conception of the democratic potential is critical. It may also become clear, that we need to work towards a coherent mix of appropriate government regulation, co-regulation, and platform-specific self-regulation in order to minimize the negative effects of the discussed threats.

References

A

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. doi: 10.1257/jep.31.2.211

Allcott, H., Gentzkow, M., & Yu, C. (2018). *Trends in the Diffusion of Misinformation on Social Media*. Retrieved from: http://arxiv.org/abs/1809.05901

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "nasty effect:" Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication*, *19*(3), 373–387. doi: 10.1111/jcc4.12009

Anderson, A. A., Yeo, S. K., Brossard, D., Scheufele, D. A., & Xenos, M. A. (2018). Toxic talk: How online incivility can undermine perceptions of media. *International Journal of Public Opinion Research*, *30*(1), 156–168. doi: 10.1093/ijpor/edw022

Arendt, F., Haim, M., & Beck, J. (2019). Fake News, Warnhinweise und perzipierter Wahrheitsgehalt: Zur unterschiedlichen Anfälligkeit für Falschmeldungen in Abhängigkeit von der politischen Orientierung [engl.: Fake news, warnings and perceived truthfulness: On the different susceptibility to false reports depending on political orientation]. *Publizistik*, *64*, 181-204. doi: 10.1007/s11616-019-00484-4

Ash, T.G., Gorwa, R., & Metaxa, D. (2019). *Glasnost? Nine ways Facebook can make itself a better forum for free speech and democracy*. Oxford, UK: Reuters Institute for the Study of Journalism.

Auxier, B. E., & Vitak, J. (2019). Factors Motivating Customization and Echo Chamber Creation Within Digital News Environments. *Social Media + Society*, *5*, 1-13. doi: 10.1177/2056305119847506

B

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

Barberá, P. (2015). *Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S.* Retrieved from: http://pablobarbera.com/static/barbera_polarization_APSA.pdf.

Barberá, P. (in press). Social Media, Echo Chambers, and Political Polarization. In N. Persily, & J. Tucker (Eds.), *Social Media and Democracy: The State of the Field*. Cambridge: Cambridge University Press.

Barrera Rodriguez, O., Guriev, S. M., Henry, E., & Zhuravskaya, E. (2018). *Facts, Alternative Facts, and Fact Checking in Times of Post-Truth Politics*. Retrieved from: https://ssrn.com/abstract=3004631

Bayer J., Bitiukova, N., Bárd, P., Alemanno, A., Szakács, J., & Uszkiewicz, E. (2019). *Disinformation and Propaganda – Impact on the functioning of the rule of law in the EU and its member states*. Brussels: European Union. Retrieved from: http://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2019)608864

Beam, M. A., Hutchens, M. J., & Hmielowski, J. D. (2018). Facebook news and (de)polarization: reinforcing spirals in the 2016 US election. *Information, Communication & Society*, *21*(7), 940–958. doi: 10.1080/1369118X.2018.1444783

Beam, M., & Kosicki, G. (2014). Personalized news portals: Filtering systems and increased news exposure. *Journalism & Mass Communication Quarterly*, *91*(1), 59–77. doi: 10.1177/1077699013514411

Bechmann, A., & Nielbo, K. L. (2018). Are We Exposed to the Same "News" in the News Feed? An empirical analysis of filter bubbles as information similarity for Danish Facebook users. *Digital Journalism*, *6*(8). 990-1002. doi: 10.1080/21670811.2018.1510741

Bell, E., & Owen, T. (2017). *The Platform Press: How Silicon Valley Reengineered Journalism.* New York: Tow Center for Digital Journalism.

Ben-David, A., & Matamoros-Fernández, A. (2016). Hate Speech and covert discrimination on social media: Monitoring the Facebook pages of extreme-right political parties in Spain. *International Journal of Communication*, *10*, 1167–1193. Retrieved from: https://ijoc.org/index.php/ijoc/article/view/3697/1585

Bennett, W. L., & Iyengar, S. (2008). A new era of minimal effects? The changing foundations of political communication. *Journal of Communication*, *58*(4), 707–731. doi:10.1111/j.1460-2466.2008.00410.x

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*, 122–139. doi:10.1177/0267323118760317

Binns, A. (2017). Fair game? Journalists' experiences of online abuse. *Journal of Applied Journalism & Media Studies*, *6*(2), 183–206. doi: 10.1386/ajms.6.2.183_1

Blom, J. N., & Hansen, K. R. (2015). Click bait: forward-reference as lure in online news headlines. *Journal of Pragmatics*, *76*, 87–100. doi: 10.1016/j.pragma.2014.11.010

Bode, L. (2016). Who sees what? Individual exposure to political information via social media. In G.W. Richardson (Ed.), *Social media and politics: A new way to participate in the political process* (pp. 189-204). Santa Barbara, CA: Praeger.

Bodó, B., Helberger, N., Eskens, S., & Moeller, J. (2019). Interested in diversity: The role of user attitudes, algorithmic feedback loops, and policy in news personalization. *Digital Journalism*, *7*(2), 206–229. doi: 10.1080/21670811.2018.1521292

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, *15*(3), 209–227. doi: 10.1007/s10676-013-9321-6

Bright, J. (2018). Explaining the Emergence of Political Fragmentation on Social Media: The Role of Ideology and Extremism. *Journal of Computer-Mediated Communication*, *23*(1), 17–33. doi:10.1093/jcmc/zmx002

Brown, A. (2018). What is so special about online (as compared to offline) hate speech? *Ethnicities*, *18*(3), 297–326. doi: 10.1177/1468796817709846

Bruns, A. (2019a). *It's not the technology, stupid: How the 'Echo Chamber' and 'Filter Bubble' metaphors have failed us*. Madrid: International Association for Media and Communication Research (IAMCR).

Bruns, A. (2019b). Filter bubble. *Internet Policy Review*, *8*(4), 1-14. doi: 10.14763/2019.4.1426

Bruns, A. (2019c). After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, *22*(11), 1544–1566. doi: 10.1080/1369118X.2019.1637447

Burkart, R. (1995). *Kommunikationswissenschaft: Grundlagen und Problemfelder - Umrisse einer interdisziplinären Sozialwissenschaft [engl.: Communication Science: Foundations and problem areas – outlines of an interdisciplinary social science].* Wien: Böhlau.

