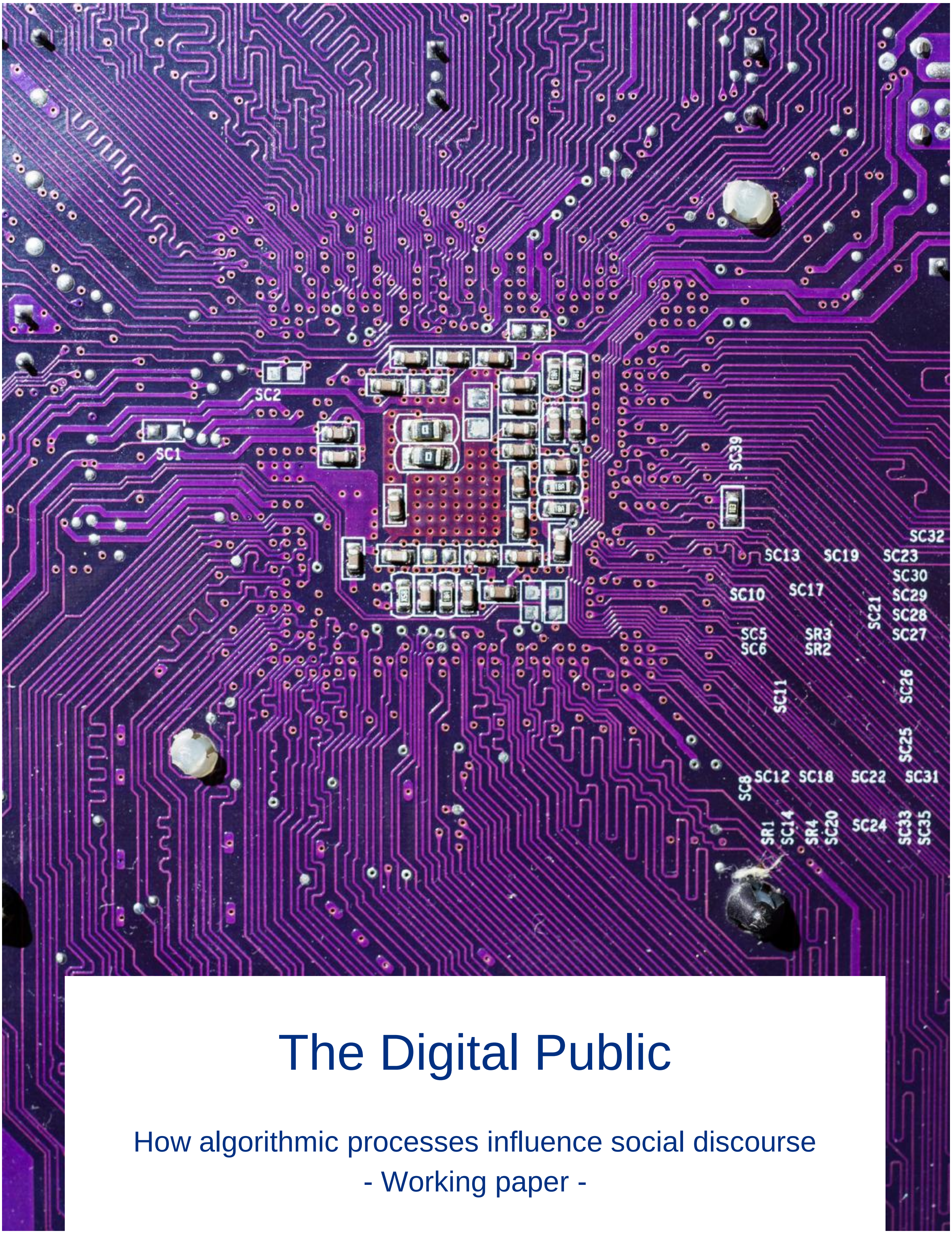
Discussion paper Ethics of Algorithms #3



The Digital Public

How algorithmic processes influence social discourse
- Working paper -

BertelsmannStiftung

The Digital Public

How algorithmic processes influence social discourse
– Working paper –

Konrad Lischka
Prof. Dr. Christian Stöcker
on behalf of Bertelsmann Stiftung

Imprint
© January 2018
Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
33311 Gütersloh
Germany
www.bertelsmann-stiftung.de

Responsible for content
Konrad Lischka, Ralph Müller-Eiselt

Authors
Konrad Lischka, Prof. Dr. Christian Stöcker

DOI 10.11586/2017049 https://doi.org/10.11586/2017049

Contents

1 Preface

A post goes viral on Facebook: The recipient of the Nobel Peace Prize has been chosen, the latest election polls have been released, a government minister has resigned – or the like. The more people who read this post or click the “like” button, the more relevant it must be for a broader audience. And the more people who choose to hide the post because they want to be exposed to fewer stories of this nature, the less relevant the subject must be for the broad majority of users. Or perhaps not?

The following is undoubtedly how most people would intuitively interpret the logic behind social networks: The more positive responses a post gets, the higher up it should be placed in the News Feed for the greatest number of users. And, conversely, the more negative responses it gets, the less prominent its placement should be. In fact, this is more or less how Facebook once deployed user reactions to judge a story’s relevance – until 2015, when someone took a closer look and ascertained that 5% of all Facebook users were responsible for 85% of all the posts blocked using the “hide” feature. These *super hiders* were somewhat of a mystery. They removed almost everything that appeared in their News Feed, even posts that they had commented on only a short time before. Were they really so dissatisfied with the algorithmically curated content that Facebook was showing them?

As it turns out, the super hiders were not at all dissatisfied, a fact subsequently revealed by a survey. They were using the “hide” feature not as it was designed, i.e. to express antipathy, but simply to organize their posts the way other people keep their in-box tidy by deleting e-mail. Once this fact came to light, Facebook changed the way it assessed relevance, no longer assuming that when users hide a post they are necessarily making a strong statement of dislike.

This story is a prime example of three developments examined in this working paper:

- Algorithmic processes influence social discourse, for example by prioritizing posts in social networks and search engine results, thereby structuring public opinion.
- Algorithmic systems also interpret the reactions people have to an item as a sign of the item’s relevance.
- Human perceptions of media content and an assessment of the content’s relevance is, depending on context, subject to cognitive distortions and therefore difficult to ascertain.

The algorithmic processes used by social networks and search engines influence which content supplied by editorial media is seen in the digital sphere by more than half of all Internet users in Germany and how they perceive it (see Section 2.2). The formation of public opinion is “no longer conceivable without intermediaries,” as Schmidt et al. (2017: 98) put it.

These developments offer opportunities and pose risks. Thanks to algorithmic decision-making (ADM), user reactions and interests are playing an ever-greater role in the public sphere. This can result in new topics and in media content provided by previously unknown individuals or organizations being disseminated and finding a broad audience more quickly. Thus, channels such as social media can also influence editorial media, since journalists use them as sources for their reportage and as a sign of the public’s interest. One risk is that people will incorrectly assume a post is representative of general opinion or unheedingly equate popularity with relevance. Because of design flaws, algorithmic processes can also systematically distort predictions of relevance, as the example of the super hiders demonstrates.

This working paper provides an overview of the known interplay between how intermediaries such as Google and Facebook are designed and the psychological mechanisms that determine users’ perceptions and behavior. To begin, the area of analysis is defined based on empiric data documenting media use, i.e. it is limited to those social networks and search engines that play a particularly decisive role in shaping public opinion (Chapter 2). Chapter 4 outlines key aspects of the structural changes affecting public opinion in the digital realm, i.e. the “bottleneck” is not the publishing of content per se, but the degree to which it receives attention. Based on this, the subsequent analysis in Chapter 5 examines the significance such processes have for public perception, i.e. what do intermediaries who employ ADM processes do differently? Chapter 6 looks at how these developments relate to the guiding principles traditionally used by editorially curated media to form public opinion. Chapter 7 outlines possibilities for intervention that can reconcile the impacts of these structural changes on public opinion with the aforementioned guiding principles.

This working paper is part of an exploration of the topic “Participation in the Age of Algorithms and Big Data” which the Bertelsmann Stiftung is using to examine how phenomena taking place in the digital sphere are influencing social participation. Previously published papers in this series include an analysis of international case studies depicting the use of ADM processes in areas such as law enforcement and education (Lischka and Klingel 2017) and a proposal for determining the potential impact of ADM processes on participation (Vieth and Wagner 2017). As we see it, social discourse is relevant for participation, since the term “participation” is understood in this paper to denote equal inclusion of individuals and organizations in political decision-making and will-building processes and the fair inclusion of all in social, cultural and economic developments. Moreover, will-building includes public opinion as formed through the media in their various forms. We are dedicating an entire working paper to this topic since the interdependencies are complex and ADM processes play a role in the everyday lives of a relatively large number of people in Germany.

This analysis is our first tentative contribution to this debate. We are publishing it as a working paper under a free license in order to add to a rapidly developing field in a way that others can easily build upon.

The manner in which the forums are structured that a society uses to enter into discourse shape the discourse itself. Winston Churchill expressed this concisely in October 1943 when, addressing Parliament on the subject of rebuilding the House of Commons, he called for an architectural design that would have the governing and opposition parties facing each other. “We shape our buildings and afterwards our buildings shape us” (Churchill 1943: 403). Instead of parliaments, online platforms and social media are now having an increasing impact on public discourse. Thus, this paper provides insight into the architecture of the digital sphere.

R. Muller-Eiselt   Konrad Lischka

Ralph Müller-	**Konrad Lischka Lischka**
Senoir Expert	Project Manager
Taskforce Digitization	Taskforce Digitization
Bertelsmann Stiftung	Bertelsmann Stiftung

2 Introduction

This working paper examines the impacts that algorithmic decision-making (ADM) systems could have on public discourse and public opinion in the future. **Chapter 3 ("What it's about")** shows that ADM processes already play a significant role in the media content many people come in contact with every day. Intermediaries such as search engines and social networks provide a growing part of the population with information that, just a few years ago, was supplied by traditional media outlets. Over 57% of Internet users in Germany keep up with current events through search engines and social networks. In the United States, 44% of adults now use Facebook as a regular source of news.

Chapter 4 ("What's changing") shows that both search engines and social networks sometimes deploy other criteria when ranking and making content available than journalists do who work for traditional publications. The ranking itself is determined by a complex interplay of user behavior and algorithmic systems: Users interact with potential content that has already been pre-selected by algorithms, doing so in a way that is often fast and emotional and that more or less lacks any true discernment. Moreover, the users' behavior is potentially geared toward objectives that are not related to gaining information. Users on social networks, for example, are often concerned with managing their identities; in other words, they do not share posts primarily because they want to make content available to other people, but because the shared information reflects their self-definition or signals their desire to be seen as a member of a specific group.

Chapter 5 ("Where it's leading") presents empiric findings and the factors having an impact in light of the aforementioned structure and the way intermediaries are currently shaping public discourse. These factors potentially promote polarization: On the one hand, they include psychological factors affecting people's perceptions and behaviors, for example the tendency users have to share news that they themselves have not read, and certain cognitive distortions such as the availability heuristic. On the other hand, the intermediaries' decisions and activities also play an essential role, especially their efforts to optimize their products in order to ensure a maximum of user interaction. The intermediaries' platforms are designed to foster a certain type of cognition, one that increases the probability of fast and unconsidered behavior. The combination of these factors, coupled with the disempowerment of traditional gatekeepers and the ubiquitous possibilities for global publication, apparently leads to tunnel vision, at least among some people, when it comes to their medial world view. People who already have extremist attitudes can use such platforms to withdraw into the company of like-minded individuals, ignoring information that could challenge their set opinions and even feeling more inclined to maintain their beliefs as they actively reject discordant ideas. This tendency is reinforced by ADM systems. A further factor, whose influence can now be proven but which cannot yet be quantified reliably, is the manipulation of relevance-signaling events through clandestine technical interventions such as the use of so-called bots. The interplay of psychological variables, intermediaries' design decisions and external distortions of the signals measured by intermediaries is apparently leading to increased polarization, at least among some users, both in terms of how they perceive content and their social and political views. It seems, however, that this process of polarization also depends on a series of other factors unrelated to ADM systems, such as how old users are and the political systems found in different countries.

The new public discourse arising from use of these platforms does not adhere to the same criteria as the one produced by traditional media, as **Chapter 6 ("How things should be")** illustrates. The new methods and media do not grant the same pivotal role to traditional social principles such as integration and respect for the truth.

Chapter 7 ("What we can do") presents initial solutions for the outlined problems which can be applied on the micro, meso and macro level. One promising approach would be to work directly with the developers and operators of platforms that use ADM systems. These efforts could, for example, make it easier to research and evaluate such systems from the outside; they could also introduce guiding principles that go beyond the ones intermediaries have previously adhered to.

Responses are also conceivable on all three levels regarding users' perceptions within the outlined processes. One activity that would undoubtedly prove beneficial is increasing awareness among the general public of the existence of ADM systems, not to mention the impact the systems can have and how they can affect user behavior. Another helpful response, especially as an antidote to the manipulation of public opinion through the willful spreading of disinformation, would be the dissemination of detailed information about the current situation and the strategies and methods employed by purveyors of disinformation. It might also make sense on the macro level to invest in independent research, with the explicit goal of producing academic studies that can shed light on and rectify methods of cognition that are intellectually less demanding and, thus, less suitable for social discourse. Examining user interfaces and the systems designed to engage and reward users, these studies would identify cognition styles that are profound, exacting and less susceptible to error.

3 What it’s about: Algorithmic processes are influencing public discourse

3.1 Introduction: Captain Kirk and the anti-vaxxers

The man that most people know as Captain Kirk provided a telling exampling in early 2017 of how the media and communications landscape has changed in the last decade. Actor William Shatner portrayed Kirk in the television series *Star Trek* and his character liked to squabble with the starship’s chief medical officer “Bones” McCoy. In 2017, Shatner got into a very public flap with another doctor, this time a real one. The way this public dispute played out and its outcome highlight an entire series of factors which demand that, in the age of digital communication, social participation be reconsidered and safeguarded.

The disagreement began after Shatner posted a tweet on Twitter, the US-based communications platform (Levinovitz 2017). He endorsed *Autism Speaks*, an organization founded in the US in 2005 to help individuals affected by autism, and their family members, to respond to the condition. According to the organization, it advocates for people with autism; in the past, however, it had been subject to criticism and calls to boycott it. For numerous reasons, many people with autism do not feel Autism Speaks represents them. Moreover, for a long time the organization maintained that autism is caused by vaccines, an assertion for which there is no scientific evidence.

Shatner has 2.5 million followers on Twitter, an audience, in other words, that could only be reached a few years ago by professional media organizations. What he says on the platform gains currency, if only because so many people read it. His Autism Speaks tweet received a great deal of attention, especially from people who are highly critical of the organization.

Shatner has been active for quite a while on behalf of people with autism and does not at all subscribe to the conspiracy theories advanced by those who oppose vaccines. He vehemently denied the accusation that he supports a questionable organization. Tweets on Twitter were, at the time, limited to 140 characters and, because of the abridged nature of the messages, the risk of a misunderstanding is considerable. If one does occur, it can, in turn, lead to heated debate. Exactly that is what happened here.

As tempers flared, a medical professional got involved who had already engaged at length with the conspiracy theories surrounding vaccines. David Gorski, an oncologist at Wayne State University in Detroit, writes for a number of blogs including *Science Based Medicine*, which advocates for evidence-based interventions and against pseudo-scientific assertions. Gorski thus has many enemies among those dogmatically opposed to vaccines.

In several tweets directed at Shatner, Gorski explained – politely, in detail and in a way that was completely open to the public – why Autism Speaks is a controversial organization. Shatner reacted as many in this type of situation would: He typed Gorski’s name into a search engine.

He then began to share with abandon what he found with his 2.5 million Twitter followers – for example an article about Gorski taken from a highly suspect website called *TruthWiki*, which is run by a vaccine opponent and conspiracy theorist. In the article, Gorski is disparaged as a “a paid shill of the powerful vaccine industry” who represents a “pseudo-scientific religion” and who has a propensity for “insane rants.” Shatner shared another article he found online which accuses the oncologist of of deliberately promoting cancer-causing medical interventions because he financially benefits from cancer patients”. Additional links to false, presumably litigable claims about the medical practitioner followed.

Gorski, who endorses science-based methods and refutes conspiracy theories, which can prove lethal, especially in cases like these, was presented to an audience of millions as a malicious, avaricious charlatan.

Many users supported Gorski. One asked Shatner why he had not read and linked Gorski’s Wikipedia entry, which correctly lays out the facts. Shatner answered that *TruthWiki* was higher up in his Google search results. You can find it “All on Google,” the actor maintained, as if that itself was a sign of quality.

The pages linked in Shatner’s tweets were all published by right-wing conspiracy theorists; a number of the sites were even created by the same person. A willingness to reject evidence-based procedures, even chemotherapy, and simultaneously advance right-wing ideas is remarkably widespread in certain circles in the United States.

Days later Shatner was still using his Twitter account to spread disparaging remarks about Gorski and he received robust, ongoing support from users who militantly oppose vaccines. Yet neither is the actor, by his own admission, against vaccines, nor does he openly subscribe to conspiracy theories. He simply felt he had been attacked and was punching back.

The case is a prime example of the many factors that have led to a radical change in how public discourse is conducted and public opinion formed.

- On today’s communications platforms, the traditional *boundary between personal and public communication has disappeared.*
- As a result, individuals, in this case William Shatner, can suddenly interact with an audience of millions. In this context, communications specialists talk about *the disempowerment of traditional gatekeepers* – journalists, press offices, etc.
- Individuals have access to completely new and powerful, but by no means infallible, methods for *locating information in the blink of an eye,* for example by using search engines such as Google.
- These tools adhere to certain *algorithmic parameters.*
- *The criteria that algorithms use to measure relevance* often *do not correspond to the criteria* that reputable journalists or researchers would use.
- In many cases the algorithms work *descriptively,* for example they show which links were clicked particularly often in the past. Yet many users *assume the results are normative* (“higher up in the Google results”).
- In the case of certain extremely controversial issues such as the false assertion that there is a correlation between autism and vaccines, small but highly motivated user groups can, by virtue of their online activities, ensure that a *massive divergence occurs between content quality and “relevance” as determined by algorithms.*
- The way that communication takes place on digital channels can lead to a *rapid escalation of what are actually trivial conflicts,* even when, in terms of content, no de facto disagreement exists between participants.

The above example took place in the United States. Some of the factors and mechanisms that gave rise to it have not played a significantly equivalent role in Germany until now. For example, just over 2% of all Internet users in Germany get their information from Twitter on a regular basis (Hasebrink, Schmidt and Merten 2016). In contrast, many others in Germany, as elsewhere, are actively creating a new form of public discourse.

This paper has thus been written to shed light on a number of fundamental questions relating to the social relevance for the public discourse of *algorithmic decision-making (ADM)* processes.

How are ADM processes changing public discourse? How can and should public discourse take place in a democratic society increasingly beholden to such processes? Do interventions exist and, if so, what are they?

First, however, a brief summary will be given based on currently available empiric data. The summary illustrates the impact that intermediaries and ADM processes are now having on public opinion and public discourse.

3.2 Intermediaries that are particularly relevant for the formation of public opinion

If one is searching for an image that immediately reveals how digitization has changed public life in the last 10 to 15 years, pictures of people riding the subway anywhere in the industrialized world would be a good choice. Photographs of rush hour from the year 2006 show a complicated dance, i.e. commuters trying to avoid disturbing the person next to them, since every other passenger has a newspaper in his or her hand. Back then, folding and holding the paper to ensure it did not invade the space of one’s fellow riders was an art that people practiced daily.

Photographs from 2016, 10 years later, reveal something completely different, even if most passengers were still concentrating on media content: People were staring at smartphones, regardless of whether the photograph had been taken in the subway in New York, Berlin or Tokyo. The year that Apple presented its first iPhone, 2007, marks a watershed in the history of the human race. Even though mobile devices already existed that could receive e-mail and access specially designed websites, only once the iPhone appeared was a new standardized format available that made it possible for people to interact with the new digital universe while they were on the go. This innovative format was a rectangular device, usually held upright, whose front consisted entirely of a touch-sensitive screen and which was outfitted with at least one camera and a number of sensors allowing the device to orient itself spatially, among other useful functions. The smartphone is the end point – for now – of a development that the British cultural critic Raymond Williams foretold with amazing prescience in 1974 using the phrase *“mobile privatization”* (Williams 1974). The term described the increasing personalization of social experiences that Williams believed would occur as the result of technological advancement. The appearance of the Sony Walkman n 1979 proved him right: Suddenly it was possible for people to move about in public within their own acoustic universe, adding a soundtrack to their everyday lives.

With the smartphone, considerable progress has been made in providing people with their own totally individualized media and communications environment. There is a fundamental difference between the newspaper perusers in the subways of 2006 and the smartphone users found there in 2016: In the case of the former, one glance is enough to reveal what they are doing; in the case of the latter, a glance reveals virtually nothing. Are they reading an article on a newspaper or magazine website? An entry in a professional journal? A comment posted on social media by a friend? Or by a celebrity such as William Shatner? A professional or personal e-mail? Are they in the process of purchasing something? Are they looking for an audio book or a song, or are they assembling a playlist to listen to through their earbuds during the rest of their journey? Are they playing a game? This list could be extended indefinitely, and each month brings even more uses for these mobile computers. At the same time, the media and communications experiences made possible by smartphones are personalized using another method, one already used back when the World Wide Web had yet to go mobile: algorithmic filtering.