Chadwick, A., & Vaccari, C. (2019). *News Sharing on UK Social Media: Misinformation, Disinformation & Correction*. Loughborough: Online Civic Culture Centre, Loughborough University.

C

Chen, G. M., & Lu, S. (2017). Online political discourse: Exploring differences in effects of civil and uncivil disagreement in news website comments. *Journal of Broadcasting & Electronic Media*, *61*(1), 108–125. doi: 10.1080/08838151.2016.1273922

Cicchirillo, V., Hmielowski, J., & Hutchens, M. (2015). The mainstreaming of verbally aggressive online political behaviors. *Cyberpsychology, Behavior, and Social Networking*, *18*(5), 253–259. doi: 10.1089/cyber.2014.0355

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658–679. doi: 10.1111/jcom.12104

Coleman, S. (2012). Believing the news: From sinking trust to atrophied efficacy. *European Journal of Communication*, *27*(1), 35–45. doi: 10.1177/0267323112438806

Coleman, S., Anthony, S. & Morrison, D. E. (2009). *Public Trust in the News. A constructivist study of the social life of the news.* Oxford: Reuters Institute for the Study of Journalism. Retrieved from: http://reutersinstitute.politics.ox.ac.uk/fileadmin/documents/Publications/Public_Trust_in_the_News.pdf

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., & Menczer, Fi. (2011). *Political Polarization on Twitter*. Menlo Park, CA: 5th International AAAI Conference on Weblogs and Social Media (ICWSM).

Cornils, M. (2020). Designing platform governance: *A normative perspective on needs, strategies, and tools to regulate intermediaries.* Berlin: AlgorithmWatch. Retrieved from: https://algorithmwatch.org/en/governingplatforms/legal-study-cornils-may-2020

Council of Europe (1997). Recommendation No. R (97) 20 of the committee of ministers to member states on "Hate Speech". Retrieved from: https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680505d5b

Council of Europe (2007). *Recommendation Rec(2007)2 of the committee of ministers to member states on media pluralism and diversity of media content.*

Courtois, C., Slechten, L., & Coenen, L. (2018). Challenging Google Search filter bubbles in social and political information: Disconforming evidence from a digital methods case study. *Telematics and Informatics*, *35*(7), 2006–2015. doi: 10.1016/j.tele.2018.07.004

Craker, N., & March, E. (2016). The dark side of Facebook®: The Dark Tetrad, negative social potency, and trolling behaviours. *Personality and Individual Differences*, *102*, 79–84. doi: 10.1016/j.paid.2016.06.043

D

DeVito, M. A. (2017). From Editors to Algorithms. *Digital Journalism*, *5*(6), 753-773. doi: 10.1080/21670811.2016.1178592

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

Dogruel, L., Stark, B., Facciorusso, D., & Liesem, K. (2020). Die Regulierung von Algorithmen aus Expertensicht. Transparenz und Diskriminierungsfreiheit – zur Vielfaltssicherung im neuen Medienstaatsvertrag [engl.: The Regulation of Algorithms from an Expert's Point of View. Transparency and Non-discrimination to Ensure Diversity in the New Media State Treaty]. *Media Perspektiven, 3,* 139-148.

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society 21*(5), 729–745. doi: 10.1080/1369118X.2018.1428656

Duggan, M., & Smith, A. (2016). *The Political Environment on Social Media*. Washington, DC: Pew Research Center. Retrieved from: http://assets.pewresearch.org/wp-content/uploads/sites/14/2016/10/24160747/PI_2016.10.25_Politics-and-Social-Media_FINAL.pdf

Dylko, I. B., Dolgov, I., Hoffman, W., Eckhart, N., Molina, M., & Aaziz, O. (2017). The dark side of technology. An experimental investigation of the influence of customizability technology on online political selective exposure. *Computers in Human Behavior*, *73*, 181–190. doi: 10.1016/j.chb.2017.03.031

E

Egelhofer, J. L., & Lecheler, S. (2019). Fake news as a two-dimensional phenomenon: a framework and research agenda. *Annals of the International Communication Association*, *43*(2), 97-116. doi: 10.1080/23808985.2019.1602782

Eickelmann, J., Grashöfer, K., & Westermann, B. (2017). #NETZDG #MAASLOS. Eine Stellungnahme zum Netzwerkdurchsetzungsgesetz [eng.: #NETZDG #MAASLOS. A statement on the Network Enforcement Act]. *Zeitschrift für Medienwissenschaft*, *9*(17), 176-185. doi: 10.25969/mediarep/2311

Erjavec, K., & Kovačič, M. P. (2012). "You Don't Understand, This is a New War!" Analysis of Hate Speech in News Web Sites' Comments. *Mass Communication and Society*, *15*(6), 899-920. doi: 10.1080/15205436.2011.619679

Eurobarometer (2018). *Fake news and disinformation online*. Retrieved from: https://ec.europa.eu/digital-single-market/en/news/final-results-eurobarometer-fake-news-and-online-disinformation

European Commission (2016). *Code of Conduct on Countering Illegal Hate Speech Online*. Retrieved from: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/countering-illegal-hate-speech-online_en

European Commission (2018a). *A multi-dimensional approach to disinformation*. Report of the High level Group on fake news and online disinformation. Retrieved from: https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation

European Commission (2018b). *Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Action Plan against Disinformation*. Retrieved from: https://op.europa.eu/en/publication-detail/-/publication/e18263d2-f962-11e8-8885-01aa75ed71a1/language-en/format-PDF/source-search

F

Facebook (2019). *NetzDG Transparency Report*. Retrieved from: https://fbnewsroomus.files.wordpress.com/2019/07/facebook_netzdg_july_2019_english.pdf

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Fiorina, M. P., & Abrams, S. J. (2008). Political Polarization in the American Public. *Annual Review of Political Science*, *11*, 563–588. doi: 10.1146/annurev.polisci.11.053106.153836

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, *80*(S1), 298–320. doi:10.1093/poq/nfw006

Fletcher, R., Cornia, A., Graves, L., & Nielsen, R. K. (2018). *Measuring the Reach of „Fake News" and Online Disinformation in Europe*. Oxford: Reuters Institute for the Study of Journalism. Retrieved from: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/Measuring%20the%20reach%20of%20fake%20news%20and%20online%20distribution%ft20in%20Europe%20CORRECT%20FLAG.pdf

Fletcher, R., & Nielsen, R. K. (2017). Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication. *Journal of Communication*, *67*(4), 476–498. doi: 10.1111/jcom.12315

G

Galtung, J., & Ruge, M. H. (1965). The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers. *Journal of Peace Research*, *2*(1), 64–90. doi: 10.1177/002234336500200104

Garimella, K., De Francisci Morales, G., Gionis, A., & Mathioudakis, M. (2018). Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web – WWW '18* (pp. 913–922). Lyon: ACM Press.

Geiß, S., Magin, M., Stark, B., & Jürgens, P. (2018). „Common meeting ground" in Gefahr? Selektionslogiken politischer Informationsquellen und ihr Einfluss auf die Fragmentierung individueller Themenhorizonte [engl.: Endangered Common Meeting Ground? Selection Logics of Political Information Sources and their Influence on the Fragmentation of Individual Issue Horizons]. *M&K – Medien & Kommunikationswissenschaft, 66*(4), 502–525. doi: 10.5771/1615-634X-2018-4-502

Gisler, A. (2016). Social Media. In fög (Eds.), *Qualität der Medien Jahrbuch 2016 [engl.: Quality of media yearbook 2016]* (pp. 99-113). Basel: Schwabe.

Glaser, J., Dixit, J., & Green, D. P. (2002). Studying hate crime with the internet: What makes racists advocate racial violence? *Journal of Social Issues*, *58*(1), 177–193. Retrieved from: https://gspp.berkeley.edu/research/selected-publications/studying-hate-crime-with-the-internet-what-makes-racists-advocate-racial-vi

Gleich, U. (2019). Auswirkungen von Echokammern auf den Prozess der Meinungsbildung: ARD-Forschungsdienst [engl.: Effects of Echo Chambers on the Process of Opinion Formation: ARD Research Service]. *Media Perspektiven*, *2*, 82–85.

Gorwa, R., & Ash, T.G. (2020). Democratic Transparency in the Platform Society. In N. Persily, & Josh Tucker (Eds.), *Social Media and Democracy: The State of the Field.* New York, NY: Cambridge University Press.

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*. *78*, 1360–1380.

Grinberg, N., Joseph, K., Friedland, L, Swire-Thompson, B., & Lazar, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, *363*, 374–378. doi: 10.1126/science.aau2706

Gruzd, A., & Roy, J. (2014). Investigating Political Polarization on Twitter: A Canadian Perspective. *Policy & Internet*, *6*, 28–45. doi: 10.1002/1944-2866.POI354

Guess, A., Nyhan, B., & Reifler, J. (2018). *Selective Exposure to Misinformation: Evidence from the Consumption of Fake News during the 2016 U.S. Presidential Campaign*. Retrieved from: https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*, online first. Retrieved from: https://advances.sciencemag.org/content/advances/5/1/eaau4586.full.pdf

H

Hagen, L. M., Wieland, M., & in der Au, A.-M. (2017). Algorithmischer Strukturwandel der Öffentlichkeit. Wie die automatische Selektion im Social Web die politische Kommunikation verändert und welche Gefahren dies birgt [engl.: Algorithmic Transformation of the Public sphere. How Automatic Selection in the Social Web is Changing Political Communication and the Dangers this Entails]. *Medien Journal*, *41*(2), 127–143. doi: 10.24989/medienjournal.v41i2.1476

Haim, M, Graefe, A., & Brosius, H.-B. (2018). Burst of the Filter Bubble? Effects of personalization on the diversity of Google News. *Digital Journalism*, *6*(3), 330–343. doi: 10.1080/21670811.2017.1338145

Hameleers, M., & van der Meer, T. G. (2020). Misinformation and polarization in a high-choice media environment: How effective are political fact-checkers? *Communication Research*, *47*(2), 227–250. doi: 10.1177/0093650218819671

Hameleers, M., Powell, T. E., Van Der Meer, T. G., & Bos, L. (2020). A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated via Social Media. *Political Communication*, *37*(2), 281–301. doi: 10.1080/10584609.2019.1674979

Hegelich, S., & Thieltges, A. (2019). Desinformation und Manipulation. Wie soziale Netzwerke die Meinungsbildung steuern: Ein Werkstattbericht [engl.: Disinformation and Manipulation. How Social Networks Steer Opinion Formation: a Workshop Report], In K. Holetschek, & U. Männle (Eds.,): *Mittelpunkt Bürger: Dialog, Digital und Analog* [*engl.: Focus Citizen: Dialogue, Digital and Analog*] (pp. 97-108). Munich: Hanns-Seidel-Stiftung e.V.

Helberger, N. (2011). Diversity by design. *Journal of Information Policy*, *1*, 441–469. doi: 10.5325/jinfopoli.1.2011.0441

Helberger, N. (2018). Challenging Diversity - Social Media Platforms and a New Conception of Media Diversity. In M. Moore, & D. Tambini (Eds.), *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple* (pp. 153-175). New York, NY: Oxford University Press.