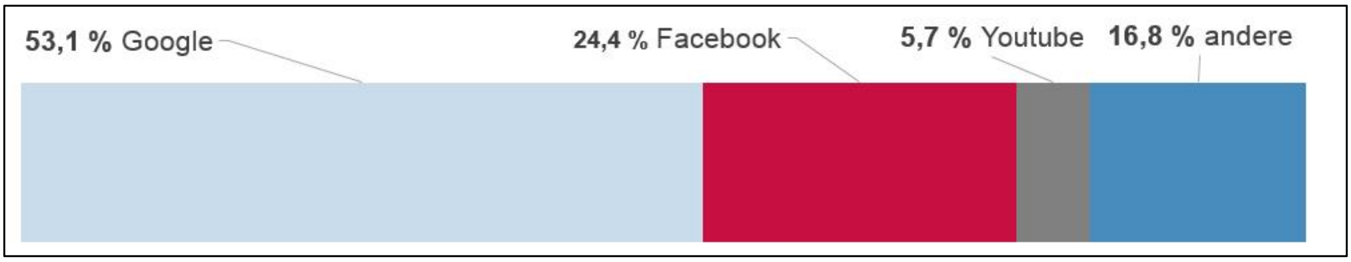
In Germany, *“Intermediäre” (intermediaries)* is the term now commonly used to refer to sites that share and filter information (Hasebrink, Schmidt and Merten 2016; Schmidt et al. 2017). Researchers at the Hamburg-based Hans-Bredow-Institut use this term to refer to search engines such as Google and Bing and video sites such as YouTube, along with more visually oriented communications platforms such as Snapchat and Instagram, and more traditional instant messaging services such as WhatsApp. According to statistics from TNS Infratest, over 57% of all Internet users in Germany regularly use such intermediaries to keep abreast of current developments – in other words, not just to chat with friends or consume entertaining content. Search engines are the most popular intermediaries, used by almost 40% of the German population, followed by social networks such as Facebook, at over 30%, and video sites such as YouTube, at over 9%. According to the statistics, 8.5% of users avail themselves of instant messaging services to stay informed, services in which ADM processes currently do not play a role. In terms of serving as a source of information, there are clear market leaders among the intermediaries, namely Google among search engines (37.9%) and Facebook among social networks (24.1%). In comparison, Microsoft’s search engine, Bing, is only used by 2% of Germany’s online surfers and Twitter by 2.1% of its social networkers to stay up to date. Overall, more than 54% of those people who get their news from the

Internet have regularly used search engines or social media to view content provided by traditional sources such as the websites operated by newspapers, magazines and TV broadcasters.

If one expands the focus beyond news-seeking purposes, the picture becomes much clearer: Over 95% of all German Internet users access at least one intermediary every day (for any number of reasons), with Google the clear leader (78.6%) followed by YouTube (42%) and Facebook (41.8%). In terms of instant messaging services, which play a subordinate role in the subject under examination here, the Facebook subsidiary WhatsApp leads the field, with almost 75% of all Internet users in Germany using the application every day.

If one asks users which products and services are particularly important to them when they are looking for news and information, the results are similar (Ecke 2016), see also Figure 1.

Figure 1: The most important intermediaries for information, according to German users



Figures in percent; daily reach = use yesterday. Base: 23.252 million Internet users 14 years or older in Germany who used at least one intermediary yesterday, n=669. Question: “You used the following as a source of or access to information on current events in the areas of politics, business and culture. Which of these sources is the most important for you?”
Source: Kantar TNS. Berlin, in Ecke, 2016: 21.

Based on these and other data, Hasebrink et al. (2016) arrived at the conclusion in 2016 that “processes for forming public opinion are no longer conceivable without intermediaries,” even though the latter are “only one element in the process of forming public opinion.” A person’s own social environment “remains important” as does the “reportage done by trusted journalistic media” (ibid.).

It must be noted that, as in almost every area affected by digitization, the figures cited here reflect the situation at only that point in time in 2016 when the data were collected; the number of users changes constantly, as does the way they use the corresponding products and services. At the beginning of 2016 in the US, where developments in this area are always slightly more advanced than in Germany, 44% of all adults were using Facebook as a news source more often than just “rarely” (Gottfried and Shearer, 2016).

Moreover, it is true on both sides of the Atlantic that the younger the surveyed target group, the higher the percentage of those who keep up to date using social networks, video platforms and search engines – and the lower the percentage of those who make use of traditional media such as daily newspapers (Mitchell et al. 2016).

A second important development is the increasing significance of social networks as a point of entry and facilitator of contacts for other content providers. US-based Parse.ly, which offers website operators analytic tools and therefore has access to traffic statistics for thousands of providers, also evaluates annually how many visitors come to the analyzed sites and from which locations. At the end of 2012, Google was the clear leader, accounting for over 40% of all traffic directed to the analyzed sites; Facebook’s share was approximately 10%. In mid-2015, Facebook surpassed Google for the first time, with both having shares of just under 40%.

People who use such intermediaries, excluding instant messengers, in order to access information always come in contact with ADM processes (Gillespie 2014). The posts shown in a Facebook News Feed, whether they originate from the user’s friends or from media sources whose Facebook pages are being followed by the user, are automatically ranked according to certain criteria. The same is true for search engines, i.e. the results returned by every Google query have been sorted according to specific criteria. What is surprising is that many users apparently still do not know that.

When researchers from California asked Facebook users in 2015 how the order of the posts on their own News Feed was determined, over 62% said that all of the posts from their friends and all the accounts they follow on Facebook are shown. Some suspected that something else might also be happening, but they had no idea what. One respondent was quoted thus: "I have like 900 and some friends and I feel like I only see 30 of them in my News Feed. So I know that there's something going on, I just don't know what it is exactly." (Eslami et al. 2015).

In sum it can be said that, for a growing part of the population in industrial nations, intermediaries – above all search engines and social networks – are a key venue for gaining and absorbing information relevant to social participation. This is all the more applicable the younger the cohort is. Although traditional media channels, especially television, continue to play a major role, developments in recent years clearly show that intermediaries and thus ADM processes are becoming increasingly important for the dissemination of information. Yet many users remain unaware that ADM processes are at work.

Recent developments have shown that even the largest of these new intermediaries, Facebook and Google, are thinking about the growing responsibility they bear. After an extensive debate on the role that *fake news* spread by Facebook might have had in influencing the 2016 US presidential election, Facebook CEO Mark Zuckerberg, who had consistently denied that his company had had any effect, said: "We don't want any hoaxes on Facebook. Our goal is to show people the content they will find most meaningful, and people want accurate news. We have already launched work enabling our community to flag hoaxes and fake news, and there is more we can do here" (Zuckerberg 2016).

On January 31, 2017 this vague announcement was followed by a more detailed one, namely that Facebook would be "incorporating new signals to better identify and rank authentic content" (Lada, Li and Ding 2017). At Facebook, the idea has thus clearly been taken onboard that the company's algorithms can be used to form public opinion, a recognition that has become apparent in the form of technical changes.

At Google, such quality assurance mechanisms have been in effect for a while. Evaluators, known as *quality raters,* regularly input predefined queries into the company's search engine and then use a series of criteria to assess the results. The handbook for these raters also includes a section on so-called *Your Money or Your Life (YMYL) pages* (Google 2017). According to the handbook, the pages in this category include those that could "potentially impact the future happiness, health, or financial stability of users." In addition to pages containing financial or medical information, this category explicitly includes news articles or "public/official information pages important for having an informed citizenry" and "webpages that include information about local/state/national government processes, people, and laws" and "news about important topics such as international events, business, politics, science, and technology." The raters are called upon to "please use your judgment and knowledge of your locale" (ibid.).

The major intermediaries are thus already behaving in a way which reflects the assumption that outcomes from their systems and processes have an effect on information in the public sphere. Before we turn to the question of what the empiric data informing this assumption look like, we must clarify the terms and concepts used here.

3.3 Conceptual background

This paper examines providers in the digital sphere that have an impact on social discourse and public opinion, or more precisely, providers that make use of ADM processes. Websites that are particularly relevant for forming public opinion can be classified according to user behavior (see Section 3.2). German Internet users cite Google, Facebook, YouTube, WhatsApp and Twitter, in that order, as their most common sources of information, direct or indirect, for current events in the political, business and cultural spheres (Ecke 2016: 21). We will therefore consider these sites here and will identify the overarching design principles relevant to our analysis.

3.3.1 Intermediaries and their design principles

In the German debate, *“Intermediäre” (intermediaries)* has become the general term now commonly used to refer to these providers. *“Informationsintermediären” (information intermediaries)* is sometimes also used to denote those providers or services that play an important role in the social discourse. The term was coined by Schulz and Dankert (2016) although it “has neither clearly defined contours, nor any basis in theory” (ibid.: 13). The term is sufficient for the tentative analysis offered here. What is meant are “digital services … that play a mediating role between users and content” (ibid.: 19). Mediation is a design principle shared by all five information intermediaries in Germany most often cited by users, as seen in Table 1.

Table 1: Design principles of key intermediaries in Germany

Characteristic	Google (search)	Facebook	YouTube	WhatsApp	Twitter
Mediation (between public and third-party services; functions such as finding and aggregating information)	yes	yes	yes	yes	yes
Delivery (complete third-party content is delivered using the intermediary’s infrastructure)	no (excerpts)	yes (partially)	yes	yes	yes (partially)
Structuring (content from third parties is ranked according to relevance determined by intermediary)	yes	yes	yes	no	yes (optional)
Algorithmic decision-making (for determining relevance and selecting displayed content)	yes	yes	yes	no	yes (optional)
Social network (users define their own identity and relationships)	no	yes	yes	yes	yes

Source: The authors.

The other design principles apply to the five providers to varying degrees. The following provides a short discussion of the differences and, derived from that, a definition of intermediaries as used in this paper.

Delivery: Most services use both models. On the one hand, they merely link to content released elsewhere by third parties; on the other hand, they provide third parties with the infrastructure for publishing and disseminating content. Facebook, for example, allows links, but also the publication of texts, videos and photos using the Facebook infrastructure; the same is true of Twitter (although only for certain elements such as photos and videos). Complex hybrid forms also exist, such as Google’s *Accelerated Mobile Pages.* In the German discussion, services with their own infrastructure for storing and delivering content are often referred to as platforms (see for example Schmidt et al. 2017: 9). Ultimately this criterion is not decisive for the scope of this paper: All providers serve as intermediaries for all content, regardless of the infrastructure the content is made available on.

Structuring public discourse and the use of *ADM processes* are a common characteristic of all providers, with the exception of the instant messaging service WhatsApp. WhatsApp employs a structuring logic that Facebook and Twitter gave up years ago: When users on WhatsApp establish a link to another user, they can see all of the latter’s posts. This makes WhatsApp a special case. Its service more resembles applications designed purely for interpersonal communication such as e-mail. In their current form, providers such as WhatsApp are not relevant as information intermediaries for the purposes of this paper. They mediate between users and third parties, but structure public discourse solely according to users’ chosen signals (contact / no contact). This hardly differs from the principle underlying a mailing list or forum. In contrast, we are focusing here on providers who use ADM for structuring purposes, i.e. for “de- and re-grouping of information” (ibid.: 20). Personalization is in most cases a form of re-grouping, even if, unlike others such as Schmidt (ibid.: 20), we do not see personalization as an organizational principle necessarily inherent to informational intermediaries. A list of the day’s 20 most important topics created using an algorithm based on user behavior is relevant, even if it is not personalized. If all of a provider’s users who speak the same language see this selection of topics on the website’s landing page, then that too influences public opinion.

Social network: In this paper, we do not use this design principle to define intermediaries. Most services encourage users to create an individual profile and to link to a network of friends that can be evaluated using algorithms. And while these signals are helpful for structuring public discourse using ADM, they are not a prerequisite for doing so. This becomes apparent when one looks at Twitter and at Google searches, which offer non-registered users a structured selection based on readily available characteristics (language, location, etc.) and evaluations of other users’ behavior.

Definition: Intermediaries who use algorithms to structure public opinion

For the purposes of this paper, three characteristics distinguish intermediaries (after Perset 2010: 9):

(1) They mediate between third parties whose interaction gives rise to public discourse and public opinion. The third parties can be private individuals, journalists or editors at media outlets, representatives of business organizations, policy makers and public administrators.

(2) They disseminate and/or make available content created by third parties. In doing so, they re-group the content based on principles they have developed for determining relevance. They define the conditions for accessing content and the matching mechanisms.

(3) They use ADM processes for assessing relevance and selecting displayed content.

3.3.2 ADM processes

We use the term *algorithmic decision-making (ADM) processes* as Zweig (2017) defines it, namely for automated processes that:

- use an algorithm to evaluate a situation or a person or to predict the probability that a situation will occur, and
- subsequently react aided by the (sometimes indirect) activation of an element, and thus directly affect the lives of human beings.

As commonly used elsewhere, we also employ “ADM process” as the shorthand for this phenomenon.[1] Intermediaries who use ADM processes to structure public opinion predict the potential demand for certain content and adjust their offerings to reflect that prediction (Napoli, 2014, S. 34). This selection determines which aspects of the public discourse people are confronted with when they use websites such as Google and Facebook.

The selection occurs de facto in a far more complex manner. As a rule, multiple ADM-driven programs interact and reference various parameters – so-called blacklists, for instance – during the decision-making process. One example would be the *module used by the German Federal Review Board for Media Harmful to Minors* (BPjM n.d.), which the board provides to intermediaries such as Google so results can be generated in which content inappropriate for minors does not appear at all.

For the purposes of this paper, the algorithmic portion of the decision-making process is of primary importance.

[1] The term is used, for example, in USACM 2017; Ananny and Crawford 2016; Goodman and Flaxman 2016; Mittelstadt 2016a; and Zarsky 2016.

3.3.3 Participation

“Participation” as used in this paper denotes the equal inclusion of individuals and organizations in political-decision and will-building processes, and the fair inclusion of all in social, cultural and economic developments. This means, first, participation in democratic processes – i.e. political equality – and, second, participation in society’s achievements: everything from “suitable living and housing conditions, social and health protection, adequate and universally accessible educational opportunities and inclusion in the labor market to various opportunities for spending one’s leisure time and determining how one lives one’s life” (Beirat Integration 2013).

One prerequisite of participation in this sense is that the material resources available to all are above the minimum level required for ensuring everyone can be part of society. The guarantee of social and political participation thus presupposes a “basic equality of social necessities” (Meyer 2016). Elements of this basic equality are described, for example, in the Universal Declaration of Human Rights and in the International Covenant on Economic, Social and Cultural Rights (Bundesgesetzblatt 1966). Targeted investments in the development of individual skills are necessary to enable equal participation in this sense (Bertelsmann Stiftung 2011: 31). It is the responsibility of the state and of the community to continually empower each individual to take advantage of the opportunities available to him or her.

4 What's changing: Structural transformation of public discourse

4.1 Key aspects of the structural change

The intermediaries in Germany who are especially relevant when it comes to public discourse all use ADM to personalize the selection and structuring of content. Facebook has been personalizing its News Feed since September 2006 (Facebook 2006). Google has been doing the same since 2009 worldwide, even for users who are not logged in (Horling and Bryant 2009). It has been the practice on Twitter since the beginning of 2015 (Rosania 2015), if users do not explicitly opt out (Twitter 2017). The principles (see Section 3.3.1) used by these intermediaries are leading to a structural change in public discourse. The following summarizes the opportunities and risks for six aspects relating to this trend that are of key importance for this paper.

Table 2: Key aspects of the structural change transforming public discourse

Aspect	Description	Opportunities	Risks
Disconnect between publication and reach	Anyone can publish what they want, even if not everyone finds an audience. Content gains attention only after people interact with ADM processes (Dobusch 2017).	Larger, more easily accessible range of media, since gatekeepers do not determine what gets published.	Less diversity because of filtering by a few predominant intermediaries and ADM processes.
"Unbinding" of publications	Reach is determined at the level of each post or article, placement within a publication results in fewer articles on intermediary sites.	Greater diversity of media. Opportunities for gaining attention are more widely spread. Media can subdivide diverse public into segments.	Publications are less frequently seen as a hallmark of quality. Unpopular but relevant issues are less likely to be seen.
Personalization	Users learn more about the areas that interest them.	More in-depth information, greater mobilization.	Filter bubbles Echo chambers
Public's greater impact on reach	User reactions influence overall ADM process and the reach of every post.	Democratization due to public's greater influence. More dynamic agenda.	Narrowing of attention to issues and tonalities favored by ADM processes.
Centralized selection	There are fewer intermediaries than editorially curated media.	More easily accessible, diverse media.	Less diversity due to filters.
Interaction between human and machine-based curation	Editorially curated media consider responses on intermediary sites as a sign of public interest.	Greater chance of dissemination of new, non-established topics and actors in editorially curated media.	Erroneous assumptions about the representative nature of algorithmically ranked statements and actors. Exoneration through inclusion in editorial media.

Without the structural transformation of public discourse outlined above, phenomena such as the rise and fall of upstart political actors, e.g. Germany's Pirates Party, would be largely inconceivable. This means that a new potential for mobilization exists when it comes to organizing movements around specific issues. As Dobusch notes:

> *Thanks to the Internet and digital technologies, voices are heard and seen today, or even arise at all, that were not previously present in the public sphere…. Never before were our societies so open to public pronouncements. Yet it is not only this openness to pronouncements of all sorts; it is also the algorithmically prepared dissemination and reinforcement via platforms such as Twitter, YouTube and Facebook that expands this openness yet again. (Dobusch 2017)*

The comparison of the publication processes below shows the crucial new role played by user reactions and algorithmic processes. Both determine, via intermediaries, how much attention content receives. We are positing the hypothesis that user reactions and ADM processes cannot be clearly arranged in a linear sequence of causality. The following section analyzes in detail how these factors interact.

Figure 2: Organization of algorithmically structured public sphere (after Dobusch 2017)

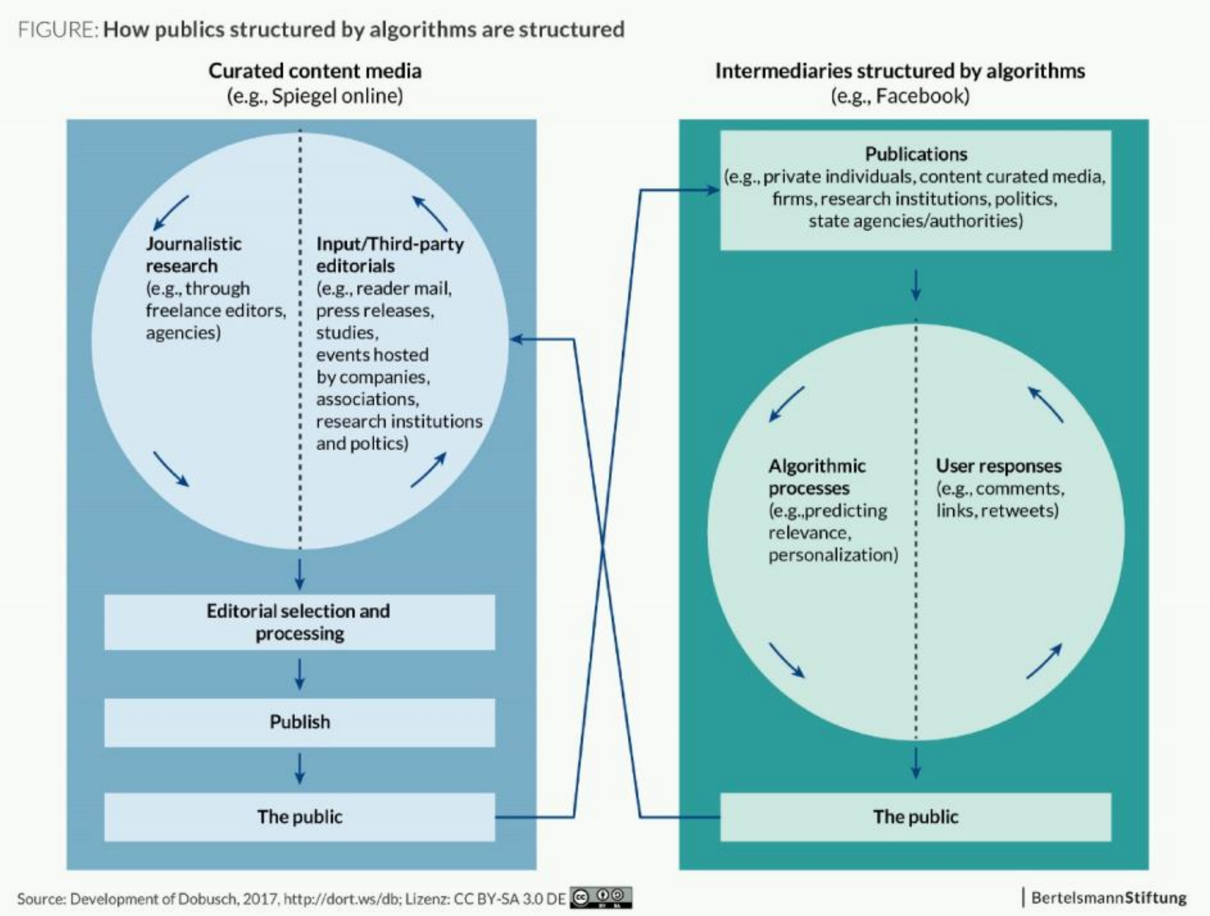


FIGURE: How publics structured by algorithms are structured
Curated content media
(e.g., Spiegel online)
Journalistic research (e.g., through freelance editors, agencies)
Input/Third-party editorials (e.g., reader mail, press releases, studies, events hosted by companies, associations, research institutions and poltics)
Editorial selection and processing
Publish
The public
Intermediaries structured by algorithms
(e.g., Facebook)
Publications (e.g., private individuals, content curated media, firms, research institutions, politics, state agencies/authorities)
Algorithmic processes (e.g.,predicting relevance, personalization)
User responses (e.g., comments, links, retweets)
The public
BertelsmannStiftung

Source: Development of Dobusch, 2017, http://dort.ws/db; Lizenz: CC BY-SA 3.0 DE

4.2 Criteria used by editors to measure relevance

The way that journalists decide what merits reporting and how news and information are weighted has been the subject of intensive debate for almost 100 years and now has a comprehensive body of research dedicated to it. The US-based journalist and media critic Walter Lippmann first defined criteria determining an event's *news value* in 1922 (Lippmann and Noelle-Neumann 1990). In the ensuring decades, numerous sets of criteria were published and discussed at length (e.g. Galtung and Ruge 1965; Schulz 1976). In 2013, Siegfried Weischenberg listed significance and interest on the part of the public as factors creating news value, factors that he then divided into in a series of subfactors (Weischenberg 2001): the extent and consequences of the event (as subfactors of significance); and proximity, prominence of the persons involved, timeliness and human interest (as subfactors of the public's interest).