Helberger, N. (2019). On the Democratic Role of News Recommenders. *Digital Journalism*, *78*, 993-1012. doi: 10.1080/21670811.2019.1623700

Himelboim, I., McCreery, S., & Smith, M. (2013). Birds of a Feather Tweet Together: Integrating Network and Content Analyses to Examine Cross-Ideology Exposure on Twitter. *Journal of Computer-Mediated Communication*, *18*(2), 40–60. doi: 10.1111/jcc4.12001

Hölig, S., & Hasebrink, U. (2019). *Reuters Institute Digital News Report 2019 – Ergebnisse für Deutschland. Unter Mitarbeit von Julia Behre [engl.: Reuters Institute Digital News report 2019 – Results for Germany. In collaboration with Julia Behre]* . Hamburg: Verlag Hans-Bredow-Institut, Juni 2019 (Arbeitspapier des HBI Nr. 47). Retrieved from: https://www.hans-bredow-institut.de/de/publikationen/reuters-institute-digital-news-report-2019-ergebnisse-fuer-deutschland

Hsueh, M., Yogeeswaran, K., & Malinen, S. (2015). "Leave your comment below": Can biased online comments influence our own prejudicial attitudes and behaviors? *Human Communication Research*, *41*(4), 557–576. doi: 10.1111/hcre.12059

Huang, H. (2017). A War of (Mis)Information: The Political Effects of Rumors and Rumor Rebuttals in an Authoritarian Country. *British Journal of Political Science*, *47*(2), 283–311. doi: 10.1017/S0007123415000253

Hutchens, M. J., Cicchirillo, V. J., & Hmielowski, J. D. (2015). How could you think that?!?!: Understanding intentions to engage in political flaming. *New Media & Society*, *17*(8), 1201–1219. doi: 10.1177/1461444814522947

I

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology. *Public Opinion Quarterly*, *76*(3), 405–431. doi: 10.1093/poq/nfs038

J

Jaursch, J. (2019). *Regulatory Reactions to Disinformation: How Germany and the EU are trying to tackle opinion manipulation on digital platforms*. Retrieved from: https://www.stiftung-nv.de/sites/default/files/regulatory_reactions_to_disinformation_in_germany_and_the_eu.pdf

Johnson, N.F., Leahy, R., Restrepo, N.J., ..., Wuchty, S. (2019). Hidden resilience and adaptive dynamics of the global online hate ecology. *Nature*, *573*, 261–265. doi: 10.1038/s41586-019-1494-7

Jourová, V. (2019). *Code of Conduct on Countering Illegal Hate Speech Online. Fourth evaluation confirms self-regulation works.* Retrieved from: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/countering-illegal-hate-speech-online_en

Jürgens, P., & Stark, B. (2017). The Power of Default on Reddit: A General Model to Measure the Influence of Information Intermediaries. *Policy & Internet*, *9*(4), 395–419. doi: 10.1002/poi3.166

Jürgens, P., Stark, B., & Magin, M. (2014). Gefangen in der Filter Bubble? Search Engine Bias und Personalisierungsprozesse bei Suchmaschinen [engl.: Caught in the Filter Bubble? Search Engine Bias and Personalization of Search Engines]. In B. Stark, D. Dörr, & S. Aufenanger (Eds.), *Die Googleisierung der Informationssuche. Suchmaschinen zwischen Nutzung und Regulierung [engl.: The "Googleization" of Information search. Search engines between usage and regulation]* (pp. 98-135). Berlin: de Gruyter.

Jürgens, P., Stark, B., & Magin, M. (2019). Two Half-Truths Make a Whole? On Bias in Self-Reports and Tracking Data. *Social Science Computer Review*. doi: 10.1177/0894439319831643

Jungherr, A. (2019). Desinformation: Konzepte, Identifikation, Reichweite und Effekte [engl.: Disinformation: Concepts, Identification, Reach and Effects.]. In Landesanstalt für Medien (Lfm) NRW (Ed.): *Was ist Desinformation? Betrachtung aus sechs wissenschaftlichen Perspektiven [engl.: What is Disinformation? Consideration from Six Scientific Perspectives]* (pp. 23-30). Düsseldorf: Landesanstalt für Medien NRW.

K

Karlsen, R., Steen-Johnsen, K., Wollebæk, D., & Enjolras, B. (2017). Echo chamber and trench warfare dynamics in online debates. *European Journal of Communication*, *32*(3), 257–273. doi: 10.1177/0267323117695734

Karlsson, M. B. (2016). Goodbye politics, hello lifestyle. Changing news topics in tabloid, quality and local newspaper websites in the U.K. and Sweden from 2002 to 2012. *Observatorio*, *10*(4), 150–165. doi.org/10.15847/obsOBS1042016940

Katz, E. (1996). And deliver us from segmentation. *Annals of the American Academy of Political and Social Science*, *546*, 22–33. Retrieved from: http://www.jstor.org/stable/1048167

King, G., & Persily, N. (2019). A New Model for Industry–Academic Partnerships. *PS: Political Science & Politics*, 1–7. doi: 10.1017/S1049096519001021

Klinger, U., & Svensson, J. (2015). The emergence of network media logic in political com-munication: A theoretical approach. *New Media & Society*, *17*(8), 1241-1257. doi: 10.1177/1461444814522952

Kohring, M., & Zimmermann,F. (2019). Aktuelle Desinformation: Definition – Konsequenzen – Reichweite [engl.: Current Disinformation: Definition – Consequences – Reach]. In Landesanstalt für Medien (Lfm) NRW (Ed.): *Was ist Desinformation? Betrachtung aus sechs wissenschaftlichen Perspektiven [engl.: What is Disinformation? Consideration from six scientific perspectives]* (pp. 15-22). Düsseldorf: Landesanstalt für Medien NRW.

Krafft, T. D., Gamer, M., & Zweig, K. A. (2018). *Wer sieht was? Personalisierung, Regionalisierung und die Frage nach der Filterblase in Googles Suchmaschine [engl.: What did you see? – Personalization, regionalization and the question of the filter bubble in Google's search engine]*. Medienanstalt Berlin Brandenburg. Retrieved from: https://www.mabb.de/files/content/document/INFORMATION/Digitale%20Welt/Abschlussbericht%20Datenspende.pdf

Kümpel, A. S., & Rieger, D. (2019). *Wandel der Sprach- und Debattenkultur in sozialen Online-Medien. Ein Literaturüberblick zu Ursachen und Wirkungen von inziviler Kommunikation [engl.: Change of language and debate culture in social online-media: A literature review on the causes and effects of uncivil communication].* Berlin: Konrad-Adenauer-Stiftung. Retrieved from: https://www.kas.de/einzeltitel/-/content/wandel-der-sprach-und-debattenkultur-in-sozialen-online-medien

Kuiken, J., Schuth, A., Spitters, M., & Marx, M. (2017). Effective Headlines of Newspaper Articles in a Digital Environment, *Digital Journalism*, *5*(10), 1300-1314. doi: 10.1080/21670811.2017.1279978

L

Lazer, D., Baum, M., Grinberg, N., Friedland, L., Joseph, K., Hobbs, W., & Mattsson, C. (2017). *Combating fake news: An agenda for research and action*. Harvard Kennedy School, Shorenstein Center on Media. Politics and Public Policy, 2.