The manner in which factors are weighted by editors of course depends on the relevant medium's focus and the way it defines itself. According to Weischenberg, significance plays a greater role in *hard news,* and interest on the part of the public is more important for *soft news.* Hard news is defined as including news from the political and business spheres, while soft news includes news about celebrities or other reportage that is at least partly entertaining.

4.3 Criteria used by intermediaries to direct attention

Facebook and Google use a different procedure to determine what is relevant than do journalists, for example. The following section concretizes what these intermediaries actually do in this regard. At first glance it seems trivial, since both organizations manage to describe the gist of their ADM processes in a single sentence:

- "The algorithm looks at your query and uses over 200 signals to decide which of the millions of pages and content are the most relevant answers for that query" (Google 2016).
- "Rather than showing people all possible content, News Feed is designed to show each person on Facebook the content that's most relevant to them" (Boland 2014).

Yet what does "relevant" mean? Relevance is a human concept and thus not directly measurable. In the following, we analyze how two intermediaries, Google and Facebook, measure relevance. We concentrate on examples that demonstrate the interplay of human perceptions and machine-based decisions, from which basic principles can be derived. For that reason, we merely outline some of the factors that Google and Facebook are known to deploy. We use information provided by the two intermediaries and by external researchers. The goal is to gain an impression of the factors relevant for the decision-making process, not to offer a statement about their absolute significance.

4.3.1 Google

Google lists "more than 200 signals" that any given ADM process uses to determine the final ranking of hits resulting from a search query (Google 2016). The analysis shows that signals can be grouped into four categories:

- Search query results
- User's characteristics
- Reactions of other users to the results
- Human evaluation

Search query results

The characteristics of a post or story which are evaluated to rank hits have little or only indirectly to do with the post or story's actual content. Google explicitly names a number of evaluative signals, including "freshness of content on a website," "words on the webpage" and "URL and title" (Google 2016). According to Google manager Andrey Lipattsev, these are the most important types of signals for ranking hits: "It's content, and links pointing to your site" (Ratcliff 2016). One of the few scientific studies on the relevance of ranking factors identifies the density of key phrases in the URL, title and HTML header as a significant criterion (Su et al. 2014). A large, mostly unregulated service industry has arisen dedicated to optimizing webpages to make them visible by search engines (Iredale and Heinze 2016). A core service offered by providers in this industry is adapting website characteristics to reflect changes in ADM processes.

User's characteristics

The user is the one who actually provides the written or spoken query. The intention behind a query depends on a number of other factors, for example the user's place of residence or language. Google talks in the abstract about "personalization" as a signal (Google 2016), although which of the user's characteristics are actually meant is not discussed. According to multiple Google representatives (Levy 2016), the third most influential ranking factor for predicting intention is the ADM process known as *RankBrain,* which looks at various metrics, including the potential significance of previously unsubmitted search queries, by extrapolating from past queries (Clark 2015). In 2008, a Google product manager explained in a brief blog post that the user's presumed location, recent search queries and the web history of the user's browser are used for personalization purposes, but only if the user has agreed to the use of these data (Garb 2008).

Reactions of other users to the results

The ADM process measures people's reactions to search results or content using a number of criteria, including the number of other websites linking to a particular site and the "authority of those links" (Google 2016). To create the ranking, the ADM process also evaluates the behavior of Google users on a particular page, although only indirectly. How many users click on which hits and how quickly they return to the hit list are seen as a signal of the hits' quality. Yet information on this behavior is not so widely available and reliable as information on links (Ratcliff 2016) and is thus included in a different manner in the ranking of results. Signals from social networks also feed into the ranking: "tweets and social mentions – which are basically more links to the page, more mentions of this context" (ibid.).

Human evaluation

Google also pays people to evaluate the quality of search results for certain queries (see Section 3.2).

Evaluators assess the quality of results for specific queries based on a 160-page handbook (Google 2017). Their evaluations are based on two key scales*: Page Quality* (the results' contents) and *Needs Met* (how well the content matches the search query).

Page Quality refers only to the returned pages. Evaluators are supposed to determine a page's expertise, authoritativeness and trustworthiness. The instructions in the handbook on how to do research read in part like a how-to manual for media skills:

> *Many other kinds of websites have reputations as well. For example, you might find that a newspaper website has won journalistic awards. Prestigious awards, such as the Pulitzer Prize award, are strong evidence of positive reputation. When a high level of authoritativeness or expertise is needed, the reputation of a website should be judged on what expert opinions have to say…. Look for information written by a person, not statistics or other machine compiled information. (Google 2017: 16)*

The Needs Met scale is meant to quantify how well the hits match the query and the user's intention. One extreme example: Users who explicitly search for *alternative viewpoints* of the Holocaust should also find links to sites denying the Holocaust: "Pages that directly contradict well-established historical facts should be rated Fails to Meet, unless the query clearly indicates the user is seeking an alternative viewpoint" (ibid.: 118). Questions such as "Did the Holocaust happen?" do not, however, fall into this category. Google changed the guidelines for evaluators after several media outlets reported on websites denying the Holocaust appearing in search results (Sullivan 2017). The company introduced an additional criterion: Evaluators can now also mark the content as *upsetting-offensive.* The new version of the manual explicitly lists Holocaust denial as an example of this category.

According to Google, the quality assessments of search engine evaluators are not directly included in the ranking of individual results in the live version, but are used, for example, in judging experimental changes in the ranking procedure (Underwood 2015). It is not publicly known how many people work as evaluators and to what extent Google uses their assessments. The US-based search machine expert Danny Sullivan estimates the number of Quality Raters, as Google calls them, to be over 10,000 worldwide (Sullivan 2017); 600 Quality Raters are believed to work in Germany (Spiegel Online 2017). Job listings suggest that at least some search engine evaluators are not employed directly by Google, but as freelancers working at home by subcontractors.

Table 3: Select overview of publicly disclosed signals for Google's ranking of search hits

Category	Signals
User's characteristics	User's typed or spoken search query

	User’s intention as predicted by RankBrain ADM process
	Personal characteristics such as place of residence
Content of search results	How long has the page existed?
	Which terms appear where and how often?
Other users’ reactions to the results	How many websites link to the page?
	Who mentions the page?
	Which terms link?
	How quickly do users return to the list of results?
	Who links to the contents on social networks?
Human evaluation	Indirect: How do search engine evaluators assess a page’s expertise, authoritativeness and trustworthiness?
	Indirect: How do search engine evaluators assess how well the results match a specific query?
Sources: Google 2016; Google 2017; Christopher Ratcliff 2016; Mimi Underwood 2015**.**	

4.3.2 Facebook

When it comes to a social network such as Facebook, there is usually no clear signal revealing a user’s interest in information, since it is likely that no such unequivocal interest exists. The algorithmically selected offerings in News Feed are meant to be potentially interesting – a characteristic that News Feed basically shares with certain tabloid media. Yet Facebook’s content derives from different sources: a mix of opinions, entertainment and personal or medial communication (Backstrom 2013). Facebook uses multiple signal types to assemble this mix:

- User’s characteristics and preferences
- Relationship between user and publisher
- Content characteristics
- Other users’ reactions
- Human evaluation

User’s characteristics and preferences

Advertisers can pay for content to be inserted in the News Feed of certain target groups based on user characteristics, such as age, place of residence, gender and interests. Since such ads can also lead people to subscribe to certain Facebook pages, such ads can have an impact on the mix of entries appearing in News Feed. Anyone who subscribes to a page based on an ad tailored to his or her interests establishes a relationship with this kind of content and could begin seeing it more often as a result.

In assembling the feed, Facebook examines which content users prefer (by using “see first”) or hide (Facebook 2016b). Both signals are based on actions that are relatively straightforward. Facebook also undoubtedly deploys signals that users are not consciously aware they are sending, such as the amount of time they view a certain entry in the feed. Users who spend more time with any one item are signaling approval without conscious interaction (Oremus 2016).

Relationship between user and publisher

The more often users interact with a friend’s Facebook profile, a page or a public figure who has posted, the more likely that content from this source will be given priority in the future (Backstrom 2013). Interactions can include likes, clicks, page views and comments, as well as slower scrolling through an author’s posts (Constine 2016).

These profiles are also given greater weight when other users’ reactions to content are evaluated: If a large number of Facebook users click on a certain post, the probability increases that it will be ranked higher in a person’s News Feed. The probability increases further when many users click on a post from a source that the user has frequently interacted with in the past (Backstrom 2013).

Content characteristics

Facebook also prioritizes based on type of post (links, videos, photos, texts) and does so using the following logic: Users who have demonstrated through past interactions that they like photos are more likely to see photos in their News Feed in the future (Facebook 2016a). Other characteristics used as signals to assess posts have rarely been documented in public. Here is one description of how other types of interactions are assessed by ADM processes to rank content: “Videos played longer than three seconds, have sound activated, or are watched all the way through are prioritized in News Feed” (ibid.).

The official reason for why Facebook gives preference to videos played with the sound on and viewed more by certain users is telling: “We’ve found videos tend to be more engaging than any other content type” (ibid.). This suggests that many – perhaps all – signals are meant to predict “engagement,” i.e. the largest number of clicks, likes, shares or comments. Yet what causes reactions is also relevant. To that extent we have to relativize the relevance of content characteristics: According to its publicly stated rationale, when ranking content Facebook seems to consider all of the content’s characteristics only in relation to their impact on users. Based on this logic, content has, independent of individual reactions, no characteristics that could justify prioritization. This reflects the so-called *News Feed Values* that Facebook adheres to: “Something that one person finds informative or interesting may be different from what another person finds informative or interesting…. Our aim is to deliver the types of stories we’ve gotten feedback that an individual person most wants to see” (ibid.).

In its so-called *Community Standards,* Facebook lists characteristics of posts that, regardless of the public’s reaction, would cause them to be removed from News Feed. These include certain types of nudity, copyright infringement and calls to engage in terrorist activities (Facebook n.d.).

Other users’ reactions

Posts with many likes, comments or shares have a higher probability of appearing higher up in News Feed – especially when a user often interacts with the people doing the sharing (Facebook 2016a). In addition, the ADM process calculates how quickly people return to Facebook after they followed a link in News Feed to an external source. The assumption is: If people quickly return, the presentation in News Feed promised more or something other than what the linked sourced provided (El-Arini and Tang 2014).

Human evaluation

Since mid-2014, Facebook has also been using human evaluators. They assess the quality of News Feed in terms of its presented stories, writing an evaluation of the placement of every post. Initially almost 1,000 people employed by a service provider in the United States were paid to do this work; since mid-2015 evaluators are at work worldwide (Oremus 2016). It is unclear how many there are and what influence their evaluations have.

Here, however, is a known example of where evaluator feedback changed the ADM ranking process: People do not click on shocking or unsettling stories in News Feed, even though they find them highly relevant. After this behavior was confirmed by the so-called Feed Quality Panel, employees developed a new signal for engagement: the time spent with a story in the feed without any additional interaction (ibid.). Facebook also takes content evaluated by humans based on specific questions and uses it for machine learning. For example, evaluators examine the informational quality of posts and the trustworthiness of the underlying source. Based on this input, an algorithm has been trained capable of recognizing worthwhile content (Kacholia 2013).

Every day, moreover, Facebook asks tens of thousands of randomly selected users around the world about the quality of their News Feed. Participants in this Feed Quality Program (Xu, Lada and Kant 2016) say whether or not they would like to see individual posts in their feed, rating them on a scale of 0 to 5 (Zhang and Chen 2016).

When users report posts that violate the Community Standards, human evaluators examine the offending entries. In Germany, they are employees of service provider Arvato[2] (Reinbold and Rosenbach 2016).

Table 4: Select overview of publicly disclosed Facebook selection criteria

Category	Signal
User’s characteristics and preferences	Past reaction to content (likes/hides)
	Time spent viewing posts
	Sociodemographic traits listed in profile
Content characteristics	Format (e.g. photo, video, text) Similar content that testers have rated positively
	Topicality
Other users’ reactions	Number of likes, shares, comments, hides
	Time spent viewing
	How quickly users return to News Feed
	Survey results on posts
Relationship between user and publisher	Friends, tags
	Clicks, scrolling behavior, page visits
	Comments, likes
Human evaluation	Evaluation of News Feed by paid evaluators
	Survey results on posts
	Human checking of offending posts

Sources: Backstrom 2013; Facebook 2016a; Facebook 2016b; Oremus 2016; Zhang and Chen 2016**.**

4.4 Conclusion: Using algorithms to determine relevance

Both of the intermediaries examined here measure relevance using similar criteria:

- How do users react to content?
- Which of the content’s characteristics correlate to the desired user reactions?

For a number of reasons, caution should be exercised when using either type of signal to measure relevance, no matter how relevance is defined.

[2] Full disclosure: Arvato is part of the Bertelsmann SE & Co. KGaA corporate group, whose shareholders include the Bertelsmann Stiftung.

4.4.1 User reactions

Diverse types of use: People can interact with posts for a wide variety of reasons, especially when it comes to intermediaries operating social networks. The types of use can be divided into three practices, as proposed by Paus-Hasebrink, Schmidt and Hasebrink (2009): relationship, identity and information management (summarized in: Schmidt et al. 2017: 21). Thus, the engagement with content and with information about social discussions is not always a function of information management. People can react to content for entirely different reasons. Perhaps they like certain stories because many of their best friends have already mentioned them. In this case, the visible reaction does not so much signal the quality of the story’s content as it does the user’s inclusion in a certain group or his or her desire to be included in the group, and that this post is a good way to make that clear. The impact that the different types of user behavior have is more likely greater in social networks than when search engines are being used.

Cognitive distortions through impulsive reactions: People send signals quickly and impulsively – for example, they click on links, “like” posts or take a bit more time scrolling through an article. A mode of thinking most likely predominates here that Kahneman refers to as “System 1,” a mode that “operates automatically and quickly, with little or no effort and no sense of voluntary control” (Kahneman 2012a: 34). The strengths and weaknesses of this mode lead to cognitive distortions in perception that are systematic, such as a general tendency to agree. It “favors uncritical acceptance of suggestions and exaggeration of the likelihood of extreme and improbable events” (ibid.: 133). If there is not sufficient correction of the evaluation of System 1 signals as signifiers of relevance, then the very human phenomenon of cognitive distortion could become part of the ADM process. Human fallibility would thus be magnified by machine-based decision-making and, to some extent, absolved. With that, processes which are extremely subjective and unconscious could become seemingly objective standards of relevance or even quality.

Reactions are not independent variables: An intermediary’s users do not determine, freely and without outside influence, which posts they prefer. They only see an algorithmically produced subset of all possible posts, a selection they then react to. Their reactions are the basis for further personalization and, thus, in terms of impact, no clear distinction can be made: Neither does the user determine what happens in the ADM process, nor does the ADM process determine the user’s reactions. The two interact and influence each other on an ongoing basis. Therefore, Facebook’s promise that “you control your experience” (Facebook 2016b) is only partially true. Individual users only control how their personalized News Feed is structured to a limited extent. The mix and weighting of all signals cannot be determined by any single user. Using a News Feed that is merely sorted chronologically is not truly an alternative, since the sheer volume of posts makes it impractical. A workable alternative could be, for example, choosing between different forms of algorithmic structuring that offer different degrees of freedom and that access different providers. This would make it possible for users to have their personal News Feed ranked by an algorithmic system that shows only links to news sources which make use of a minimum of transparently defined editorial principles. (Possible criteria here would include sources that maintain the firewall between editors and publisher, explicitly provide information on who is responsible for content, participate in industry self-regulation, etc.) Another alternative would be a system that evaluates the reactions of other users over a longer period of time, a period that the primary user determines himself or herself (e.g. posts visible at the top of the feed are those frequently recommended by friends in the last 48 hours).

4.4.2 Content characteristics in relation to user reactions

This analysis of the criteria shows that both intermediaries have their own particular understanding of relevance: Relevance is personalized. The primary goal is optimally matching content to an individual’s preferences. The efforts to measure relevance, as thus defined, are intended to increase satisfaction among users. Satisfaction becomes evident through greater and more intensive use of the platform.

That satisfaction and relevance can be conflated is, however, mere assumption. It could be argued, in contrast, that social discourse which fosters participation does not result from satisfying as many individual needs as exactly as possible, but, for example, from striking a balance among as many needs as possible.

This view of relevance as satisfying individual needs can be seen in factors such as Google’s Needs Met (NM) metric or in Facebook’s focus on user engagement. Google does collect information on some signals unrelated to users’ reactions: the distribution of key words in source material, for example, and page quality as assessed by human search engine evaluators. Yet how these factors affect the ranking of search results has not been publicly disclosed.

The focus on matching personal preferences means the quality of posts is always defined by the recipients’ satisfaction. One potential result is a relativization of the truth, as described by Facebook in its so-called *News Feed Values*:

> *We are not in the business of picking which issues the world should read about. We are in the business of connecting people and ideas — and matching people with the stories they find most meaningful. Our integrity depends on being inclusive of all perspectives and view points, and using ranking to connect people with the stories and sources they find the most meaningful and engaging. (Facebook 2016b)*

With this, ADM designed to match individual preferences is elevated to the paramount principle: Matching trumps respect for the truth.

This process must be fundamentally scrutinized, but its implementation also requires a critical appraisal. If satisfaction is severely limited to page views, length of use, likes and similar metrics, then a number of conceivable criteria are necessarily left out as a means of measurement: impartiality, increased knowledge, a constructive exchange with individuals of opposing views, building consensus and identifying the truth, among others. Observing and analyzing the behavior of many does not lead to valid statements about everyone. As Etzioni and Etzioni (2017) aptly put it, describing ethical considerations: “Observing people will not teach machines what is ethical but what is common” (ibid.: 5).

The drawbacks of this approach can be seen in extreme cases such as the widespread presence of fake news on Facebook during the presidential election in the United States (Silverman 2016) and the prominent placement of posts denying the Holocaust in Google search results following neutral queries (Mierau 2016). These examples show that lies also provoke strong reactions; lies also meet human needs and reflect human interests. If the originator of the lie is primarily concerned with achieving the greatest reach on intermediary channels using algorithmically sorted content, then a lie can be even more efficacious than the truth, since its author can optimize it to ensure a maximum number of interactions and maximum engagement (see for example Silverman and Alexander 2016).

In the following chapter, we discuss a number of factors and analyze their effect within the system outlined above and as part of intermediaries’ activities shaping public opinion.

5 Where it’s leading: The changes currently transforming public discourse

Chapter 4 outlined the structural change transforming public discourse. Chapter 5 examines the new forms that public discourse and public opinion are taking within this new structure in light of current conditions.

5.1 What you see is all there is

In a social network, a brief excerpt from a story or video can serve as the basis for forming an opinion, as can comments or recommendations of other users; the underlying content need not play a role in how the opinion is formed. Users who see the headline, introductory text, thumbnail image and 30 comments may not even look at the content itself.

Some empiric data suggest that people recommend posts in social networks without having actually read them. The research on this point has some methodological constraints, namely that it examines individual posts, and not the more general behavior of users or viewers. It can determine if there is a correlation between the number of mentions of an article in a social network and the number of views or the depth of the engagement with the article in a given online medium. Using such data, it is not possible to say with certainty how users behave, but it can be presumed: There is no appreciable correlation between the number of mentions a post gets on Twitter, for example, and the number of times it is viewed or how thoroughly it is read.

In 2013, the web analytics company Chartbeat examined how often an article is mentioned on Twitter and how that relates to reader engagement, scroll depth and the share of completely read stories. Only a weak correlation exists: “Both at Slate and across the Web, articles that get a lot of tweets don’t necessarily get read very deeply. Articles that get read deeply aren’t necessarily generating a lot of tweets” (Manjoo 2013).

Another study comes to the same conclusion, namely one that looked at the diffusion of articles from the BBC, CNN, Fox News, New York Times and Huffington Post on Twitter during one month in the summer of 2015. The researchers analyzed when and how often links to articles from these media were retweeted on Twitter and how often the related URLs supplied by URL-shortening service bit.ly were actually clicked (Gabielkov et al. 2016: 2). The statistical analysis suggests that articles which are shared are not necessarily read. To quote the study’s authors: “There seems to be vastly more niche content that users are willing to mention in Twitter than the content that they are actually willing to click on” (ibid.: 9).