Lee, J. K. (2009). *Incidental exposure to news: limiting fragmentation in the new media environment* (dissertation). University of Texas, Austin (USA), Texas.

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*, 353–369. doi: 10.1016/j.jarmac.2017.07.008

Li, J. (2020). Toward a Research Agenda on Political Misinformation and Corrective Information. *Political Communication*, *37*(1), 125–135. doi: 10.1080/10584609.2020.1716499

Lischka, J. A. (2018). Logics in social media news making: how social media editors marry the Facebook logic with journalistic standards. *Journalism.* doi:10.1177/1464884918788472

Löber, L., & Roßnagel, A. (2019). Das Netzwerkdurchsetzungsgesetz in der Umsetzung. Bilanz nach den ersten Transparenzberichten [engl.: The Network Enforcement Act in its implementation. Résumé after the first transparency reports]. *Multimedia und Recht* (MMR), 71-76.

Loosen, W., & Scholl, A. (2017). Journalismus und (algorithmische) Wirklichkeitskonstruktion. Epistemologische Beobachtungen [engl.: Journalism and (Algorithmic) Reality Construction. Epistemological Observation]. *Medien & Kommunikationswissenschaft*, *65*(2), 348–366. doi: 10.5771/1615-634X-2017-2-348

Lupu, N. (2015). Party Polarization and Mass Partisanship: A Comparative Perspective. *Political Behavior*, *37*(2), 331–356. doi: 10.1007/s11109-014-9279-z.

M

Magin, M., Geiß, S., Jürgens, P., & Stark, B. (2019). Schweigespirale oder Echokammer? Zum Einfluss sozialer Medien auf die Artikulationsbereitschaft in der Migrationsdebatte [engl.: Spiral of Silence or Echo Chamber? On the Influence of Social Media on the Willingness to Speak Out in the Migration Debate]. In P, Weber, F. Mangold, M. Hofer, & T. Koch (Eds.): *Meinungsbildung in der Netzöffentlichkeit [engl.: Opinion Formation in the Networked Public Sphere]* (pp. 93–114). Baden-Baden: Nomos.

Magin, M., Steiner, M., & Stark, B. (2019). Suche im Netz – einseitige oder vielfältige Informationsquelle? Eine Inhaltsanalyse der politischen Informationsvielfalt von Suchmaschinenergebnissen [engl.: Searching the net – one sided or diverse source of information? A content analysis of the diversity of information of search engine results]. *Media Perspektiven*, *9*, 421-429.

Martens, B., Aguiar, L., Gomez-Herrera, E., & Mueller-Langer, F. (2018). *The digital transformation of news media and the rise of disinformation and fake news- An economic perspective; Digital Economy Working Paper 2018-02; JRC Technical Reports*. Retrieved from: https://ec.europa.eu/jrc/communities/sites/jrccties/files/dewp_201802_digital_transformation_of_news_media_and_the_rise_of_fake_news_final_180418.pdf

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology*, *27*, 415–444. doi: 10.1146/annurev.soc.27.1.415

Medienanstalten, die (2019). Vielfaltsbericht 2019 der Medienanstalten [engl.: Diversity Report 2019 of the Media Institutions]. Berlin: die Medienanstalten – ALM GbR

Medienstaatsvertrag (2019). *Staatsvertrag zur Modernisierung der Medienordnung in Deutschland [engl.: State Treaty on the modernisation of the media order in Germany]*. Retrieved from: https://www.rlp.de/fileadmin/rlp-stk/pdf-Dateien/Medienpolitik/Mod-StV_MStV_und_JMStV_2019-12-05_MPK.pdf

Messing, S., & Westwood, S. J. (2012). Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. *Communication Research*, *41*, 1042-1063. doi: 10.1177/0093650212466406

Moeller, J., & Hameleers, M. (2019). Different Types of Disinformation, Its Political Consequences and Treatment Recommendations for Media Policy and Practice. In Landesanstalt für Medien (Lfm) NRW (Ed.): *Was ist Desinformation? Betrachtung aus sechs wissenschaftlichen Perspektiven [engl.: What is Disinformation? Consideration from six scientific perspectives]* (pp. 15-22). Düsseldorf: Landesanstalt für Medien NRW.

Moeller, J. & Helberger, N. (2018). *Beyond the filter bubble: concepts, myths, evidence and issues for future debates. A report drafted for the Dutch Media Regulator*. Amsterdam: University of Amsterdam. Retrieved from https://dare.uva.nl/search?identifier=478edb9e-8296-4a84-9631-c7360d593610

Moeller, J., Helberger, N., & Makhortykh, M. (2019). *Filter Bubbles in the Netherlands? University of Amsterdam*. Retrieved from: https://www.cvdm.nl/wp-content/uploads/2019/05/Rapport-Filterbubbelsin-Nederland.pdf

Moeller, J., Trilling, D., Helberger, N., Irion, K., & De Vreese, C. (2016). Shrinking Core? Exploring the Differential Agenda Setting Power of Traditional and Personalized News Media. *info*, *18*(6), 26–41. doi: 10.1108/info-05-2016-0020

Moeller, 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. doi: 10.1080/1369118X.2018.1444076

Mosseri, A. (2018). *Bringing People closer together*. Retrieved from: https://about.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/

Müller, P., & Denner, N. (2019). *„What can be done to counter Fake News"? Eine Analyse anhand der Entstehungsbedingungen und Wirkweisen gezielter Falschmeldungen im Internet [engl.: "What can be done to counter Fake News?" An analysis based on the conditions and effects of calculated false reports on the Internet].* Gutachten im Auftrag der Friedrich-Naumann-Stiftung für die Freiheit [engl.: Report commissioned by the Friedrich-Naumann Foundation for freedom] (2nd edition). Retrieved from: https://shop.freiheit.org/#!/Publikation/797

N

Napoli, P. M. (1999). Deconstructing the diversity principle. *Journal of Communication*, *49*(4), 7–34.