These findings imply that intermediaries are more than just a distribution channel. If the number of shares that articles receive is greater than the actual number of times the articles get read, then two phenomena are presumably taking place:

- *Competition for attention:* The success factors for maximizing reach in the relevant networks affect how editorial media present their content in those networks, since a post’s headline, introduction and thumbnail are just as useful for advertising purposes as for informational objectives. They are meant to generate attention and motivate readers to click on the link to the source material. This suggests that intermediaries’ methods are being anticipated, while subject matter is being overstated and truncated and content exaggerated (see also Section 5.2).
- *Priming:* Ranking and comments by others affect how people perceive articles. Users who discover a story through another reader’s recommendation which calls it “horrible” are exposed to the assessment before they experience what has been assessed – if they even proceed to read the assessed content itself.

In both cases, users often read only the preview and comments instead of the complete story. This can lead to a premature conclusion, known as the WYSIATI effect: "What you see is all there is" (Kahneman 2012a: 113).

5.2 Emotion generates reach

As findings from a number of studies suggest, there is a correlation between the sentiment expressed in a post, the reach of the post in social networks and the sentiment expressed in posts by subsequent readers.

A study by Stieglitz and Dang-Xuan (2012) examining the posts and comments on the Facebook pages of German political parties in 2011 shows that posts expressing negative emotions provoke more reactions than those that are not inherently emotional. The authors' software text analysis counted the frequency of words with a negative (e.g. "disappointed") and positive (e.g. "hurray!") connotation and assessed the intensity of the expressed emotions on a scale of 1 to 5. The analyzed sample consisted of 5,636 posts and 36,397 comments from 10,438 users on the Facebook pages of German political parties (CDU, CSU, SPD, FDP, B90/The Greens, The Left, Pirate Party) between March and September 2011. One key finding is that negativity engenders more reactions. The more negative the Facebook posts are (as measured by the frequency and intensity of the relevant terms), the more comments they receive. No such correlation exists for posts with positive content.

The same researchers examined the correlation between the emotions expressed and the reactions to them on another social network (Twitter) and partially confirmed their original findings (Stieglitz and Dang-Xuan 2013). In this study, the difference between the reactions to positive and negative posts is not as large as in the Facebook study. At the same time, the authors did not look at the number of comments a post gets, but the number of retweets. They evaluated165,000 tweets from the official accounts of the CDU, CSU, SPD, FDP, B90/The Greens, The Left and Pirate parties in the run-up to the 2011 state elections in Berlin, Baden-Württemberg and Rhineland–Palatinate using the same methodology as in the Facebook study (ibid.: 225). The findings:

- The more emotionally a tweet is phrased, the more frequently it is retweeted. This effect is in some cases stronger when the emotion expressed is negative than when it is positive (elections in Berlin, not in Baden-Württemberg or Rhineland–Palatinate).
- The more emotionally a tweet is phrased, the less time passes before its first retweet. There is no correlation here between the type of the sentiment expressed and the amount of time that passes before the retweet.

The Facebook study by Stieglitz and Dang-Xuan (2012) reveals yet another effect: The comments that emotional posts receive express similar emotions. The more negative a post is phrased, the more negative the reactions, and the more positive it is, the more positive the resulting comments are. As Stieglitz and Dang-Xuan themselves summarize: "Our results indicated that positive as well as negative emotions might diffuse in the following discussion" (ibid.: 13).

A much larger experiment carried out by Facebook at the beginning of 2012 confirmed a connection between the emotions expressed in posts read by users and the subsequent communications behavior of the users themselves (Kramer, Guillory and Hancock 2014). For one week, Facebook showed 689,003 users posts from their News Feed based on the sentiment expressed in them. One cohort saw less positive posts, another less negative posts, and a third posts that were selected at random. There was a small but discernible effect: Users who saw fewer positive posts themselves posted content that was measurably more negative. Those who saw fewer negative posts contributed more positive content (ibid: 8790).

In addition to these key findings, the work of Kramer et al. shows yet again the power Facebook has on its users: As soon as the intermediary alters the ranking of its posts, it has several immediate effects, even changing the mood of those using the site (see Chapter 5.3).

Based on the empiric results, it is impossible to definitively ascertain which impact the intermediaries’ ADM processes are responsible for, and which can be ascribed to human behavior. The two cannot be separated during machine-based processing: If people share and comment on posts more often, then the ADM process can take that as a signal for the content’s relevance. The more retweets and comments a post receives, the more relevant it is assumed to be, including for a broader target group. Based on this assumption, emotional posts could possibly be shown in more feeds than other more neutral content – not because the sentiment expressed has been analyzed, but because the process assumes relevance given the volume of user reactions occurring within a certain amount of time.

Regardless of whether the effect is positive or negative, the empiric results show that people’s reactions to posts can only be used to predict relevance to a limited extent, since their reactions clearly also depend on factors that no one would consider to be valid criteria signifying quality when it comes to the social debate. Highly negative posts and many heated reactions to them do not necessarily testify to relevance in terms of social discourse, often only to loudness for the sake of being loud (compare the discussion of cognitive distortions in Section 3.4.1).

The question of which criteria are appropriate for measuring relevance and which are actually used by intermediaries cannot be answered without analyzing the **goals** and **impacts** of ADM processes. Empirical research on emotional posts provides food for thought in both areas.

In terms of **goals:** What type of user behavior is an intermediary’s selection process designed to promote? Greater engagement? A higher click rate? More comments? A constructive discourse? Such goals are likely to end up in conflict with each other, and the empiric findings of Kramer et al. (2014) illustrate one such conflict. In the Facebook experiment examining how emotions are affected, the authors discovered that people who see fewer emotional posts in their News Feed become less active afterward. As Kramer et al. note: “We also observed a withdrawal effect: People who were exposed to fewer emotional posts (of either valence) in their News Feed were less expressive overall on the following days, addressing the question about how emotional expression affects social engagement online” (ibid.: 8792). If intermediaries want to increase user engagement, they could feasibly decide to increase the visibility of emotionally charged stories. “To reward that which keeps us agitated” is how experts expressed it in a Pew research survey (Rainie, Anderson and Albright 2017: 12).

Which other **impacts** emotional posts in social networks have has been the subject of little research in the past. It does not seem unreasonable to assume that stories with negative content could lead to destructive behavior. Based on an experiment involving 667 participants and an analysis of over 16.5 million comments on CNN.com, a study on comments by trolls[3] (Cheng et al. 2017) comes to two key conclusions:

- People who are in a bad mood (in the experiment, users’ moods were influenced by a test and then assessed using a questionnaire) are more likely to take on the role of troll in online discussions.
- People tend to become a troll in a discussion if previous comments evince similar behavior.

A direct correlation cannot be shown between the troll research and empirical studies examining emotions expressed in posts. We do not know if the mood expressed in Facebook posts has an effect only on the mood of users as expressed in Facebook posts or on other behavior as well. It seems plausible, however, that such an effect could exist. If so, insults and personal attacks would be more likely to occur when negative posts achieve a greater reach.

[3] Defined as “flaming, griefing, swearing, or personal attacks, including behaviour outside the acceptable bounds defined by several community guidelines for discussion forums” (Cheng et al. 2017: 2)

5.3 Distorted cognition shapes social media's data set

C*onfirmation bias* is a psychological phenomenon originally identified by researchers investigating cognition and memory. It refers to the tendency people have to fill in gaps in their memories with things they already believe and to believe ideas that match beliefs they already hold (see for example Klayman 1995). Kahneman puts it thus:

> *Contrary to the rules of philosophers of science, who advise testing hypotheses by trying to refute them, people (and scientists, quite often) seek data that are likely to be compatible with the beliefs they currently hold. The confirmatory bias of System 1 favors uncritical acceptance of suggestions and exaggeration of the likelihood of extreme and improbable events (Kahneman 2012b)*.

System 1 refers to a mode of operating quickly and automatically, with no effort or sense of control (see Section **Fehler! Verweisquelle konnte nicht gefunden werden.**).

In addition to *confirmation bias*, other cognitive distortions exist that presumably also affect ADM processes in terms of the content present in the media and communications context. The following is a brief summary of one such distortion.

The *availability heuristic* (see Tversky and Kahnemann 1974) describes the tendency people have to overestimate the frequency of events for which they can easily recall concrete examples (*ease of processing*). For instance, people will overestimate the probability of a plane crash occurring shortly after seeing multiple news stories reporting a recent crash. In addition, the probability of an event is thought to be higher than it actually is if a person has personal memories or knows of other dramatic examples relating to it (see Kahneman 2012b).

This distortion leads, for example, to married couples, when asked which percentage of the housework they do, giving numbers that add up to more than 100%. Both spouses tend to remember those moments when they did the housework more than those when their partner did and thus overestimate their own efforts. Similarly, the probability of meeting a violent death, e.g. through accidents, is often massively overestimated compared to what are in fact the most common causes of death such as stroke. The reason is that many people more easily recall the former thanks to media reports and their own emotional reactions to them. In discussions such as those on social media, whose content is ranked by ADM systems, this type of cognitive distortion can proliferate easily and effectively. For example, some people repeatedly see stories on crimes committed by Muslim immigrants because they extensively interacted with such articles in the past or because they are part of a social group that actively comments on such happenings. Over time, such users will overestimate the probability that Muslims commit crimes.

Another cognitive distortion plays a role in this last example, namely *base rate neglect.* This means that the probability of a certain crime being committed by a Muslim or a person of non-native heritage will probably be overestimated because the overall presence of the relevant demographic group is not considered during the implicit calculation of probability. In Germany, for example, there are many more non-Muslims than Muslims, something that is not considered when people predict the probability that crimes will be committed by Muslims, potentially resulting in an estimated probability that is too high.

Both the availability heuristic and *base rate neglect* are often seen in the literature as one overarching concept, which US-based researcher Herbert Simon has called *bounded rationality* (see Simon 1979). The debate that has been taking place for four decades essentially asks the question of whether heuristics such as the one described by Tversky and Kahneman must inevitably be irrational and misleading or, if under certain conditions (for example when people do not have sufficient information available to them) they might be beneficial (Gigerenzer and Gaissmaier 2011). This paper does not offer sufficient space for a comprehensive assessment of this complex issue.

What can be said is that extensive empiric evidence exists showing that people often do not process information in a rational manner and that cognitive distortions occur in particular when decisions and judgments must be made quickly, often emotionally and without time to reflect, i.e. “automatically.” A comparatively new science, one based on findings from traditional learning theory, is *captology,* a digitally supported method of behavior modification (see Section 5.4). Captology’s explicit goal is to more readily influence behavior through various means, for example by simplifying and accelerating cognitive and behavioral processes as much as possible (Eyal 2014). Thus, much suggests that the design decisions made by intermediaries emphasize cognitive styles that in turn give rise to cognitive distortions.

5.4 Platform design influences human behavior

When intermediaries change the way their systems function, it has ramifications on the user experience and on users’ behavior. In other words, with every design change, intermediaries influence the signals that they themselves measure in order to assess relevance, something that has been clearly demonstrated by numerous studies.

One example: When Facebook released a new feature called People You May Know (PYMK) in March 2008, it considerably changed both the form and manner of users’ networking, nearly doubling the number of new links between users (Malik and Pfeffer 2016). The researchers come to the conclusion that “it is not just the process of joining social networking sites that creates observed network properties, but also the ways in which platforms design influences users.” The PYMK mechanism is based on an algorithmic evaluation of Facebook users’ digitally portrayed circle of friends, which is then used to suggest other users that the original user might possibly know. This evidently led to a significant rise in the ties between users within the network. According to Malik and Pfeffer, the social ties within a social media network are a “non-naturalistic measure of social relationships.”

This result makes another important point clear: Intermediaries’ design decisions – here, the introduction of the PYMK feature – potentially have an immediate and immense effect on user behavior, for example the number of ties to new “friends” that are made. The design decisions made by intermediaries thus constantly influence the signals that the intermediaries themselves use to measure relevance.

An impressive example of the power intermediaries have was provided in 2012 by a research group in which Facebook employees were also involved. In 2010, a post had been disseminated by Facebook to 61 million of its users calling on them to learn about the candidates and issues and then vote in the coming US congressional elections. Here, in the authors’ own words, is what happened:

> *The results show that the messages directly influenced political self-expression, information seeking and real-world voting behaviour of millions of people. Furthermore, the messages not only influenced the users who received them but also the users’ friends, and friends of friends. (Bond et al. 2012)*

The authors believe that 340,000 people or 0.14% of the American electorate voted as a direct result of the campaign. Naturally, such an effect is desirable in a democracy. Yet it also shows that intermediaries with a reach of this magnitude can, under certain circumstances, decisively influence election outcomes, especially when one considers that the additional people inspired to vote by such tactics may not be spread evenly across the political spectrum. Robert Epstein and Ronald Robertson warn of the potential social and political power that other intermediaries have, namely search engine operators. In a series of experiments using voters in the US and India, they showed that actively manipulated search engine results could influence the electorate’s voting behavior (Epstein and Robertson 2015). The authors call their finding *search engine manipulation effect* and say that 20% of previously undecided voters could be influenced using search engines to shift their preference in one direction or another. This figure has been disputed by others in light of methodological concerns (Zweig 2017). Yet even the author critical of the work of Epstein and Robertson, socioinformatics specialist Katharina Zweig, sees the results as meaningful, especially for elections in countries with first-past-the-post systems. However, according to Zweig, the actual effect is somewhere between 2% and 4%.

Both Google and Facebook could thus have a huge influence on the information seen by users and, as a result, could directly impact social processes. Even less visible changes to a platform can have a massive effect on users' perceptions and behavior. As Kosinski et al. (2015) write:

> *As users are more likely to interact with content and people suggested to them by Facebook, their behavior is driven not only by their intrinsic goals and motivations, but also (to some unknown extent) by the Facebook algorithms constantly adjusting their exposure to content and friends. For example, friends' photos that appear on a given user's Facebook news feed are clearly more likely to be liked. Essentially, largely unknown effects of personalization represent a general class of confounding variables characteristic for observational research and deserve further study. (Kosinski et al. 2015)*

Cases have also been documented for Google in which fundamental alterations to the search algorithm have caused massive changes in the resulting rankings and, with that, in whether certain business models would succeed. This, in turn, altered the behavior of site operators and, of course, users. According to Google, in 2011 it made changes designed to "reduce rankings for low-quality sites," thereby placing them further down in the results (Cutts 2011). The codename for this algorithm change was *Panda*. The goal was to remove from the search results so-called *content farms,* websites which offer content that is of little value but that matches what search engines are looking for. Content farms are strictly money-making ventures; they are created to attract visitors using search-engine-optimized content and monetize their visits through advertising displayed on the low-content pages.

Panda and other wide-scale changes to the search algorithm – *Penguin* and *Hummingbird* – carried out in subsequent years led in some cases to furious reactions among players in the search engine optimization industry (see for example Jenyns 2013). Yet again it became clear that Google can destroy overnight business models that, as in this case, are designed to misuse its search engine.

For the purposes of this paper, what is important is the fact that every fundamental intervention in the search algorithms necessarily causes massive – and in some cases positive – changes in users' behavior and thus in the media content they see. Diverse studies have shown that the overwhelming majority of users only click on links that appear on the first page of results. For example, research carried out by van Deursen and van Dijk (2009) shows that 91% of subjects never got further than the first page of the search results (ibid.). Conversely, analysis carried out by providers engaged in search engine optimization and marketing shows that even the last link on the first page of results, the one ranked 10th, is only clicked by a very small percentage of users. Different studies have produced a range of click rates, from 1% to 3.6%, for 10th-place sites (Dearringer 2011).

Changes to the system generating such results thus have an enormous effect for providers on whether their sites are seen and, for users, on the content that gets seen.

The influence intermediaries have is great when it comes to the user experience and thus for society as a whole, but also for businesses and organizations that rely on intermediaries to diffuse their content. This inevitably leads to the question: Are these decisions and changes relevant for social participation? What is clear is that when today's intermediaries become part of the equation, complex interactions result between their design decisions and a range of other factors, including psychological mechanisms, content characteristics and external attempts at manipulation. Even the political system of the country in which the user lives plays a role in this complex, dynamic system.

Intermediaries are of course aware of the factors discussed in Sections 5.1 to 5.3 which influence the probability that users will interact with digital sites and the intensity of that interaction when they do. Moreover, for the last 15 years a new branch of psychology dedicated to human-machine interactions has been examining the question of how digital technology can influence human behavior. Its spiritual father, B.J. Fogg, christened this discipline *captology* (shortened from *computers as persuasive technologies).* In the preface to the first edition of Fogg’s book *Persuasive Technology,* the well-known social psychologist Philip Zimbardo, one of Fogg’s teachers, wrote that it was “the very first book on a totally new field of intellectual inquiry, one that has important practical implications for a host of people across many different domains, all centered around persuasion” (Fogg 2003).

In the book, which is now considered a milestone in academia’s attempts to address human-machine interaction, Fogg shows how classic learning-theory ideas and processes used in behavioralism and behavioral therapy can, with the assistance of digital systems, have a more efficient and effective impact on human behavior. Neither Facebook nor Instagram existed when Fogg’s book was first published, and smartphones as we know them today had yet to be invented. Yet numerous founders and employees of famous Silicon Valley’s companies attended his courses at Stanford University, including Instagram founder Mike Krieger. In Fogg’s course, Krieger even worked with a fellow student to develop an application, based on Fogg’s ideas, for sharing photos (Leslie 2016).

Today, the methods developed by Fogg are used by almost all developers of software products for consumers. Nir Eyal, the author of *Hooked: How to Build Habit-forming Products,* is an alumnus of Fogg’s laboratory at Stanford (Eyal 2014). Eyal has been lauded by, above all, investors, start-up founders and representatives of the advertising and marketing industry, but less so by academics. His applied research on digital behavior modification has been successful in a way that now disturbs his teacher. Fogg has been quoted as saying, “I look at some of my former students and I wonder if they’re really trying to make the world better, or just make money. What I always wanted to do was un-enslave people from technology” (Leslie 2016).

Captology considers components known from traditional learning theory such as *trigger*, *action*, *reward* and *habit*, differentiating them and applying them to concrete scenarios known from digital applications. This brief extract from *Hooked* serves to illustrate the point:

> *Facebook provides numerous examples of variable social rewards. Logging in reveals an endless stream of content friends have shared, comments from others, and running tallies of how many people have “liked” something. The uncertainty of what users will find each time they visit the site creates the intrigue needed to pull them back again. While variable content gets users to keep searching for interesting tidbits in their News Feeds, a click on the “Like” button provides a variable reward for the content’s creators. (Eyal 2014: 78)*

A key goal of captologists’ optimization efforts is to remove barriers between users and a given type of behavior. Referencing Fogg’s theories, Eyal puts it thus:

> *The pattern of innovation shows that making a given action easier to accomplish spurs each successive phase of the web, helping to turn the formerly niche behavior of content publishing into a mainstream habit. As recent history of the web demonstrates, the ease or difficulty of doing a particular action affects the likelihood that a behavior will occur. To successfully simplify a product, we must remove obstacles that stand in the user’s way. (Eyal 2014: 51/52)*

This goal is illuminating when considered against the background of System 1 (fast, intuitive, automatic, not difficult, susceptible to error and manipulation) and System 2 (slow, deliberative, ordered, requiring effort), the discrete cognitive processing systems posited by Kahneman (2012) and others. A core design principle advanced by Silicon Valley’s theoretical thought leaders is explicitly meant to promote System 1 cognition and actively avoid System 2 cognition. That is advantageous if one measures success using metrics such as clicks, shares, likes, frequency of publication or other quantitative benchmarks of user behavior. If one prioritizes other criteria instead, such as understanding, deeply engaging with content, or the quality of shared content, then one’s design decisions would undoubtedly be of a different nature (see Section 7.2).

Hypotheses derived from such learning-theory fundamentals are constantly being explored by intermediaries, who permanently experiment by manipulating their systems with the explicit goal of optimizing user interactions:

"At Facebook, we run over a thousand experiments each day," wrote Facebook researcher Eytan Bakshy (2014) in a blog post. According to Bakshy, some of these experiments were meant to optimize short-term outcomes, while others served as the basis for long-term design decisions.

Google is also known to carry out complex experiments during daily operations with real users before it makes even the smallest change to the user interface – to decide, for example, which shade of blue will be used for the links in advertisements. According to media reports in 2014, such a color change increased the company's revenue by $200 million (Hern 2014*)*.

In terms of the topic addressed in this paper, this approach is significant for one reason in particular: As they conduct experiments meant to optimize their products, intermediaries constantly change variables that effect measurements of their own relevance criteria.

When the frequency and intensity of interaction is increased, the number of clicks on Facebook's "like" button grows, as does the click rate for links in Google hits. The question, however, is which type of engagement with content these variables reflect. To be sure, the systems are not adjusted in order to optimize informed public discussion, but to serve the providers' commercial interests.

5.5 Measured reach is not necessarily actual reach

One factor, which casts the numerous benchmarks and outcomes described above in a new light, has yet to be discussed here: the possibility of deploying automated processes to influence and distort many of the metrics that intermediaries use to measure "relevance." Presumably the most significant tool currently being used to this end is the so-called *bot,* an automated software app designed to be mistaken for a human user. Some bots serve commercial, others criminal purposes. They "click" on ads to generate revenue, for example, or to deplete a competitor's advertising budget as the result of meaningless traffic. Some search the web for weak spots that would allow hackers to attack, while others engage in *scraping:* extracting information from webpages, for example those listing competitors' prices, to gain a competitive advantage (see for example Zeifman 2017).

In the context of this paper, two types of bots are especially relevant: so-called *auto clickers,* which simulate real traffic on webpages, and the latest development in this field, so-called *social bots*.[4] The latter are partially autonomous software apps that behave on social networking platforms as if they were real humans, but which are simply following predetermined commands (Kollanyi, Howard and Woolley 2016). For example, such bots automatically issue tweets on Twitter that are often padded with certain hashtags, or they retweet posts from other accounts to help them achieve greater reach and what seems like relevance. Bessi and Ferrara estimate that during the 2016 presidential election in the US, some 400,000 bots were active on Twitter alone, sending 3.8 million tweets and making up "about one-fifth of the conversation" about the election. Kollanyi, Howard and Woolley (2016) even estimate that in the run-up to the election some one-third of the pro-Trump activity on Twitter was driven by bots and other highly automated accounts, compared to about one-fifth of the pro-Clinton activity.

Howard and Kollanyi (2016) also found when examining the *Brexit* campaign determining whether the United Kingdom would leave the European Union that a substantial share of Twitter posts had been generated automatically: Less than 1% of all Twitter accounts that used relevant hashtags like #*Brexit* or #*StrongerIn* were responsible for almost one-third of all tweets relating to these topics. In the run-up to the vote, the observed bots

[4] Full disclosure: One of the authors of this paper, Christian Stöcker, is currently participating in a research project (PropStop), supported by the German Federal Ministry of Education and Research, that is examining ways to identify and thwart automated propaganda attacks (see https://www.wi.uni-muenster.de/de/news/1975-bmbf-projekt-zur-propaganda-erkennung-online-medien-gestartet).

were, the authors write, used largely for “amplifying messages,” for example so that “the family of hashtags associated with the argument for leaving the EU dominates” (ibid.).

Hegelich and Janetzko (2016) examined a botnet that focuses on the Ukrainian/Russian conflict and evidently serves to disseminate pro-Russian messages. The authors come to the conclusion that the bots’ behavior “is not guided by a simple deterministic structure of command and obedience between a human botmaster and an army of bots.” Instead, it results from “complex algorithms leading to a high degree of autonomy of the bots.” The algorithms behaved in a way that at first glance they seemed to be human users while following abstract rules such as “take a popular tweet and add the following hashtags.” These camouflaging strategies made it “extremely hard to identify the bots and to understand their political aim.” Such botnets, the authors conclude, are “a new development of high political relevance” (ibid.).

The studies cited above all examine Twitter, largely since bots can be relatively easily implemented on the platform, and recognized as such. At the same time, Facebook and other intermediaries are also home to accounts that seem to be normal users but are in fact automated apps serving specific political or commercial ends (see for example Lill et al. 2012).

In terms of the topic being examined in this paper, the phenomenon of *social bots* is relevant for one reason in particular: Bots are well suited to distorting the signals intermediaries use to measure relevance. Fake news is one example: In the US election, various purely fictitious posts resulted in high interaction rates at Facebook, much higher than reliable reports in traditional media. The best known example was the spurious claim that Pope Francis endorsed Donald Trump for president (see Silverman 2016). Yet it is entirely possible that the post’s metrics for reach and interaction were hugely distorted by bots; no one has yet been able to clarify how many of the likes and shares it received came from human users.

As this brief overview of studies shows, automated propaganda systems are already in wide use influencing politically contested issues. It therefore seems plausible that bots that behave in accordance with the maxim described by Hegelich and Janetzko (2016) – “take a popular tweet and add the following hashtags” – could make use of intermediaries to help stories that contain disinformation reach a much wider audience. This, in turn, could affect how intermediaries’ ADM systems rank the relevant links or posts. The result could be an upward spiral during which the sorting algorithms and the algorithm-driven bots reinforce each other. The ensuing digital public discussion would have nothing to do with reality. In other words, measured reach is not necessarily actual reach.

5.6 Polarization is increasing, but for more reasons than just the Internet

The term *“filter bubble”* was coined by the American author Eli Pariser (2011). Pariser defines filter bubble thus: “this unique, personal universe of information created just for you by this array of personalizing filters”*,* Pariser’s use of “personalizing filters” roughly matches what in this paper we call ADM processes: filtering and sorting of content that is tailored for individual users based on algorithmic processes.

The idea that digital communication services could help constrict the medial view of the world did not originate with Pariser. This idea was advanced as early as 2001 by American legal scholar Cass Sunstein (2001; 2008). Sunstein speaks of “information cocoons,” i.e. “communications universes in which we hear only what we choose and only what comforts and pleases us” (Sunstein 2008). At the same time, he is mostly concerned with personalization on the basis of an individual’s selective decisions, and less with the impact of algorithmic systems. Sunstein presented a concern that repeatedly resounds even in more recent work on this issue: that personal universes of information could lead to political radicalization and that the process of democratic will-building could be compromised if users no longer see a sufficient diversity of opposing viewpoints on intermediaries’ sites. According to Sunstein, society could lose its “social glue,” analogous to a fear expressed by Jürgen Habermas, who in 2008 warned that public life is disintegrating “in the virtual realm into a huge number of random, fragmented groups held together by special interests. This, seemingly, is how the public sphere as it currently exists in national contexts is being undermined" (Habermas 2008).

Evidently Pariser's vision of a society fragmented by widespread personalization was a cause of dissatisfaction at Facebook. In 2015, a study appeared in *Science* which deployed a massive amount of Facebook data to see if users availing themselves of an intermediary's services are in fact exposed mostly to ideological content reflecting their own attitudes (Bakshy, Messing and Adamic 2015). The three authors were Facebook employees, which in and of itself was enough to generate controversy in the academic community. *Science* therefore complemented the study with another article authored by an independent researcher, David Lazer of Northeastern University, who praised Facebook for being willing to engage in discussion on the subject while also warning that "this creates the risk that the only people who can study Facebook are researchers at Facebook" (Lazer 2015).

In their study, the three Facebook authors come to the conclusion that filter bubbles are more of an imaginary problem: "Our work suggests that individuals are exposed to more cross-cutting [political] discourse in social media than they would be under the digital reality envisioned by some" (Bakshy u. a., 2015). In discussing this point, the authors cite Pariser's book.

Yet the Facebook researchers' data in fact show that the algorithm led to a certain distortion in terms of the political leanings of the viewed content: "The risk ratio comparing the probability of seeing cross-cutting content relative to ideologically consistent content is 5% for conservatives and 8% for liberals" (Bakshy u. a., 2015). According to the authors, the user's own choices have a much greater impact: "Individual choices more than algorithms limit exposure to attitude-challenging content" (ibid.).

Both this conclusion and the methodological approach taken by the Facebook researchers were the object of severe criticism from some members of the social science community after the article appeared (for a summary, see Lumb 2015). One point subject to particular criticism was that Bakshy, Messing and Adamic had attempted to downplay the algorithmic ranking's considerable impact by comparing it to the effect that the users' own choices had. One must be viewed separately from the other, maintains sociologist Zeynep Tufekci (2015). According to Tufekci, what is ultimately novel, remarkable and significant is the effect the algorithmic ranking has on the mix of content that is seen.

What presumably plays a key role in whether or not a user's view of the world is restricted by Facebook is the **interplay of individual choices and ADM systems.**

A group of researchers from the Institute for Advanced Study in Lucca, Italy, and colleagues from other organizations have examined the diffusion of rumors, and conspiracy theories in particular, within social networks. By evaluating Facebook data, the group came to the conclusion that users have a tendency to aggregate in communities of interest, and they largely end up seeing content reflecting their common viewpoint. This results in "confirmation bias, segregation and polarization." In the context of social media, *confirmation bias* (see 5.3) leads to "proliferation of biased narratives fomented by unsubstantiated rumors, mistrust, and paranoia" (Michela Del Vicario et al. 2016).

In another study, a number of the researchers cited in the last paragraph also showed that conspiracy theorists in particular react paradoxically when they are confronted with information that challenges their viewpoint: They ignore facts that refute the conspiracy theory, or they surround themselves more closely with other like-minded individuals from their echo chamber (Zollo et al. 2015). The work done by the Italian researchers suggests one thing above all: The content that people in social networks share plays an important role in their identity management (see for example Schmidt 2016). People share things that reflect their own world view (see also An, Quercia and Crowcroft 2013).

An important issue in this regard is what happens when people encounter information that reinforces or challenges their own positions, especially in the case of individuals who are already radicalized. In a study of participants in US-based online discussion forums for neo-Nazis, Magdalena Wojcieszak (2010) came to the conclusion that users of such platforms use the content they find there to “rebut counter-arguments and generate rationales that strengthen their predilections” (ibid.). The communications researcher quotes an anonymous member of one of the right-wing forums she examined as follows: “We are existing in a world filled with influence, but are mostly immune to it because we have educated ourselves.” Education here refers, for example, to gathering arguments for why the Holocaust did not happen or why the “white race” is superior. At least for persons who have already been radicalized to some degree, the availability of environments offering extremist messages – be they Internet forums or Facebook groups – can apparently lead to further radicalization (ibid.).

Empirical research is still relatively rare that offers clear proof for or against a far-reaching impact of digital communications platforms or ADM systems on democratic decision-making processes. There are indications of constriction by content-proffering intermediaries (see for example An, Quercia and Crowcroft 2013), while several authors have called for moderation, since “there is little empirical evidence that warrants any worries about filter bubbles” (Borgesius et al. 2016).

One of the few large-scale studies on the topic comes to an ambivalent conclusion. Flaxman, Goel and Rao (2016) examined the media consumption of 50,000 Internet users living in the United States. Their main finding is:

> *We showed that articles found via social media or web-search engines are indeed associated with higher ideological segregation than those an individual reads by directly visiting news sites. However, we also found, somewhat counterintuitively, that these channels are associated with greater exposure to opposing perspectives. Finally, we showed that the vast majority of online news consumption mimicked traditional offline reading habits, with individuals directly visiting the home pages of their favorite, typically mainstream, news outlets. We thus uncovered evidence for both sides of the debate, while also finding that the magnitude of the effects is relatively modest. (Flaxman, Goel and Rao 2016)*

In contrast to Pariser (2011) and Sunstein (2001; 2008), the authors believe that “while social media and search do appear to contribute to segregation, the lack of within-user variation seems to be driven primarily by direct browsing [of online news sources]” (ibid.). The “partisan” reportage that users see comes from sources that Flaxman et al. tend to consider “mainstream,” from “the *New York Times* on the left to Fox News … on the right.” Ideologically extreme sites like *Breitbart News* “do not appear to quantitatively impact the dynamics of news consumption” (ibid.). Yet it must be noted that the data for the study were collected between March and May 2013, i.e. long before the final phase of the most recent US presidential election.

A study that appeared in 2017 comes to a decidedly different conclusion regarding the significance of *Breitbart News* and similar publications, namely that the politically extreme news outlets dictated issues to more moderate outlets like Fox News with its “hyper-partisan” reports, and even massively influenced coverage by more liberal media, “in particular coverage of Hillary Clinton.” According to the study, social media played a key role here as a diffusion channel. Information gleaned from 1.25 million stories that appeared online on social media between April 1, 2015 and the election in November 2016 reveals that “a right-wing media network anchored around Breitbart developed as a distinct and insulated media system, using social media as a backbone to transmit a hyper-partisan perspective to the world” (Benkler et al. 2017).

Widespread consensus exists among researchers working in this field that media users, especially in politically highly polarized societies like the US, are increasingly limiting themselves ideologically in terms of where they get their news. Iyengar and Hahn (2009), for example, show that Republicans in the US gravitate towards Fox News and away from more liberal media such as CNN and NPR, while Democrats tend to do exactly the reverse. Given the greater choice resulting from digital media, this mechanism could “contribute to the further polarization of the news audience.” Politically extreme publications such as Breitbart News have only become available to a wider audience because of the possibility of digital distribution. According to another study by Iyengar and Westwood

(2015), this polarization is, however, not simply a result of the Internet: The "divergence in affect toward the in and out parties – affective polarization – has increased substantially over the past four decades."

Iyengar and Westwood had participants in the study who are acknowledged Republicans and Democrats give grants to what proved to be fictitious high-school graduates. The grades of the fictitious applicants were of secondary importance, while their purported party affiliation played a key role in decisions about who would receive support: Republicans gave scholarships to Republicans, Democrats to Democrats. The often very aggressive political rhetoric used by political players in the US and the willingness to denigrate political opponents have led members of both parties to "feel free to express animus and engage in discriminatory behavior toward opposing partisans", Iyengar and Westwood write. A two-party, first-past-the-post system like the one in the US could conceivably reinforce this explicit form of polarization (see for example Trilling, van Klingeren and Tsfati 2016).

Boxell and colleagues (2017) also note that the growing polarization cannot be ascribed to the Internet alone:

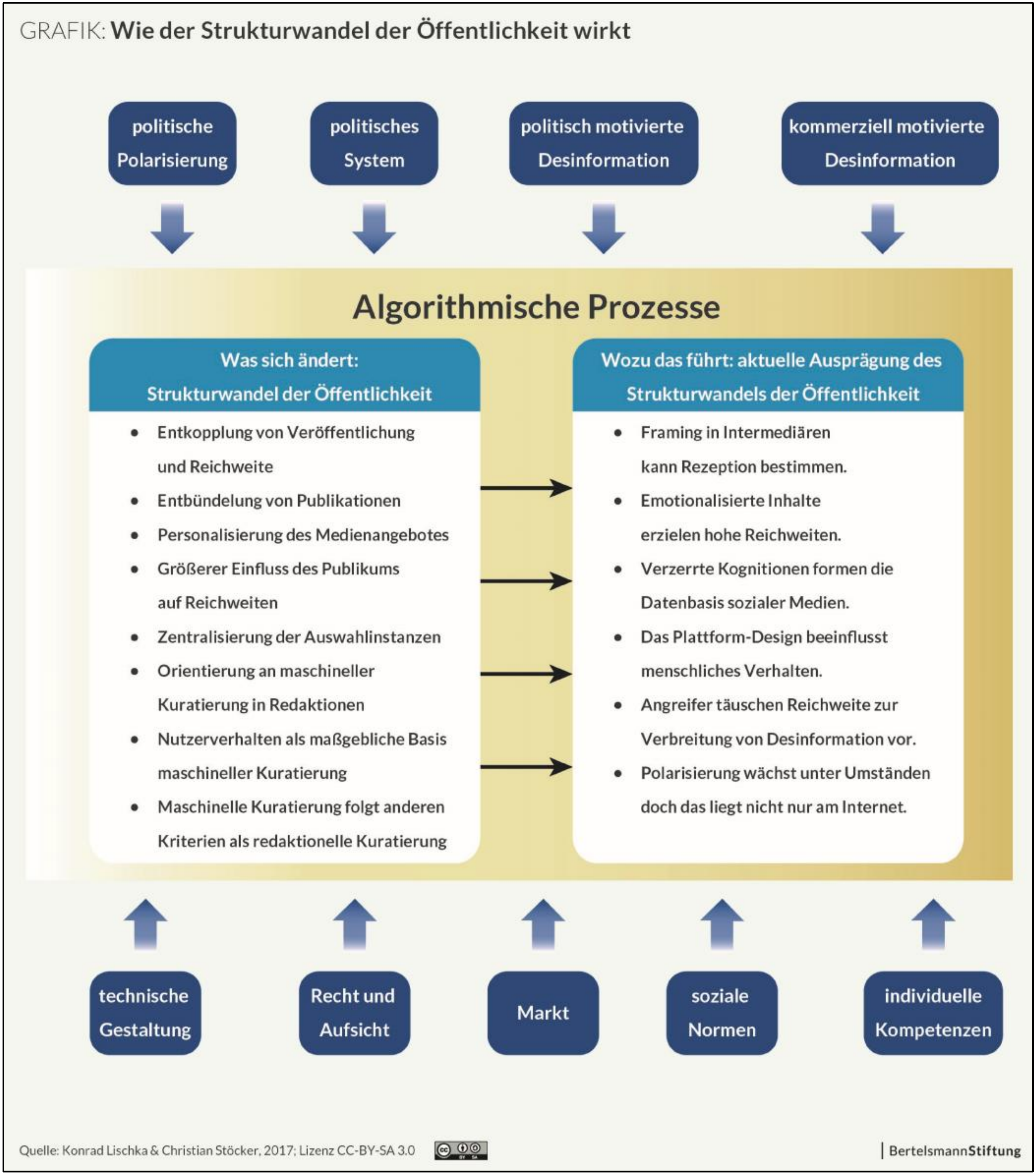
> *We find that the increase in polarization is largest among the groups least likely to use the internet and social media. A normalized index of our nine polarization measures increases by 0.18 index points overall between 1996 and 2012. Among respondents aged 75 and older, the increase is 0.38 index points, whereas for adults under age 40, the increase is 0.05 index points. Across intermediate age groups, the growth in polarization is consistently higher among older respondents. Polarization increases more for the old than the young in eight of the nine individual measures. A similar pattern emerges for groups of respondents divided by our broader index of predicted internet use. (Boxell et. al, 2017)*

In sum, there are clear indications that complex interactions exist between intermediaries, ADM systems, expanded and diversified news sources, the political system found in any given country and the phenomenon of political polarization. As Borgesius et al. (2016) emphasize: "The effect of personalised news on polarisation is conditional on the political system." A recent study from the Netherlands, which does not have a two-party system, asserts that no clear correlation exists between polarization and information reflecting individual political attitudes. In other words: Whether or not someone in the Netherlands holds extreme political views does not depend on his or her being exposed to selective media content evincing a certain political slant (Trilling, van Klingeren and Tsfati 2016).

5.7 Conclusion: Complex interactions reinforced by algorithmic processes

As the previous sections show, public discourse and public opinion result from the complex interactions of various factors as refracted through digital intermediaries. On the individual level, certain user behaviors play a role and can, in turn, be encouraged by media offerings and how those offerings are designed. Thus, media content is often shared only based on its headline and introductory text without the sharer having actually engaged with it thoroughly. Moreover, content that causes a strong emotional reaction is often passed along to others and discussed at length.

Figure 3: Digital discourse – Structural change, influencing factors and current situation



GRAFIK: **Wie der Strukturwandel der Öffentlichkeit wirkt**

politische Polarisierung

politisches System

politisch motivierte Desinformation

kommerziell motivierte Desinformation

Algorithmische Prozesse

Was sich ändert: Strukturwandel der Öffentlichkeit

- Entkopplung von Veröffentlichung und Reichweite
- Entbündelung von Publikationen
- Personalisierung des Medienangebotes
- Größerer Einfluss des Publikums auf Reichweiten
- Zentralisierung der Auswahlinstanzen
- Orientierung an maschineller Kuratierung in Redaktionen
- Nutzerverhalten als maßgebliche Basis maschineller Kuratierung
- Maschinelle Kuratierung folgt anderen Kriterien als redaktionelle Kuratierung

Wozu das führt: aktuelle Ausprägung des Strukturwandels der Öffentlichkeit

- Framing in Intermediären kann Rezeption bestimmen.
- Emotionalisierte Inhalte erzielen hohe Reichweiten.
- Verzerrte Kognitionen formen die Datenbasis sozialer Medien.
- Das Plattform-Design beeinflusst menschliches Verhalten.
- Angreifer täuschen Reichweite zur Verbreitung von Desinformation vor.
- Polarisierung wächst unter Umständen doch das liegt nicht nur am Internet.

technische Gestaltung

Recht und Aufsicht

Markt

soziale Normen

individuelle Kompetenzen

Quelle: Konrad Lischka & Christian Stöcker, 2017; Lizenz CC-BY-SA 3.0

| BertelsmannStiftung

These signals are included in the algorithmic ranking of content, thereby increasing the probability that content which is particularly suitable for rousing emotion will reach an even wider audience. Cognitive distortions known to psychologists such as the availability heuristic presumably interact with mechanisms of this sort: How users view the world is potentially determined by content that neither they nor the people who shared it have actually viewed completely. The goals and principles underlying intermediaries’ design choices reinforce such distortions and any actions carried out without any notable cognitive engagement. One key design principle is ensuring that all desired activities are executed as simply as possible; any barriers that might lead to a cognitive deceleration are identified and removed, if possible.

For commercial or propagandistic reasons, external actors become involved in this process using automated mechanisms such as bots which are designed for manipulative purposes, ensuring that certain content and ideas receive more attention than would otherwise be the case. Ultimately, the interplay of individual user choices, algorithmic ranking systems and psychological factors leads, at least among a subset of users, to increased polarization of the content users see and their social and political attitudes. ADM processes are only one factor in this complex construct, but one that interacts with all other factors, thus potentially magnifying both shortcomings in human cognition and concerted efforts to manipulate the relevant processes through the use of technology.

6 How it should be: Principles for guiding public discourse

The above analysis of the main factors causing a structural change in public discourse suggests that intermediaries are exerting greater influence on the formation of public opinion and social debate. Using two intermediaries as examples, Facebook and Google, we have ascertained that known ranking signals and the method used for evaluating relevance for each individual user essentially serve as a general guiding principle for the structuring of online content. We now want to compare this principle with those used by editorial media in Germany to form public opinion. We use the following chapter to identify these principles, focusing on those that underlie the norms used to inform professional ethics and in institutional and regulatory contexts.