Napoli, P. M. (2019). *Social Media and the Public Interest: Media Regulation in the Disinformation Age*. New York, NY: Columbia University Press.

Nechushtai, E., & Lewis, S. C. (2019). What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior*, *90*, 298–307. doi: 10.1016/j.chb.2018.07.043

Nelson, J. L., & Taneja, H. (2018). The small, disloyal fake news audience: The role of audience availability in fake news consumption. *New Media & Society*, *20*(10), 3720–3737. doi:10.1177/1461444818758715

Nelson, J. L., & Webster, J. G. (2017). The Myth of Partisan Selective Exposure: A Portrait of the Online Political News Audience. *Social Media + Society*, 1-13. doi: 10.1177/2056305117729314

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2019). *Reuters Institute Digital News Report 2019.* Oxford: Reuters Institute for the Study of Journalism.

Nilsson, M. L., & Örnebring, H. (2016). Journalism under threat. Intimidation and harassment of Swedish journalists. *Journalism Practice*, *10*(7), 880–890. doi: 10.1080/17512786.2016.1164614

Noelle-Neumann, E. (2001). *Die Schweigespirale: öffentliche Meinung – unsere soziale Haut [engl: The Spiral of Silence: public opinion – our social skin]*. München: Langen Müller.

Ø

Ørmen, J. (2018). Testing the Myth of Enclaves: A discussion of research designs for assessing algorithmic curation. In S. A. Eldridge, & B. Franklin (Eds.): *The Routledge Handbook of Developments in Digital Journalism Studies* (pp. 132–142). London: Routledge.

O

Oksanen, A., Kaakinen, M., Minkkinen, J., Räsänen, P., Enjolras, B., & Steen-Johnsen, K. (2018). Perceived societal fear and cyberhate after the November 2015 Paris terrorist attacks. *Terrorism and Political Violence*, 1-20. doi: 10.1080/09546553.2018.1442329

Otto, L., Glogger, I., & Boukes, M. (2017). The softening of journalistic political communication: A critical review of sensationalism, soft news, infotainment, and tabloidization and a comprehensive framework model. *Communication Theory*, *27*(2), 136–155. doi: 10.1111/comt.12102

P

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you.* New York, NY: Penguin Press.

Post, S., & Kepplinger, H. M. (2019). Coping with audience hostility. How journalists' experiences of audience hostility influence their editorial decisions. *Journalism Studies*, *20*(16), 2422-2442. doi: 10.1080/1461670X.2019.1599725.

Prior, M. (2009).The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure. *Public Opinion Quarterly*, *73*(1), 130-143. doi: 10.1093/poq/nfp002

Prior, M. (2013). Media and Political Polarization. *Annual Review of Political Science, 16*, 101–127. doi: 10.1146/annurev-polisci-100711-135242

Prochazka, F., Weber, P., & Schweiger, W. (2018). Effects of civility and reasoning in user comments on perceived journalistic quality. *Journalism Studies*, *19*(1), 62–78. doi: 10.1080/1461670X.2016.1161497

Puschmann, C. (2018). Beyond the Bubble: Assessing the Diversity of Political Search Results. *Digital Journalism*, *7*(6), 824-843. doi: 10.1080/21670811.2018.1539626

R

Rashidin, N., Tsiveriotis, G., & Brown, P. (2019). *Platforms and Publishers: The End of an Era*. New York, NY: Tow Center for Digital Journalism at Columbia Journalism School. Retrieved from: https://www.cjr.org/tow_center_reports/platforms-and-publishers-end-of-an-era.php

Reicher, S. D., Spears, R., & Postmes, T. (1995). A social identity model of deindividuation phenomena. *European Review of Social Psychology*, *6*(1), 161–198. Retrieved from: https://www.researchgate.net/profile/Tom_Postmes/publication/240237685_A_Social_Identity_Model_of_Deindividuation_Phenomena/links/00b495252e2c20b382000000.pdf

Reinemann, C., Stanyer, J., & Scherr, S. (2016). Hard and soft news. In C. de Vreese, F. Esser, & D. N. Hopmann (Eds.), *Comparing political communication* (pp. 131–149). New York, NY: Routledge.

Rösner, L., & Krämer, N. C. (2016). Verbal venting in the social web: Effects of anonymity and group norms on aggressive language use in online comments. *Social Media + Society*, *2*(3), 1-13. doi: 10.1177/2056305116664220

Rowe, I. (2015). Deliberation 2.0: Comparing the deliberative quality of online news user comments across platforms. *Journal of Broadcasting & Electronic Media*, *59*(4), 539–555. doi: 10.1080/08838151.2015.1093482

S

Scharkow, M., Mangold, F., Stier, S., & Breuer, J. (2020). How social network sites and other online intermediaries increase exposure to news. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1918279117

Schudson, M. (2011). *The Sociology of News*. New York, NY; London: W. W. Norton.

Schweiger, W., Weber, P., Prochazka, F., & Brückner, L. (2019). *Algorithmisch personalisierte Nachrichtenkanäle. Begriffe, Nutzung, Wirkung [engl.: Algorithmically personalized news channels. Terms, usage, effects]*. Wiesbaden: Springer VS.

Settle, J. E. (2018). *Frenemies: How Social Media Polarizes America*. Cambridge: Cambridge University Press.

Shoemaker, P. J., Cohen, A. A. (2005). *News around the world: Content, practitioner and the public*. London, New York: Routledge.

Shoemaker, P. J., & Reese, S. D. (2014). *Mediating the message in the 21st century: a media sociology perspective* (3rd edition). London, New York: Routledge.

Siapera, E., Moreo, E., & Zhou, J. (2018). *Hate Track. Tracking And Monitoring Racist Speech Online.* Dublin: IHREC. Retrieved from: https://www.ihrec.ie/app/uploads/2018/11/HateTrack-Tracking-and-Monitoring-Racist-Hate-Speech-Online.pdf

Smith, N., & Graham, T. (2019). Mapping the anti-vaccination movement on Facebook. *Information, Communication & Society*, *22*(9), 1310–1327. doi: 10.1080/1369118X.2017.1418406

Sobieraj, S. (2018). Bitch, slut, skank, cunt: Patterned resistance to women's visibility in digital publics. *Information, Communication & Society*, *21*(11), 1700–1714. doi: 10.1080/1369118X.2017.1348535

Stark, B. (2013). Fragmentierung Revisited: eine theoretische und methodische Evaluation im Internetzeitalter [engl.: Revisiting Fragmentation: a theoretical and methodological evaluation in the Internet Age]. In W. Seufert & F. Sattelberger (Eds..), *Langfristiger Wandel von Medienstrukturen [engl.: Long-term transformation of media structures]* (pp. 199-218). Baden-Baden: Nomos Verlag.

Stark, B., & Magin, M. (2019). Neuer Strukturwandel der Öffentlichkeit durch Informationsintermediäre: Wie Google, Facebook & Co. die Medien und den Journalismus verändern [engl.: The 'New Structural Transformation of the Public Sphere' through information intermediaries: How Google, Facebook & Co. change media and politics]. In M. Eisenegger, L. Udris, & P. Ettinger (Eds.), *Wandel der Öffentlichkeit und der Gesellschaft. Gedenkschrift für Kurt Imhof [engl. Change of public sphere and society. Commemorative volume for Kurt Imhof]* (pp. 377-406). Wiesbaden: Springer VS.

Stark, B., Magin, M., & Jürgens, P. (2014). Navigieren im Netz – Befunde einer qualitativen und quantitativen Nutzerbefragung [engl.: Navigating the web. Results of a qualitative and quantitative user survey]. In B. Stark, D. Dörr, & S. Aufenanger (Eds.), *Die Googleisierung der Informationssuche. Suchmaschinen zwischen Nutzung und Regulierung [engl.: The "Googleization" of Information search. Search engines between usage and regulation]* (pp. 20–74). Berlin: de Gruyter.

Stark, B., Magin, M., & Jürgens, P. (2017). *Ganz meine Meinung? Informationsintermediäre und Meinungsbildung – Eine Mehrmethodenstudie am Beispiel von Facebook [engl.: I second that!? Information intermediaries and opinion formation – a mixed-methods study taking the example of Facebook]*. Düsseldorf: LfM.

Stark, B., Magin, M., & Jürgens, P. (in press). Maßlos überschätzt. Ein Überblick über theoretische Annahmen und empirische Befunde zu Filterblasen und Echokammern [engl.: Excessively overrated. An overview of theoretical assumptions and empirical findings on filter bubbles and echo chambers]. In M. Eisenegger, R. Blum, P. Ettinger & M. Prinzing (Eds.), *Digitaler Strukturwandel der Öffentlichkeit: Historische Verortung, Modelle und Konsequenzen [engl.: The Digital Structural Transformation of the Public Sphere – Historical Perspectives, Models and Consequences]*. Wiesbaden: Springer VS.

Su, L. Y.-F., Xenos, M. A., Rose, K. M., Wirz, C., Scheufele, D. A., & Brossard, D. (2018). Uncivil and personal? Comparing patterns of incivility in comments on the Facebook pages of news outlets. *New Media & Society*, *20*(10), 3678-3699. doi: 10.1177/1461444818757205

Sunstein, C. R. (2001). *Republic.com*. Princeton, NJ: Princeton University Press.

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

Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. H. (2017). Processing political misinformation: Comprehending the Trump phenomenon. *Royal Society Open Science*, *4*(3). doi:10.1098/rsos.160802

T

Tewksbury, D. (2005). The seeds of audience fragmentation: Specialization in the use of online news sites. *Journal of Broadcasting & Electronic Media*, *49*(3), 332–348. doi: 10.1207/s15506878jobem4903_5

Tewksbury, D., & Rittenberg, J. (2008). Online News Creation and Consumption: Implications for Modern Democracies. In A. Chadwick & P.N. Howard (Eds.), *The Handbook of Internet Politics* (pp. 186–200). London: Routledge.

Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political Communication*, *33*, 460–480. doi: 10.1080/10584609.2015.1102187

Thorson, K., & Wells, C. (2016). Curated Flows: A Framework for Mapping Media Exposure in the Digital Age. *Communication Theory*, *26*(3), 309–328. doi: 10.1111/comt.12087

Tucker, J., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S. Stukal, D., & Nyhan, B. (2018). *Social Media, Political Polarization and Political Disinformation: A Review of the Scientific Literature*. Report prepared for the Hewlett Foundation. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3144139

Twitter (2019). *Netzwerkdurchsetzungsgesetzbericht: Januar – Juni 2019 [engl.: Report of the Network Enforcement Act: January – June 2019]*. Retrieved from: https://cdn.cms-twdigitalassets.com/content/dam/transparency-twitter/data/download-netzdg-report/netzdg-jan-jun-2019.pdf

U

Umbricht, A., & Esser, F. (2016). The push to popularize politics, understanding the audience-friendly packaging of political news in six media systems since the 1960s. *Journalism Studies*, *17*, 100–121. doi:10.1080/1461670X.2014.963369

V

van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., de Vreese, C., Matthes, J., ..., Stanyer, J. (2017). Political communication in a high-choice media environment: a challenge for democracy?. *Annals of the International Communication Association, 41*(1), 3-27. doi: 10.1080/23808985.2017.1288551

van Dijck, J., & Poell, T. (2013). *Understanding Social Media Logic*. ID 2309065, SSRN Scholarly Paper, 12. August. Rochester, NY: Social Science Research Network. Retrieved from: https://papers.ssrn.com/abstract=2309065.