A comparison of the guiding principles known to be deployed by particularly relevant intermediaries and those used by editorial media is, we believe, not only appropriate but necessary. For many users, intermediaries fulfill functions comparable to those addressed by editorial media – functions, for example, that are relevant to our research, such as providing guidance on political, economic and cultural happenings, and facilitating participation in the social discourse. The need for comparison remains even if some users focus on other functions.

In terms of the principles guiding public discourse, it seems more helpful to analyze the constituent parts of what is communicated than its means of transmission. No fundamental differences are to be found here between intermediaries and editorial media which might prohibit a comparison of principles. A suitable system for doing this has been proposed by Jan-Hinrik Schmidt (2016: 286), who differentiates between three modes of communication:

- Publication (*one to many*): e.g. editorial media targeting a wide online audience (Spiegel Online)
- Conversation (*one to one, one to few*): e.g. chat groups
- Human-machine interaction: e.g. databank queries, search machines

If we look at the modes of communication employed by the platforms analyzed in this paper, we see an increasing convergence: *One-to-many communication* predominates among editorially curated media, although they also engage to some extent in *one-to-few communication* in their forums and social media channels (e.g. responses to criticism or reader contributions in comments).

At first glance, *one-to-few communication* predominates in social networks. People share their personal experiences, mentioning one-to-many sources while adding a personal comment, a recommendation or criticism for their social cohort in the network. Editorially curated media, however, also engage in one-to-many communication using their social media accounts. When a site like *Buzzfeed* or the *Huffington Post* posts videos or stories directly on Facebook, this mode of communication can hardly be differentiated from publication using the site's own app or website. Both modes of communication are, moreover, inseparable from human-machine interaction when it comes to intermediaries such as Facebook: A user only sees the comment posted by a friend about an article on Spiegel Online because the ADM ranking process allowed it.

Since the type of communication employed by editorially and algorithmically curated media and intermediaries is converging, there is much to suggest that their guiding principles should converge as well. We are not arguing for principles that are identical or for having one system adopt principles used by another. A comparison and analysis are warranted, nonetheless. Editorially curated media and algorithmically curated intermediaries differ in multiple respects. However, when intermediaries and editorial media assist people by providing guidance on political, economic and cultural events and by helping them participate in the social discourse, an argument can be made for their adhering to similar principles. We will now outline, using select principles, what German society expects of its public discourse. According to these principles, social debate is more than the aggregation of individual choices, and quality is more than just an exact reflection of individual preferences. It is a process that involves more than just ensuring each and every individual preference is satisfied. These principles, derived from Germany's Basic Law, illustrate what a positive structuring of public discourse might look like. They illustrate what a focus on participation and participation itself both mean when it comes to public opinion and social discourse and what a positive structuring of public discourse could be like if it is designed to include everyone.

6.1 Freedom

Germany’s Basic Law guarantees freedom of expression and freedom of information as basic rights, from which related rights can be derived, such as the right to a positive structuring of public discourse. Such a structuring is characterized by qualities such as respect for the truth and integration, instead of simply a “negative” structuring, i.e. the absence of government interference in freedom of expression (see Section 6.3.2). The two positions can be contradictory: The freedom of expression enjoyed by journalists and editors at Germany’s public broadcasting stations is, for example, constrained by the principles governing the broadcasters’ operations, i.e. the principles meant to achieve a positive structuring of public discourse – in particular, the right the public has to comprehensive and truthful information, as laid out by Germany’s Constitutional Court based on the Basic Law. As stated in a ruling by the Constitutional Court: “It is the legislature’s responsibility to resolve these contradictions” (Bundesverfassungsgericht 1981: 321).

In other words, in certain cases freedom of expression is limited, for example to ensure freedom of information. The absence of interference does not ensure a positive structuring of public discourse. This interpretation of freedom as a basic principle rests on a specific understanding of basic rights: “What is crucial here is the view that individuals are free, above all, because they live in a society that, as the result of certain prerequisites and the limitation of certain actions, ensures the highest degree of potential freedom for all its members” (Heesen 2016: 52). The Constitutional Court has explicitly stated that when a medium plays a significant role in forming public opinion, the mere right to ward off government influence and be subject to the free play of market forces is not sufficient: “In light of this situation, it would not be compatible with the constitutional requirement for ensuring broadcasting freedom to merely exclude government interference and allow broadcasters to be exposed to the free play of forces” (Bundesverfassungsgericht 1981: 323).

With that, the concept of freedom of expression and information as laid out in Germany’s Basic Law and as interpreted by the country’s Constitutional Court differs fundamentally from the concept as it applies to media in the United States. The principle of freedom of expression is very wide-ranging in the US, having been expanded by the Supreme Court in a judgment (US Supreme Court 1969: Brandenburg v. Ohio 395 U.S. 444) that held that general public endorsement of illegal acts and violence (in the case of a Ku Klux Klan member versus African Americans) is permissible insofar as it is not a direct appeal for others to take immediate action. Public discourse is, in this sense, a marketplace of ideas that should largely be driven by competition among the players within it. It is predicated on the logic that counterarguments are the preferred defense against lies, racist comments or incitement. Those who are attacked must defend themselves. More, not less free speech is the appropriate response – that is the core argument advanced by Justice Louis Brandeis in his concurring opinion to a 1927 Supreme Court ruling (in a case involving a founding member of the Communist Labor Party):

> *If there be time to expose through discussion the falsehood and fallacies, to avert the evil by the processes of education, the remedy to be applied is more speech, not enforced silence. Only an emergency can justify repression. Such must be the rule if authority is to be reconciled with freedom. (Justitia US Supreme Court 1927: 377)*

Facebook has made reference in Germany, as elsewhere, to the possibility of responding to freely expressed opinions with counter-opinions (Herbold 2016).

6.2 Diversity

Public opinion can only be formed freely and thoroughly if "the constitutive diversity of opinion [required] for a free democracy is represented" within the sum total of the offerings made available by the media (Bundesverfassungsgericht 1981: 321). Dörr and Deicke, for example, see this call to pluralism as one of the most significant structural requirements that follow from the democratic principle laid out in the Basic Law:

> *Diversity presupposes the existence and communication of various opinions and opinion-related information, so that they might serve as the foundation for a communications process among the public. It is even the foundation for the development of the individual and political autonomy presupposed by Art. 1 paragraph 1 of the Basic Law, since this development requires the individual to have knowledge of facts, as well as opinions and attitudes. (Dörr and Deicke 2015: 13)*

Two models of pluralism greatly influence the implementation of this basic principle in Germany:

- *Internal plurality:* A service must be inherently balanced in its diversity (e.g. services offered by a public broadcaster).
- *External plurality:* The entire range of services must be diverse in their entirety (e.g. the press).

This is the principle underlying broadcasting requirements in Germany, such as requirements for regional programming that depend on the size of a broadcaster's reach, and air time that must be reserved for independent broadcasters.

6.3 Truth

6.3.1 Press: Truth-finding as a process and guiding ethical principle

In almost all codes of conduct used by European journalists, the truth is seen as a "key and crucial element" (Bentele 2016: 63), for example as laid out in Section 1 of the press code maintained by the Deutscher Presserat (German Press Council): "Respect for the truth, preservation of human dignity and accurate informing of the public are the overriding principles of the Press. In this way, every person active in the Press preserves the standing and credibility of the media" (Deutscher Presserat 2017: 2).

The principles are the starting point from which concrete applications follow, as discussed below. What is important here is an acknowledgement of the fundamental nature of this norm: The code of conduct does not state that the truth can be reported in every instance. The phrase "respect for the truth" instead formulates a goal that reporting processes should be optimized to achieve.

The press code used by the German Press Council contains a number of concrete examples further elucidating this principle, such as:

- "Unconfirmed reports, rumors or assumptions must be reported as such" (ibid.: 3).
- "If a publication concerns the publisher's own interests, this must be clearly identifiable" (ibid.: 5).
- "The Press may call a person a perpetrator [during investigations and court proceedings] if he/she has made a confession and there is also evidence against him/her or if he/she committed the crime in public view" (ibid.: 7).

These concrete applications are the result of decisions made by the council following complaints that the press code may have been violated. Using its adjudicatory practice, the council continues to develop the code, further concretizing and updating the profession's ethical principles. The council is sponsored by publishers' and journalists' associations. It was established in 1956 as a system of professional self-regulation in order to forestall a planned national press law. The members of the council's complaints board are appointed by the council's sponsors (in plenum) in keeping with the concept of the council as a self-regulatory ethics organization.

The approach taken by the press code and press council presupposes that identifying the truth is a process, and providing accurate information is clearly predicated on using certain tried-and-true methods of acquiring knowledge. This is reminiscent of scholarly research: Statements must be verifiable, comprehensible and falsifiable so that they might be considered true until refuted.

6.3.2 Broadcasting: A basic right to accurate information

In its third ruling on broadcasting, the German Constitutional Court made it clear that, given the right to freedom of information as laid out in the Basic Law (Section 5 paragraph 1 sentence 1 clause 2), a "right to comprehensive and accurate information" exists (Bundesverfassungsgericht 1981: 321). The Constitutional Court has ruled that freedom of broadcasting means more than just non-interference in the traditional sense; mere freedom from government encroachment does not necessarily make a "free and comprehensive formation of public opinion" possible (ibid.: 320). What is required instead is a "positive structure."

This structure must be put in place by the legislature. In terms of Germany's broadcasting system, what is paramount, according to the ruling by the Constitutional Court, is that "free, comprehensive and accurate opinion is ensured in the sense laid out here" (ibid.: 321). The country's legally mandated programming principles thus require its public broadcasters to achieve this goal. For example, the law governing broadcaster WDR mandates in Section 5 paragraph 4 that WDR "must respect the truth."

Broadcasting councils made up of representatives of different social groups (varying from state to state) are responsible for ensuring these programming principles are upheld by public broadcasters. The concept underlying the councils' composition is different from the one used by the press council in that the councils are not meant to develop ethical guidelines for members of the profession, but to carry out a controlling function for society as a whole. To that degree, the broadcasting councils should reflect society's make-up, an ideal that has not been achieved, as Donges pointedly notes: "If one looks at the bodies within the public broadcasters, for example, one sees that society in general is not represented there, but its older, masculine and high-status subsection" (Donges: 97).

Outsiders may submit complaints within this system as well. In contrast to the press council, however, the broadcasting councils cannot further refine or update the broadcasting principles, they can only respond to each individual case.

6.4 Integration

The ideal of a deliberative public sees social integration as a core characteristic of any positive structuring of public discourse. Deliberation – the exchange of arguments within the social discourse – is meant to lead to a balancing of viewpoints and consensus. This ideal envisions deliberation "through which people recognize the potential diversity of positions and opinions and thus enter into a dialogue with each other whose goal is achieving understanding" (Schmidt et al. 2017: 26).[5]

The ideal of a deliberative public sees the clash of opinions not as a repetitive, ongoing cycle, but as a process with various phases of development. The prerequisites for ensuring the process succeeds are exchange, rationality and argumentation:

[5] "This characterization is based on ideas advanced by Jürgen Habermas (1981), without doing justice to the complexity of his theory of the public sphere" (Schmidt et al.: 26).

> *The idea of public deliberation is precisely that of striving for consensus through dissent. Criticism and problematization are just as much a part of the discourse as the attempt to rationally overcome dissent. And we can hardly speak of a serious discussion if the participants do not intend in some form or other to convince each other with their arguments and, when appropriate, allow themselves to be convinced. (Peters 2002: 31 f.)*

The ideal of social integration through public deliberation is reminiscent of the mandate Germany's public broadcasters have of promoting social cohesion at the federal and state level. They are supposed to do this by providing "a comprehensive overview of international, European, national and regional happenings in all key areas of life" (die medienanstalten 2016: 19).

This ideal fundamentally informs the discussion of the function and impact of intermediaries in Germany, for example the discussion of filter bubbles and echo chambers. When it comes to these phenomena, the question of diversity must be posed anew: Although intermediaries offer access to a great diversity of content and source material, they use personalized criteria to prioritize only a small and homogenous subsection of all the information potentially available. The question arising from the guiding principles discussed here is: Which heterogeneity is required in the prioritizing done by intermediaries for social integration to succeed, assisted by public deliberation?

Social integration is also a guiding principle for public discourse on the European level. For example, the High Level Expert Group on Media Diversity and Pluralism has warned that, because of personalized media content, people could be exposed to less diverse opinions, which could in turn compromise public debate and the democratic decistion making process (Vike-Freiberga et al. 2013). The Council of Europe has also issued warnings regarding the impact on public opinion of algorithmic ranking done by search engines (Council of Europe 2012a) and social networks (Council of Europe 2012b).

6.5 Conclusion: Applicability of guiding principles to algorithmically structured public discourse

The following section discusses the degree to which the standards for public discourse outlined above as guiding principles and ascribed to editorial media could also be applied to algorithmically structured services. We analyze if these principles can be deployed, and how they can be interpreted, when intermediaries guide people as they engage with current political, economic and cultural events and participate in social discourse.

Our ideas are based on findings relating to structural changes in public discourse (see Chapter 4) and the impact of algorithmic processes on, and their interaction with, a variety of factors (see Chapter 5): the availability via such services of large volumes of media content that is, in some cases, ideologically extreme; the manner in which some users select and recommend content; the ranking of content based on signals that are sometimes unclear to outsiders; the design decisions made by intermediaries and intermediaries' surreptitiously conducted experimental variations; the use of automated systems meant to manipulate public discussion; and other external factors such as the political system found in each country.

6.5.1 Freedom as the freedom to determine relevance

As seen in analyses of their services, intermediaries such as Facebook and Google structure public discourse according to their own principles and values. These design decisions (examined in Section 4.3) are, in our opinion, ultimately editorial in nature, and include:

- Criteria according to which the intermediary weights relevance
- Signals it uses to operationalize and measure these criteria
- The weight it gives individual factors

The weighting and selection of opinions, information and statements relating to current events is a form of expression, something that Volokh states concisely in an acclaimed white paper commissioned by Google:

> *For example, a newspaper also includes the materials that its editors have selected and arranged, while the speech of DrudgeReport.com or a search engine consists almost entirely of the selected and arranged links to others' material. But the judgments are all, at their core, editorial judgments about what users are likely to find interesting and valuable (Volokh and Falk 2012: 4 f.).*

In this sense, intermediaries are *gatekeepers of information diversity*, as Morganti et al. (2015) put it. Philip Napoli (2014a) goes further and speaks of *automated media*.

In the case of intermediaries, users' influence on structure is greater than for an editorially curated medium. Yet the signals sent consciously and unconsciously by users are dependent on the intermediary's decisions about which signals are possible and which will be reinforced. For that reason, user choices are not an independent variable (see Section 5.4), but are constantly interacting with intermediaries' efforts to shape content (see Section 4.4).

Given the relevance of intermediaries for forming public opinion (see Section 3.2), it seems appropriate to balance the intermediaries' freedom of expression and the public's right to comprehensive, accurate information (see Section 6.1). No conclusive framework has been devised for how those intermediaries providing information are to be included in the system developed by Germany's Constitutional Court for ensuring freedom of information and, subsequently, freedom of broadcasting. At the same time, recent rulings provide a few clues, which Fetzer summarizes thus:

> *Content and services disseminated via the Internet, directed at the general public and relevant for forming individual and collective opinions are covered by content protected by Article 5 paragraph 1 sentence 2 alternative 2 of the Basic Law [freedom of the press and broadcasting, alternatively free development of one's personality]. (Fetzer 2015: 8)*

6.5.2 Diversity as diversity of algorithmic processes

Applying the principle of diversity to intermediaries reveals an implicit assumption: Diverse offerings will be used diversely. That is the rationale underlying the requirement that various viewpoints be presented in the media (cf. Burri 2013). Certain broadcasting regulations in Germany focus on threshold values of actual use, but beyond that there is no actual engagement with the relationship between how the various products and services on offer are used. Current empiric research provides no conclusive answer to the question of how usage in the digital sphere has evolved as the range of offerings has increased. There are no empiric studies on the diversity of user behavior. We do not know, for example, how large and how varied personalized News Feeds are. Little research has been carried out in this area, among other reasons because it is difficult for outsiders to investigate how intermediaries operate.

These questions all apply on the level of how information is used, i.e. editorially curated content. How much content is there, and how diverse is it? A second level of diversity must be differentiated here, one that seems to be exceedingly relevant for intermediaries: the diverse processes used for selecting and ranking content. Every intermediary conceptualizes relevance and weights signals accordingly. This journalistic activity is not controlled by users. If users would like their Facebook News Feed to be ranked using an algorithmic process, they cannot choose from multiple processes. As it applies to intermediaries, three levels of diversity must be considered:

- Algorithmic ranking (process)
- Products and services (output)
- Use (outcome)

It should be noted that process and output influence use – and use can affect ranking and output. Algorithmic processes make this interaction possible and mediate between all levels.

The narrowing of process diversity cannot be equated with phenomena known from editorially curated media. Here it is not a case – or not currently a case – of partisan viewpoints (e.g. conservative vs. liberal intermediaries) serving as a basis for prioritization. Instead, it is a case of whether the range of relevance criteria used by intermediaries is limited or not. For some intermediaries, posts that call forth negative emotions have a measurably greater reach than more neutral posts (see Section 5.4). This is the result of how the platforms are designed: User interfaces are optimized for easily accessible reactions that promote a certain form of interaction, and the algorithmic system uses these interactions as a sign of relevance. Other processes for measuring relevance would produce other results and would not, in sum, favor negative posts to the same degree. This, however, would only be possible if a greater selection of algorithmic processes were available for use.

Less diversity is undoubtedly present given that very few intermediaries aim to achieve an extremely wide reach and extremely high user-engagement times, and given that each of these providers only permits and makes use of its own processes for ascertaining relevance.

Thus, a more broadly based analytic foundation is required for interventions that can promote diversity among intermediaries than is needed for ensuring diversity at traditional broadcasters and in the press. Diversity must also be examined and addressed in terms of the diversity of algorithmic ranking processes (process level). Moreover, another approach is needed for analyzing actual use and the actual content made available by intermediaries than for analyzing editorial media. If researchers are to investigate how algorithmic-driven personalization tools impact the diversity of content and use, they must be able to access the relevant underlying data.

6.5.3 Respect for the truth

If algorithmic tools for determining relevance are used to structure public discourse, then they must be assessed based on whether or not they evince respect for the truth. There are no cogent arguments for not applying the principles found in the Basic Law and in judicial rulings to intermediaries that help form public opinion. The key question is: To what degree must these providers commit themselves to respecting the truth? The answer is clearly more than Facebook does in its *News Feed Values,* which do not reference this principle at all (Facebook 2016b), but probably not to the extent that editorially curated media are regulated or regulate themselves.

The discussion about strategically posted, viral fake news in Germany shows that intermediaries can also be held to this standard. We do not use the term “fake news” but refer instead to “Falschinformationen”: false information that is “intentionally produced and diffused, and formulated in a way that takes advantage of the logics used by social media” (Reinbold 2017).

When it comes to intermediaries and respect for the truth, one must also ask to what degree the evaluated signals make it possible to falsify widespread user reactions (see Section 5.5). Not only must ranked content be held to the standard of respecting the truth, so must the reactions of the (purported) public as disseminated by intermediaries.

6.5.4 Social integration

This principle is the most difficult to transfer to intermediaries from editorially curated media. By definition, it conflicts with the personalization of ADM processes, even if concerns about filter bubbles and echo chambers are only partially justified given current empiric findings (see Section 5.6). In terms of these phenomena, the principle of social integration pertains more to the outcome level than the principle of diversity does, since integration refers to actual use and the quality of the resulting reactions. If, for example, an intermediary’s offerings lead to users being confronted with diverse content, but this content increases the existing polarization among social groups, then the relevant algorithmic process promotes diversity, but not integration.

The integrative effect of a deliberative public discussion does not relate solely to personalization, selection and prioritization, but also to the discourse that ensues. This level is often absent from debates about filter bubbles and their consequences: All intermediaries which promote discourse through the social components of their platforms have a direct influence on the quality of social discourse as a result of their efforts to shape and facilitate interactions. Numerous findings suggest that this has a considerable impact on user behavior (for an overview see Diakopoulos 2016b).

A platform that designs comments and possibilities for interaction in a way that serves the ideal of deliberative public discourse would perhaps evaluate other signals and use them as success metrics. A fundamental difference in design can be seen in those optimized for balancing consensus, such as *MediaWiki* (the operating software used by Wikipedia) and *LiquidFeedback*, as opposed to those that promote unlimited ongoing comments. In their analysis of Germany's Pirate Party, Lobo and Lauer provide insight into how an intermediary's design can influence public opinion:

> *In about 2010, Twitter began functioning as an emotional vent for the Pirate base, as a swingboat for maximizing outrage…. The main social media platforms are perfectly suited for mobilization, for creating groups and for disseminating information in the blink of an eye – but not for productive, political discussion; that is why such arenas for debate must be structured completely differently than Twitter and Facebook are. What are needed are unique, functional platforms that promote digital democracy. (Lobo and Lauer 2015: loc. 2436 ff.)*

These days, Facebook reaches more than half of the 23.3 million Germans who, on an average day, keep abreast of current events using algorithmically curated sites. They not only learn what has happened that day, but also experience the purported social discussion about it. The current debate about issues such as hate speech and disinformation in social networks suggests that the social discussion taking place there is often not seen as productive and constructive, but, on the contrary, as aggressive and often repellent. When considering social participation, this seems to be a missed opportunity: Never before in history have so many people gathered in what are for the most part public spaces to participate in debates on a variety of topics, including political and social issues. Yet the outcome is not currently seen as enriching deliberative democracy, but largely as hampering and even endangering it.