Vogler, D., Eisenegger, M., Schneider, J., Hauser, L., & Udris, L. (2019). Qualität von Schweizer Informationsmedien im Zeitverlauf [engl.: Quality of Suisse information media over time]. In fög (Ed.), *Qualität der Medien Jahrbuch 2019* [engl.: Quality of media yearbook 2016] (pp. 85-96). Basel: Schwabe.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*, 1146–1151. doi: 10.1126/science.aap9559

W

Waddell, T. F. (2018). What does the crowd think? How online comments and popularity metrics affect news credibility and issue importance. *New Media & Society*, *20*(8), 3068–3083. doi: 10.1177/1461444817742905

Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, *85*(3), 423–441. doi: 10.1080/03637751.2018.1467564

Wardle, C., & Derakhshan, H. (2017). *Information Disorder. Toward an interdisciplinary framework for research and policymaking*. Straßbourg: Council of Europe. Retrieved from: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Weber, P., Mangold, F., Hofer, M., & Koch, T. (2019). *Meinungsbildung in der Netzöffentlichkeit [engl.: Opinion formation in the net public]*. Baden-Baden: Nomos.

Webster, J. G. (2005). Beneath the veneer of fragmentation: Television audience polarization in a multichannel world. *Journal of Communication*, *55*(2), 366–382. doi: 10.1111/j.1460-2466.2005.tb02677.x

Webster, J. G. (2010). User information regimes: How social media shape patterns of consumption. *Northwestern University Law Review*, *104*, 593–612.

Webster, J. G., & Ksiazek, T. B. (2012). The dynamics of audience fragmentation: Public attention in an age of digital media. *Journal of Communication*, *62*(1), 39–56. doi: 10.1111/j.1460-2466.2011.01616.x

Welbers, K., & Opgenhaffen, M. (2019). Presenting News on Social Media: Media Logic in the communication style of Newspapers on Facebook. *Digital Journalism*, *7*(1), 45–62. doi: 10.1080/21670811.2018.1493939

Williams, H. T. P., McMurray, J. R., Kurz, T., & Lambert, F. H. (2015). Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change*, *32*, 126–138. doi: 10.1016/j.gloenvcha.2015.03.006

Wollebæk, D., Karlsen, R., Steen-Johnsen, K., & Enjolras, B. (2019). Anger, Fear, and Echo Chambers: The Emotional Basis for Online Behavior. *Social Media & Society*, *5*(2), 1-14. doi: 10.1177/2056305119829859

Y

Yang, J. H., Rojas, H., Wojcieszak, M., Aalberg, T., Coen, S., Curran, J., ..., Tiffen, R. (2016). Why Are "Others" So Polarized? Perceived Political Polarization and Media Use in 10 Countries. *Journal of Computer-Mediated Communication*, *21*, 349-367. doi: 10.1111/jcc4.12166

YouTube (n.d.). *Entfernungen von Inhalten aus YouTube auf der Grundlage des NetzDG - 01. Januar 2019 – 30. Juni 2019 [engl.: Removal of content from YouTube on the basis of the Network Enforcement Act – January 1st 2019 – June 30th 2019]*. Retrieved from: https://transparencyreport.google.com/netzdg/youtube?hl=de

Z

Ziegele, M., Breiner, T., & Quiring, O. (2014). What creates interactivity in online news discussions? An exploratory analysis of discussion factors in user comments on news items. *Journal of Communication*, *64*(6), 1111–1138. doi: 10.1111/jcom.12123

Ziegele, M., & Jost, P. (2016). Not funny? The effects of factual versus sarcastic journalistic responses to uncivil user comments. *Communication Research,* 1–30. doi: 10.1177/0093650216671854

Ziegele, M., Jost, P., Bormann, M., & Heinbach, D. (2018): Journalistic counter-voices in comment sections: Patterns, determinants, and potential consequences of interactive moderation of uncivil user comments. *Studies in Communication and Media, 7*(*4*), 525-554. doi: 10.5771/2192-4007-2018-4-525

Zimmer, A., & Kunow, K. (2019). Intermediäre und Meinungsbildung [engl.: Intermediaries and opinion formation]. In die Medienanstalten (Ed.), *Vielfaltsbericht 2019 der Medienanstalten [engl.: Diversity Report2019 of the Media Institutions]* (pp. 46-59). Berlin: die Medienanstalten – ALM GbR.

Zimmermann, F., & Kohring, M. (2018). „Fake News" als aktuelle Desinformation. Systematische Bestimmung eines heterogenen Begriffs [engl.: „Fake News" as current Disinformation. Systematic Definition of a Heterogeneous Term]. *Medien & Kommunikationswissenschaft*, *66*(4), 526–541. doi: 10.5771/1615-634X-2018-4-526

Zimmermann, F., & Kohring, M. (2020). Mistrust, disinforming news, and vote choice: A panel survey on the origins and consequences of believing disinformation in the 2017 German parliamentary election. *Political Communication*, *37*(2), 215-237. doi: 10.1080/10584609.2019.1686095

Zollo, F. (2019). Dealing with digital misinformation: A polarised context of narratives and tribes. *EFSA Journal*, *17*(S1), 1-15. doi: 10.2903/j.efsa.2019.e170720

Zuiderveen Borgesius, F. J., Trilling, D., Moeller, J., Bodó, B., de Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles?. *Internet Policy Review, 5*(1). doi: 10.14763/2016.1.401

Zuiderveen Borgesius, F.J., Moeller, J., Kruikemeier, S., Ó Fathaigh, R., Irion, K., Dobber, T., Bodo, B., & de Vreese, C. (2018). Online Political Microtargeting: Promises and Threats for Democracy. *Utrecht Law Review*, *14*(1), 82–96. doi: 10.18352/ulr.420

Are Algorithms a Threat to Democracy?
The Rise of Intermediaries:
A Challenge for Public Discourse

Professor Dr. Birgit Stark and Daniel Stegmann, M.A.
with Melanie Magin, Assoc. Prof. & Dr. Pascal Jürgens

26 May 2020

Publisher:
AW AlgorithmWatch gGmbH
Linienstraße 13
10178 Berlin
Germany
Contact: info@algorithmwatch.org

Proofreading:
Graham Holliday

Layout:
Beate Autering, beworx.de

Published as part of the research project Governing Platforms
Website: algorithmwatch.org/en/governingplatforms

A project by

ALGORITHM WATCH

in partnership with

EPC EUROPEAN POLICY CENTRE

MAINZER MEDIENINSTITUT

with support by

civitates

The sole responsibility for the content lies with the author(s) and the content may not necessarily reflect the positions of Network of European Foundations NEF, Civitates, or the Partner Foundations.