7 What we can do: Approaches to intervention

As the previous findings show, the guiding principles of freedom, truth, diversity and social integration are fundamentally suitable for analyzing intermediaries which provide insight into political, economic and cultural happenings and which facilitate participation in social discourse. The application of these principles to the digital sphere and its mediatory algorithmic processes needs, to some extent, a different analytic approach and a different operationalization.

Options for making these guiding principles actionable and effective can essentially be found in two areas:

- Algorithmic decision-making
- Human perception

Both areas have a significant impact on the formation of public opinion in the digital sphere (see Section 4.1) and are closely interconnected. According to current findings, causality is rarely clear: Cognitive distortions influence which signals people send when using intermediary sites; ADM processes evaluate these signals as proxies for relevance, thereby potentially scaling the distortions seen in the data, which then become the basis for rankings of certain media content and, in turn, user responses. In addition, relevance signals can be manipulated externally through the deployment of technology. In terms of interventions, we differentiate between three levels applicable to both areas:

- Macro level: Social framework, guiding principles, regulation, government actors
- Meso level: Businesses, public institutions, self-regulatory bodies
- Micro level: Media users, developers, journalists

If we employ this matrix to analyze (presumably) clearly known phenomena such as *fake news,* multifaceted, shifting factors can be seen underlying the phenomena on various levels. For example, it is easier to use intermediaries to reach a wide audience without deploying established media brands. It is easier to fake a critical mass of interest for certain content using manipulated user profiles. And it is easier, with optimized content, to make use of cognitive distortions to generate a critical mass of user reactions. This range of interdependent factors suggests that one intervention alone will not suffice to rectify the situation and that a systematic examination is necessary to identify appropriate responses (see Table 5).

Table 5: Factors for disseminating consciously posted fake news on intermediaries

Level	Algorithmic decision-making	Human perception
Macro	Lack of diversity of ADM processes Insufficient evaluation	Widespread impact of intermediaries that are designed using principles which promote specific cognitive distortions
Meso	Selection and weighting of signals Lack of adherence to the principle of “respect for the truth” among intermediaries Lack of independent bodies (e.g. analogous to the German Press Council)	Lack of measures for reducing the impact of cognitive distortions
Micro	Identification of individual “fake news” posts as highly relevant Decisions made by developers and evaluators	Diffusion of fake news; possible causes: misinterpretation by ADM processes, dominance of other needs besides the desire for information (reinforcement of group membership, user identity, etc.)

Source: The authors.

In the following, we present approaches for introducing potential improvements in the areas shown in the table above. We define "improvements" as ensuring that the principles discussed above relating to public discourse are given greater consideration when algorithmic processes are used to structure public discourse. We do not develop and recommend specific solutions, but identify instead effective approaches for further investigation (see Table 6).

Table 6: Approaches and interventions

Level	**Algorithmic decision-making**	**Human perception**
Macro	Using guiding principles to increase social understanding Ensuring a broad range of relevance metrics Ensuring external auditing	Research and external evaluation
Meso	Promoting external evaluation Development of alternative relevance metrics Institutional anchoring and dynamic development of guiding principles (e.g. professional ethics, "Algorithm Council" as self-regulatory body, impact review of ADM changes)	Ensuring a range of forums for user interaction Inoculation
Micro	Anchoring of guiding principles for individuals (e.g. in educational programs) Promoting individual skills for dealing with algorithmic systems	Diverse individual input Sensitization to the consequences of cognitive distortion

Source: The authors.

To have a social debate about the type of public discourse society wants, ADM processes are required that are transparent, explainable, verifiable and correctable. Also needed are responsible actors who are skilled in using the relevant applications. If we do not know where we are headed when we optimize ADM processes and what the goal is, it will be impossible to discuss whether these processes are good or bad in terms of participation. It will also be impossible to know which principles can provide concrete guidance on promoting participation, thus shaping public discourse in a positive sense.

7.1 Shaping ADM processes to promote participation

7.1.1 Using guiding principles to increase social understanding (macro level)

A social agreement on the principles needed to guide public discourse in the digital sphere is the precondition for shaping the discourse for the common good. In Germany, however, the various intermediaries and ADM processes have not been anchored in the country's social agreements for ensuring diversity (see Section 7.1.2).

There has been no ruling by the Constitutional Court on the degree to which *positive efforts* by broadcasters to ensure freedom of information must also apply to the digital sphere. It has still not been definitively clarified in which instances federal or state institutions are responsible for enforcing the country's telecommunications and media laws. Federal or state institutions could precipitate a clarification by actively addressing this nebulous situation. If diversity is to be promoted among intermediaries on the macro level, then greater awareness is required about what limits diversity in this area. Also needed are effective measures for shaping what are in essence public services in the digital sphere – followed by legislative changes, if necessary.

7.1.2 Ensuring a broad range of relevance metrics (macro level)

If only few conceptions of relevance exist in the digital sphere coupled with few corresponding algorithmic processes within the social discourse, then society's potential diversity is reduced accordingly. What is required here is an enlarged understanding of diversity and how it can be ensured, i.e. diversity of algorithmic processes, of provider models and of goals for structuring public discourse in the digital sphere.

One fact is key for identifying approaches that can increase diversity on the macro level: There is no technical or functional reason that requires people to use only one algorithmic process (i.e. that of the operator) for assigning relevance when engaging with a website. This is predominantly the case for intermediaries in the digital sphere, yet when one looks back at the history of the World Wide Web and the Internet, it is a comparatively recent development. In the mid-aughts, open standards such as *rich site summary (RSS)* and *Outline Processor Markup Language (OPML),* not to mention simple aids such as CSV address-book files, made it possible to transmit collective source material and social links. That allowed users to identify their own structured signals ("Which sources do I like and which people do I know?") when engaging with various providers of algorithmically structured services. The degree to which portability is still possible today can be seen in services such as *Nuzzel* which, based on *application program interfaces (APIs)*, avail themselves of the social infrastructure provided by Twitter and Facebook to offer, despite the infrastructure's very limited usability, a clearly alternative selection of information.

Other possibilities exist. Theoretically, a social graph or search index could serve as infrastructure and be evaluated without restriction using alternative processes and goals. As a result, intermediaries could deploy models that promote diversity both within their own operations and externally. What is required is relinquishing the idea that the relationships captured by an intermediary's infrastructure (e.g. between accounts or between websites indexed on the Internet) must be evaluated using one sole process.

If intermediaries are to ensure diversity, other aspects (beyond having a range of editorial media on offer) must be considered that have rarely been examined until now: the diversity of processes that can be used for structuring public discourse, and the diversity of actual use. The latter concept must be considered in greater detail. Information about individual stories and posts is just as relevant as the overall reach of select media products, as captured, for example, by Germany's MedienVielfaltsMonitor (MediaDiversityMonitor). This is true since only on the level of specific posts can an understanding be gleaned of how certain effects such as filter bubbles limit diversity (see Section 5.6). When it comes to how the algorithmic structuring of public discourse affects diversity, the relevant media oversight authorities are ineffective given the current lack of research and expertise.

Mittelstadt outlines the areas where interventions are needed and could conceivably be deployed (for more on external accessibility and evaluation see Section 7.1.2):

> *One possibility for managing algorithmic auditing would be a regulatory body to oversee service providers whose work has a foreseeable impact on political discourse by detecting biased outcomes as indicated by the distribution of content types across political groups. (Mittelstadt 2016b: 4998)*

Having not yet taken place, a social debate is needed to determine which instruments can ensure diversity on this level. The spectrum of possible, and as yet undeployed, instruments is broad: No proposals, for example, have been made for creating a public framework to govern intermediaries which are not designed to maximize income by promoting reach and advertising. Ideas are still lacking for how alternative models of service provision can be created that both serve business interests and ensure diversity. In contrast, in the Netherlands the *Stimuleringsfonds voor de Journalistiek,* funded by the Ministry of Culture, shows through its early support of services such as *Blendle* that possibilities indeed exist. Approaches used in the third sector for intermediaries have yet to be tried in Germany; in the United States, the *Wikimedia Foundation* and the *Mozilla Foundation* are prime examples demonstrating that alternative models can result in more diverse products and services.

7.1.3 Ensuring external auditing (macro level)

Researchers who do not work for the relevant intermediaries face a number of problems when investigating and evaluating processes at algorithmically curated online services. *Application programming interfaces (APIs),* for example, were designed to assist external developers create games or other applications which run on sites such as Facebook and which sometimes access information stored by the site operator such as a user's list of friends. Theoretically, APIs can also be used by researchers to carry out analysis or test hypotheses based on the large datasets collected by platform operators. Practically speaking, however, access to these interfaces for such purposes is often limited.

> *In contrast to the Twitter API, using the API of Facebook means that researchers have to request permission to collect nonpublic data from the participants through a Facebook app. (Lomborg and Bechmann 2014)*

One explanation for the restrictions confronting researchers is provided by Puschmann and Ausserhofer:

> *Facebook has greatly restricted access to user data through the API out of privacy concerns, as have other platforms. When dubious actors acquire large amounts of data that are clearly not used for the API's intended purpose, this often leads to a tightening of policies by the API's operators, if only because providing and sustaining the performance of an API is not trivial computationally. (Puschmann and Ausserhofer 2017 in: Schäfer and van Es 2017)*

Over the years, Twitter, too, has limited the options for accessing its APIs, as noted by Puschmann and Ausserhofer:

> *Initially offering broad access to data in the first years of its operation in order to encourage development of derivate services, such as software clients for unsupported platforms, the company reasserted its control by making access to data more restrictive in several successive steps over recent years. (Puschmann and Ausserhofer 2017 in: Schäfer and van Es 2017)*

Puschmann and Burgess use Twitter to summarize the difficult situation regarding availability of user and operational data on online platforms:

> *Platform providers and users are in a constant state of negotiation regarding access to and control over information. Both on Twitter and on other platforms, this negotiation is conducted with contractual and technical instruments by the provider, and with ad hoc activism by some users. The complex relationships among platform providers, end users, and a variety of third parties (e.g. marketers, governments, researchers) further complicate the picture. These nascent conflicts are likely to deepen in the coming years, as the value of data increases while privacy concerns mount and those without access feel increasingly marginalized. (Puschmann and Burgess 2013, in: Weller 2013)*

Another problem ensues from the terms and conditions set by intermediaries. Purely to prevent manipulation, for example by bots, Facebook, among others, forbids the creation of profiles, known as *sock puppets,* not associated with an actual human user. This, however, greatly limits the possibilities researchers have of observing and checking how ranking algorithms work by isolating variations.

More empirical work is needed if an evidence-based discourse is to take place on the influence of algorithmically structured intermediaries. Currently, researches can only audit and evaluate intermediaries' operations to a limited degree. What is required are:

- Transparency: Information on the data used, how it is weighted and the relevant impacts
- Explainability: Making decisions (and criteria) comprehensible as a prerequisite for discourse
- Verifiability: Auditing and assessment of algorithmic decisions by independent third parties

Much can be learned about the social evaluation of ADM processes from the debate on the subject that has taken place the United States, one that has progressed further than the German debate. One aspect that must be concretized is what should be expected of intermediaries with a wide reach. Only if clearly formulated quality standards are in place will it be possible to ascertain the degree to which intermediaries fulfill, on a detailed level, the relevant principles promoting public discourse.

Clear transparency requirements are set out by Diakopoulos (2016a: 60), who identifies the following five categories for external evaluation of information on how ADM processes work:

- Where are people involved?
- Which data does the process reference?
- What does the model look like?
- Which conclusions does the process come to?
- Where and when is the process used?

7.1.4 Promoting external evaluation (meso level)

A number of suggestions have been made for facilitating external evaluation, although their execution must still be worked out. One example is progressive transparency vis-à-vis institutions and the public. In cases where providing complete transparency could prove disadvantageous, Tutt proposes making processes transparent successively:

> *On the lighter end, an agency could require that certain aspects of certain machine-learning algorithms (their code or training data) be certified by third-party organizations, helping to preserve the trade secrecy of those algorithms and their training data. Intermediately, an agency could require that companies using certain machine-learning algorithms provide qualitative disclosures (analogous to SEC disclosures) that do not reveal trade secrets or other technical details about how their algorithms work but nonetheless provide meaningful notice about how the algorithm functions, how effective it is, and what errors it is most likely to make. (Tutt 2016: 18)*

Algorithmic processes that are continually enhanced must be evaluated differently than static processes. Sandvig et al. (2014) offer five approaches for researching such systems externally:

1. Code audit
2. Non-invasive user audit
3. Scraping audit (extracting data from accessible sources)
4. Systematic tests using sock puppets (see above for the difficulties relating to this approach)
5. Collaborative audit with volunteers or paid testers / Crowdsourcing audit

Sandvig's crowdsourcing approach is reminiscent of the representative sample of households used in Germany to ascertain television ratings. As this analogy shows, the expert debate on researching and assessing ADM processes in general can be used to identify and adapt practical approaches for auditing intermediaries' algorithmic processes.

These examples are by no means an exhaustive representation of the now extensive debate on researchers' use of the data collected by operators. They do, however, shed light on possibilities for regulation of algorithms. Transparency requirements might have to be laid out to ensure the necessary datasets are made available. Research-friendly rules and options for accessing operators' databanks would go a long way towards increasing the transparency of those ADM systems impacting public discourse, without requiring that the algorithms themselves be made public, a move that would not be advisable for a number of reasons (see for example Diakopoulos 2016a).

Through targeted efforts that facilitate finding solutions, the private and social sectors could drive the development and testing of prototypes for the required tools. As Sandvig writes, designing algorithmic systems and processes so that they promote broad-based participation and thus serve the common good requires commitment on the part of the public sector and civil society: “Regulating for auditability also implies an important third-party role for government, researchers, concerned users, and/or public interest advocates to hold Internet platforms accountable by routinely auditing them. This would require financial and institutional resources that would support such an intervention: a kind of algorithm observatory acting for the public interest” (Sandvig et al. 2014: 18).

7.1.5 Development of alternative relevance metrics (meso level)

Depending on the findings of such evaluations, efforts would be necessary to improve signals, for example those documenting relevance as defined by the guiding principles. Numerous projects exist dedicated to finding solutions to specific challenges; they could be examined for their relevance and transferability to Germany. As part of the *Trust Project* at the University of Santa Clara, for example, researchers are developing a list of indicators signaling journalistic quality which can be extracted from a website’s HTML source code (Filloux 2017). If it functions consistently and provides valid results, the list would be an important addition to the options for evaluating the relevance of intermediaries such as Facebook.

7.1.6 Institutional anchoring and dynamic development of guiding principles (meso level)

In the area of editorially curated media in Germany, the principles guiding public discourse are ensured and developed by numerous bodies dedicated to regulation, co-regulation and self-regulation. Institutions such as the German Press Council and the country’s broadcasting boards are prime examples of instruments that guarantee the following criteria are met:

- **Appropriateness**: Before a new ADM process is deployed, an understanding must be reached as to its goals, social impact and suitability within a broad-based dialogue.
- **Responsibility**: It must be clear who is responsible at every step along the way for a process’s proper deployment, and the relevant parties must ensure their responsibilities are being met.

It must be ascertained how institutions such as the press council delegate to intermediaries which shape the public discourse. In the US, initial attempts at laying a foundation for a system of professional ethics have been made by the *Association for Computing Machinery* (ACM) (USACM 2017), participants in the workshop *Fairness, Accountability, and Transparency in Machine Learning* (FAT/ML) (2016) and developers of the *Asilomar AI Principles*. Although these efforts treat ADM processes in general, they are a good starting point for transferring findings to ADM processes that shape public discourse. The development of the press code and press council in Germany shows that the practice of issuing judgments on an ongoing basis and examining concrete examples has an effect on the individual level, since journalists often reference decisions made by the press council when deciding how to act. Similar effects will hopefully also be discernible in the medium term as professional ethics for ADM developers become available.

Binding predictions of social consequences could conceivably be used as a concrete instrument and promulgator of professional ethics, as the FAT/ML group proposes for ADM processes: “When the system is launched, the statement should be made public as a form of transparency so that the public has expectations for social impact of the system” (FAT/ML 2016).

7.1.7 Anchoring of guiding principles for individuals (micro level)

Professional ethics developed and institutionalized as described above must be adopted on the individual level by all persons who design algorithmic systems. To achieve this, professional ethics in its various guises could be included in educational and training programs (Zweig 2017).

7.1.8 Promoting individual skills for dealing with algorithmic systems (micro level)

Developers, operators and the public at large need information about the foundations, the potential shortcomings and the impacts of algorithmic processes in general, as well as the assumptions underlying specific processes with which they are confronted and the unintended consequences of those processes. What are needed here are measures that sensitize people to the relevant issues, along with educational programs and advisory services (such as those offered by consumer protection groups). Also required are instruments such as standardized, public and generally comprehensible statements by the developers of a given algorithmic process as to their basic assumptions, any unintended consequences of the process, etc. (Zweig 2017).

7.2 Responding to systematic distortions in human perception

7.2.1 Research and external evaluation (macro level)

Application-oriented research on optimizing user interfaces for intermediaries' digital platforms is currently focused on lowering barriers and thresholds. As a rule, the goal is to make user interactions as simple and smooth as possible in order to maximize interaction frequency (see Section 5.4). Eyal lists a series of exemplary questions that, according to tenets laid out by B. J. Fogg, designers must answer in order to "increase the likelihood that a behavior will occur":

> *Is the user short on time? Is the behavior too expensive? Is the user exhausted after a long day of work? Is the product too difficult to understand? Is the user in a social context where the behavior could be perceived as inappropriate? Is the behavior simply so far removed from the user's normal routine that its strangeness is off-putting? (Eyal 2014)*

Eyal even recommends explicit, cognitive heuristics, such as anchoring and framing effects, that make targeted use of distortions (see also Kahneman 2012a; Tversky and Kahnemann 1974) in order to activate certain behaviors simply and effectively.

In other words, captology, the science of digital behavior modification, is largely based on the optimization of System 1 processing (Kahneman 2012b), i.e. rapid, less strenuous, potentially emotional forms of cognition that are prone to error and are unsuitable for addressing many more complex tasks such as engaging with social contexts.

What would therefore be welcome is a research program that tends in the opposite direction: research of applications that favor a demanding, deep and thorough engagement with content (System 2 processing). What are not needed are answers to questions such as how extremely simple behaviors can be activated as frequently and reliably as possible – clicking on a "like" button, for example – but questions such as how it is possible to ensure that users only share or comment on a post if they truly engage with and, ideally, understand it, i.e. if they have made a true cognitive effort before taking action.

7.2.2 Ensuring a range of forums for user interaction (meso level)

The shaping of digital products and services influences which type of human interaction predominates. Digital offerings that are largely optimized for fast, less strenuous, potentially emotional forms of cognition (see Section 7.2.1) promote certain forms of engagement and discussion. As Lobo and Lauer (2015) ascertain: "The dominant social media platforms are outstandingly suited to mobilization, to group-building and to lightning-quick information diffusion – but not to productive, political discussion" (ibid.: 2436 ff.).

A starting point for intervention here is the creation of alternative platforms for narrowly defined areas of application, carried out by nonprofit organizations. This does not mean creating new social networks, but using clearly limited target groups and operational areas to identify where the need exists for digital products whose algorithmic processes promote other cognitive processes – products, in other words, that are optimized to meet other objectives besides activating simple behavior as often and reliably as possible ad infinitum. An alternative goal could be achieving consensus within a structured process that has a predefined end. Examples of undertakings with such a goal include the Aula Project supported by the Bundeszentrale für politische Bildung (German Federal Agency for Civic Education), in which students, using a platform based on liquid democracy, develop ideas for improving their school and then build a majority, negotiate compromises and vote (Dobusch 2015). Another example is the *Coral Project*, a collaborative effort of the Mozilla Foundation, Knight Foundation, New York Times and Washington Post, which develops software and materials to improve the discussion culture in forums (The Coral Project 2016).

7.2.3 Inoculation through platform design (meso level)

According to several researchers, preventive steps can be taken to counteract disinformation and its impact. Van der Linden et al. (2017) have shown using a large random sample that when warnings are given, especially detailed ones, signaling that disinformation pertaining to man-made climate change is imminent, the disinformation's impact can in fact be neutralized. The authors speak of an "inoculation" against disinformation as it is propagated with considerable vehemence by climate-change deniers, especially in the US. At the same time, it is unclear to what extent this finding can be applied in more general terms to other issues and scenarios. Moreover, the same inoculation method has clearly been used successfully by disseminators of disinformation themselves, with both US President Donald Trump continuing to rail against the "dishonest media" which have reported critically on his administration, and the battle cry of "Lügenpresse" (lying press) being declaimed in Germany by members of the New Right to discredit German media across the board. Both instances can be seen as attempts at "inoculation," i.e. to turn the tables and cast doubt on critical reportage.

One concept that has recently proven popular in the relevant research literature for explaining certain distortions in perception and processing, especially regarding politically infused information, is *politically motivated reasoning* (see for example Taber, Cann and Kucsova 2009). In the words of Tabel et al.: "Citizens' prior attitudes toward the people, groups, or issues implicated in political arguments strongly bias how they process those arguments, through selective exposure or selective judgment processes" (ibid.).

This concept is based on long-standing socio-psychological ideas such as the reduction of cognitive dissonance (Festinger 2001). A research group recently highlighted findings that provide a measure of hope regarding the problems resulting from *politically motivated reasoning:* Kahan et al. (2016) claim to have shown that *scientific curiosity* can serve as a protective factor capable of offsetting the distorting effects of politically motivated reasoning. According to the authors, a personality trait exists – *scientific curiosity,* something that has been much debated among experts for quite some time – which can be defined as the joy of being surprised by results that contradict one's world view. Scientific curiosity is a characteristic that Kahan et al. maintain is exhibited to a greater or lesser degree by all people and can, if especially present, counteract the distortions resulting from politically motivated reasoning. However, research in this field is still in its infancy – and it is unclear and even questionable if scientific curiosity can be awakened through education in people who do not naturally exhibit it.

A further approach for offsetting some of the factors discussed in Section 5.2 is having intermediaries change the design of their offerings in order to encourage more reflective, considered interactions in lieu of the currently preferred System 1 methods (see Section 5.4). Possibilities here include the targeted activation of certain types of so-called *meta cognition,* i.e. thinking about one's own thinking (Alter et al. 2007), thinking about actively delaying certain reactions to content, and thinking about methods that encourage greater emotional distance to the subject at hand (see for example Costa et al. 2014; Geipel, Hadjichristidis and Surian 2015).

NRKbeta, a website of NRK, the Norwegian public broadcasting company, is currently testing a system to achieve exactly this goal. Before the website's users post a comment on certain stories, they must first answer three multiple-choice questions relating to the relevant content (Lichterman 2017). Without availing himself of psychological findings or theories, one of the site's editors explained the approach thus: "If you spend 15 seconds on it, those are maybe 15 seconds that take the edge off the rant mode when people are commenting" (ibid.). In addition, the editor said, the system ensures that people who comment on a story have actually read it.

7.2.4 Diverse individual input (micro level)

A first step toward inoculation against possible participation-related distortions in personal media environments would undoubtedly be an attempt to introduce diversity into the media that individual users are exposed to. Distorted depictions of public discourse and disinformation seen on intermediary sites which is designed to be misleading will of course be less impactful if users also access other sites. Against this background, activities that give school students a basic awareness of the importance of engaging with diverse, quality media could presumably also have an immunizing effect.

7.2.5 Sensitization to the consequences of cognitive distortion (micro level)

Responses such as those proposed in the preceding sections will take time. For that reason and others, measures for controlling and monitoring the impacts of ADM processes should be supplemented by activities designed to sensitize users to the issues resulting from those processes. Possibilities here include specially designed training courses or continuing education seminars. Regardless of how the research develops in this area, educational programs that emphasize facts, objectivity and a thorough examination of source materials will play a crucial role. Indeed, they will gain considerably in importance over time as a means of inoculating the public against strategically deployed disinformation such as the rejection of scientific and scholarly consensus.

A first step towards this sensitization could and should be to inform users of algorithmically curated services that machine-based ranking processes are at work. Diverse empiric evidence exists, for example, that many Facebook users are wholly unaware of the fact that their News Feed is based on specific criteria into which they themselves have no insight. Meredith Morris, for instance, has shown that some users were annoyed by how often they saw pictures of newborn babies in their News Feed and thus came to the conclusion that new mothers were flooding Facebook with images of their offspring (Morris 2014). Yet this assumption was not supported by empiric verification: The frequency of baby pictures on Facebook as perceived by users stemmed from the fact that these images received many likes and comments and were therefore placed higher in the ranking by the algorithm.

In an extensive qualitative study involving 40 participants, Eslami et al. (2015a; 2015b) ascertained that less than half of the participants were aware that Facebook uses a ranking algorithm (see Section 3.2). The researchers developed *FeedVis*, an application that makes it possible for users to see both their personalized, algorithmically curated News Feed and an unfiltered version in which individual posts appear in their entirety and in reverse chronological order. FeedVis also allows users to set their own priorities for ranking incoming posts, for example by showing contacts more frequently that have previously been underrepresented, or by reducing the frequency of other contacts.

Eslami et al. (2015b) report that in one of their studies 83% of the participants subsequently changed their Facebook behavior after seeing their unfiltered News Feed using FeedVis. Several began actively using the network’s settings options, others changed the way they interacted with contacts on the platform in order to signal to the algorithm which people or pages they wanted to see content from more often. In follow-up interviews, several participants said they had become more selective about when they clicked the “like” button “because it will have consequences on what [they] see/don’t see in the future.” One participant even stopped using Facebook completely since, after learning about the algorithmic curation, she said she felt “like I was being lied to” (ibid.).

If they were demonstrated as part of classroom instruction or continuing education programs, such comparisons would be an easily implementable method of explaining the mechanisms used by the relevant platforms, a method that could both increase awareness and change behavior and, thus, presumably have a long-term impact.

A non-personalized but very compelling method of visualizing the distortions arising from personalization can be seen in projects such as Jon Keegan’s *Blue Feed, Red Feed* (Keegan 2016). Based on the *Science* study cited in Section 5.6 carried out by researchers working for Facebook, Keegan developed two Facebook accounts for the *Wall Street Journal* which contained only stories shared by either “very conservatively” or “very liberally” aligned

Facebook accounts as determined by data collected by Bakshy et al. (2015). The findings, which continue to be updated, provide insight into the media experience that two different, fictional Facebook users have when they engage with a variety of issues. Similar depictions could undoubtedly be created for users in other countries or for different viewpoints, political or otherwise.

Another measure relating to increasing awareness of algorithmic curation is Facebook's own efforts in recent months to remove posts that are clearly fake news once they have been identified as such by external organizations such as the independent research organization *Correctiv* (Schraven 2017). The clear identification of stories shown to be manipulative could potentially have a sensitizing effect by making users aware that their algorithmically ranked News Feeds might also contain fake news. However, any impact that a confrontation with factual counterarguments might have, especially on already radicalized individuals, remains unclear (see Section 5.6).

8 Conclusion

Essentially this paper answers three key questions relating to public discourse and public opinion in the digital age:

1. *Media transformation: How is public discourse changing because of the new digital platforms through which many people now receive socially relevant information?*
   When all age groups are considered, intermediaries driven by algorithmic processes, such as Google and Facebook, have had a large but not defining influence on how public opinion is formed, compared to editorially driven media such as television. These intermediaries judge the relevance of content based on the public's immediate reaction to a much greater degree than do traditional media.
2. *Social consequences: In terms of quality and diversity, is the information that reaches people via these new channels suitable for a democratic decision-making process and does it promote participation?*
   Use of these intermediaries for forming public opinion is leading to a structural change in the public sphere. Key factors here are the algorithmic processes used as a basic formational tool and the leading role that user reactions play as input for these processes. Since they are the result of numerous psychological factors, these digitally assessed and, above all, impulsive public reactions are poorly suited to determining relevance as defined by traditional social values. Until now, these values, such as truth, diversity and social integration, have served in Germany as the basis for public opinion as formed by editorial media.
3. *Solutions: How can the new digital platforms be designed to ensure they promote participation?*
   Algorithms that sort content and personalize how it is assembled form the core of the complex, interdependent process underlying public discourse and the formation of public opinion. That is why solutions must be found here first. The most important areas that will need action in the foreseeable future are facilitating external research and evaluation, strengthening the diversity of algorithmic processes, anchoring guiding principles (e.g. by focusing on professional ethics) and increasing awareness among the public.

1. Media transformation: Intermediaries are a relevant but not determining factor for the formation of public opinion.

Numerous studies have shown that so-called *intermediaries* such as Google and Facebook play a role in the formation of public opinion in numerous countries including Germany. For example, 57% of German Internet users receive politically and socially relevant information via search machines or social networks. And although the share of users who say that social networks are their most important news source is still relatively small at 6% of all Internet users, this figure is significantly higher among younger users. It can thus be assumed that these types of platforms will generally increase in importance. The formation of public opinion is "no longer conceivable without intermediaries," as researchers at the Hamburg-based Hans Bredow Institute put it in 2016.

The principles that these intermediaries use for shaping content are leading to a structural shift in the public sphere. Key aspects are:

- *Decoupling of publication and reach:* Anyone can make information public, but not everyone finds an audience. Attention is only gained through the interaction of people and algorithmic decision-making (ADM) processes.
- *Detachment from publications:* Each story or post has its own reach.
- *Personalization:* Users receive more information about their specific areas of interest.
- *Increased influence of public on reach:* User reactions influence ADM processes in general and the reach of each article or post.
- *Centralization of curators:* There is much less diversity when looking at intermediaries than in the sphere of traditionally curated media.

- *Interplay of human and machine-based curation:* Traditionally curated media disseminate content via intermediaries and use the resulting reactions on intermediary sites as a measure of public interest.

The comparison of processes in the figure below shows the key role now being played by user reactions and algorithmic processes in public communication and the formation of public opinion. User reactions and algorithmic processes determine the attention received via intermediaries. Our hypothesis is that it is impossible to define a clear direction of causation between these reactions and processes.

Figure 2: Organization of an algorithmically structured public sphere

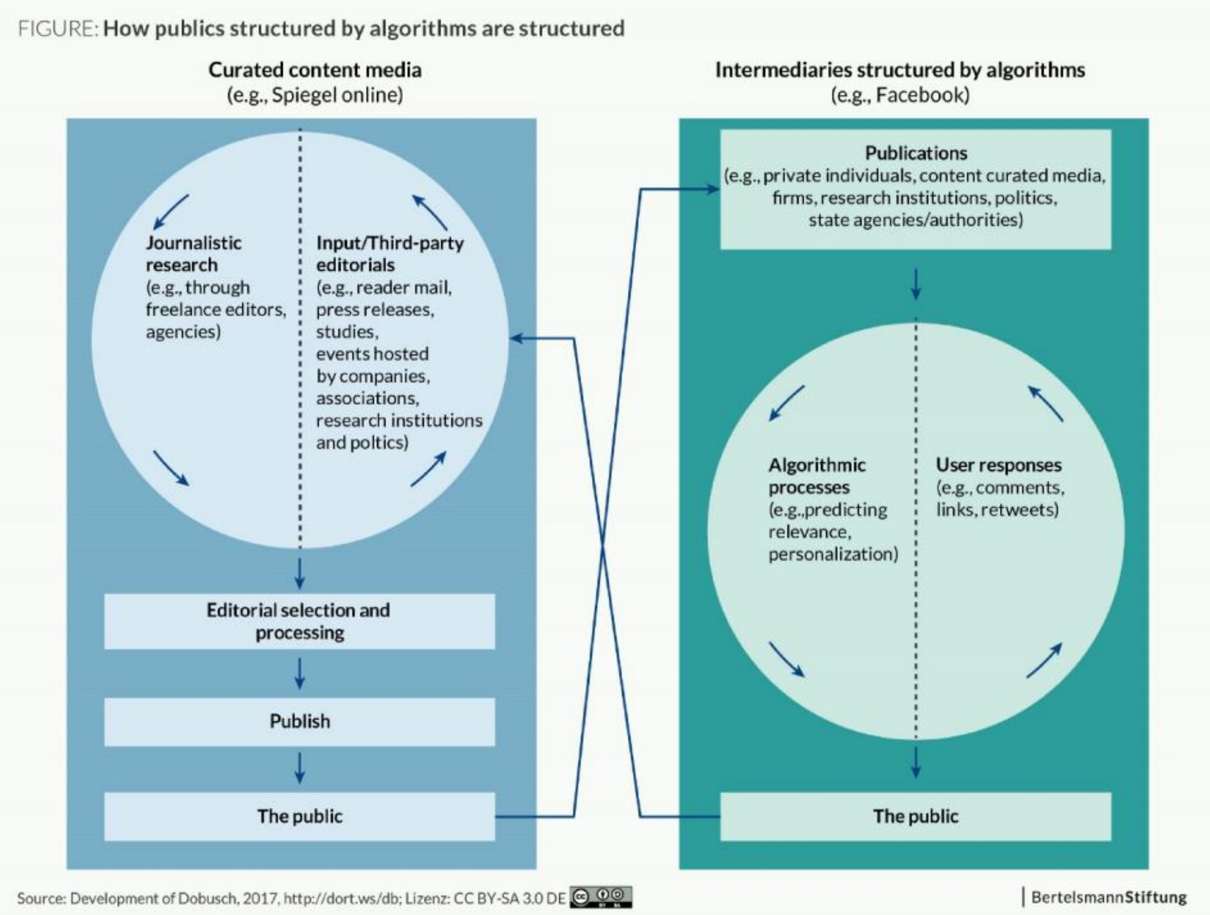


FIGURE: How publics structured by algorithms are structured
Curated content media
(e.g., Spiegel online)
Journalistic research
(e.g., through freelance editors, agencies)
Input/Third-party editorials
(e.g., reader mail, press releases, studies, events hosted by companies, associations, research institutions and poltics)
Editorial selection and processing
Publish
The public
Intermediaries structured by algorithms
(e.g., Facebook)
Publications
(e.g., private individuals, content curated media, firms, research institutions, politics, state agencies/authorities)
Algorithmic processes
(e.g.,predicting relevance, personalization)
User responses
(e.g., comments, links, retweets)
The public
| BertelsmannStiftung

Source: Development of Dobusch, 2017, http://dort.ws/db; Lizenz: CC BY-SA 3.0 DE

Google, Facebook and other intermediaries are already playing an important role in the public discourse, even though as a rule these platforms were not originally conceived to supply consumers with media content from journalistic organizations. They use technical systems instead to decide whether certain content taken from a huge pool of information could be interesting or relevant to a specific user. These systems were initially intended – in the case of search engines – to identify websites, for example, that contain certain information or – in the case of social networks – to display in a prominent position particularly interesting messages or photos from a user’s own circle of friends. In many cases they therefore sort content according to completely different criteria than editors at a daily newspaper or a magazine would. “Relevant” means something different to Google than it does to Facebook, and both understand it differently than the editors at SPIEGEL ONLINE or Sueddeutsche.de.

The intermediaries use numerous variables to calculate the relevance of individual items. These variables range from basic behavioral metrics such as scrolling speed or how long a page is viewed to the level of interaction among multiple users in a social network. When someone with whom a user has repeatedly communicated on Facebook posts content, the probability is higher that the user will be shown this content than they would if someone posts with whom the user has, theoretically, a digital connection, but with whom the user has never truly had contact. The signals that other users send – often unknowingly – are also included in assessments of relevance, whether they be the insertion of links, clicks on links, clicks on the “like” button, forwards, i.e. so-called shares or the number of comments that certain content receives.

2. Social consequences: As used by the most relevant intermediaries currently forming public opinion, algorithmic processes evaluate users’ reactions to content. They promote and reinforce a manner of human cognition that is susceptible to distortion, and they themselves are susceptible to technical manipulation.

The metrics signaling relevance – about which platform operators are hesitant to provide details because of competition-related factors or other reasons – are potentially problematic. This is true, first, because the operators themselves are constantly changing the metrics. Systems such as those used by Google and Facebook are being altered on an ongoing basis; the operators experiment with and tweak almost every aspect of the user interface and other platform features in order to achieve specific goals such as increased interactivity. Each of these changes can potentially impact the signals that the platforms themselves capture to measure relevance.

A good example is the People You May Know feature used by Facebook, which provides users with suggestions for additional contacts based on assessments of possible acquaintances within the network. When this function was introduced, the number of new links made every day within the Facebook community immediately doubled. The network of relationships displayed on such platforms is thus dependent on the products and services that the operators offer. At the same time, the networks of acquaintances thus captured also become variables in the metrics determining what is relevant. Whoever makes additional “friends” is thus possibly shown different content.

A further problem stemming from the metrics collected by the platform operators is the type of interaction for which such platforms are optimized. A key design principle is that interactions should be as simple and convenient as possible in order to maximize the probability of their taking place. Clicking on a “like” button or a link demands almost no cognitive effort, and many users are evidently happy to indulge this lack of effort. Empiric studies suggest, for example, that many articles in social networks forwarded with a click to the user’s circle of friends could not possibly have been read. Users thus disseminate media content after having seen only the headline and introduction. To some extent they deceive the algorithm and, with it, their “friends and followers” into believing that they have engaged with the text.

The ease of interaction also promotes cognitive distortions that have been known to social psychologists for decades. A prime example is the availability heuristic: If an event or memory can easily be recalled, it is assumed to be particularly probable or common. The consequence is that users frequently encounter unread media content that has been forwarded due to a headline, and the content is thus later remembered as being “true” or “likely.” This is also the case when the text itself makes it clear that the headline is a grotesque exaggeration or simply misleading.

Other psychological factors play an important role here, for example the fact that people use social media in particular not only for informational purposes, but also as a tool for identity management, with some media content being forwarded only to demonstrate the user’s affiliation with a certain political camp, for example. Moreover, the design of many digital platforms explicitly and intentionally encourages a fleeting, emotional engagement with content. Studies of networking platforms indeed show that *content which rouses emotion* is commented on and shared particularly often – above all when negative emotions are involved.

Such an emotional treatment of news content can lead to increased *societal polarization,* a hypothesis for which initial empirical evidence already exists, especially in the United States. At the same time, however, such polarizing effects seem to be dependent on a number of other factors such as a country’s electoral system. Societies with first-past-the-post systems such as the US are potentially more liable to extreme political polarization than those with proportional systems, in which ruling coalitions change and institutionalized multiparty structures tend to balance out competing interests. Existing societal polarization presumably influences and is influenced by the algorithmic ranking of media content. For example, one study shows that Facebook users who believe in conspiracy theories tend over time to turn to the community of conspiracy theorists holding the same views. This process is possibly exacerbated by algorithms that increasingly present them with the relevant content. These systems could in fact result in the creation of so-called *echo chambers,* at least among people with extremist views.

Technical manipulation can also influence the metrics that intermediaries use to ascertain relevance. So-called bots – partially self-acting software applications that can be disguised as real users, for example in social networks – can massively distort the volume of digital communication occurring around certain topics. According to one study, during the recent presidential election in the US, 400,000 such bots were in use on Twitter, accounting for about one-fifth of the entire discussion of the candidates' TV debates. It is not clear to what extent these types of automated systems actually influence how people vote. What is clear is that the responses they produce – clicks, likes, shares – are included in the relevance assessments generated by ADM systems. Bots can thus make an article seem so interesting that an algorithm will then present it to human users.

In sum, it can be said that the relevance assessments that algorithmic systems create for media content do not necessarily reflect criteria that are desirable from a societal perspective. Basic values such as truthfulness or social integration do not play a role. The main goal is to increase the probability of an interaction and the time users spend visiting the relevant platform. Interested parties whose goal is a strategic dissemination of disinformation can use these mechanisms to further their cause: A creative, targeted lie can, on balance, prove more emotionally "inspiring" and more successful within such systems – and thus have a greater reach – than the boring truth.

3. Solutions: Algorithmic sorting of content is at the heart of the complex interdependencies affecting public discourse in the digital sphere. This is where solutions must be applied.

The last section of this paper contains a series of possible solutions for these challenges. A first goal, one that is comparatively easy to achieve, is making users more aware of the processes and mechanisms described here. Studies show that users of social networking platforms do not even know that such ranking algorithms exist, let alone how they work. Educational responses, including in the area of continuing education, would thus be appropriate, along with efforts to increase awareness of disinformation and to decrease susceptibility to it, for example through fact-finding and verification education.

Platform operators themselves clearly have more effective possibilities for intervention. For example, they could do more to ensure that values such as appropriateness, responsibility and competency are adhered to when the relevant systems are being designed and developed. A medium-term goal could be defining industry-wide professional ethics for developers of ADM systems.

Moreover, researchers who do not work for platform operators should be in a position to examine and evaluate the impact being made by the operators' decisions. Until now it has been difficult if not impossible for external researchers or government authorities to gain access to the required data, of which operators have vast amounts at their disposal. Neither the design decisions made by platform operators nor the impacts of those decisions on individual users are transparent to any significant degree. Systematic distortions, for example in favor of one political viewpoint or another, are difficult to identify using currently available data. More transparency – through a combination of industry self-regulation and, where necessary, legislative measures – would make it possible to gain an unbiased understanding of the actual social consequences of algorithmic ranking and to identify potential dangers early on. Making it easier to conduct research would also stimulate an objective, solution-oriented debate of the issue and could help identify new solutions. Measures like these could also make it easier to design algorithmic systems that increase participation. This would foster a more differentiated view of algorithmic processes and could increase trust in those systems that are designed to benefit all of society.

9 References

Alter, Adam L., Daniel M. Oppenheimer, Nicholas Epley and Rebecca N. Eyre (2007). "Overcoming intuition: Metacognitive difficulty activates analytic reasoning." *Journal of Experimental Psychology. General* (136) 4. 569–576. https://doi.org/10.1037/0096-3445.136.4.569 (accessed June 2, 2017).

An, Jisun, Daniele Quercia and Jon Crowcroft (2013). "Fragmented social media: A look into selective exposure to political news." *Proceedings of the 22nd International Conference on World Wide Web.* New York NY: ACM. 51–52. http://dl.acm.org/citation.cfm?id=2487807 (accessed June 2, 2017).

Ananny, Mike and Kate Crawford (2016). "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability." *New Media & Society* 15 (7). 1005–1021. https://doi.org/10.1177/1461444816676645 (accessed June 2, 2017).

Association for Computing Machinery US Public Policy Council (USACM) (2017). Statement on Algorithmic Transparency and Accountability. Jan. 12, 2017. https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf (accessed June 2, 2017).

Backstrom, Lars (2013). "News Feed FYI: A window into News Feed." March 6, 2013. https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed (accessed March 1, 2017).

Bakshy, Eytan (2014). "Big experiments: Big data's friend for making decisions." April 3, 2014. https://www.facebook.com/notes/facebook-data-science/big-experiments-big-datas-friend-for-making-decisions/10152160441298859/ (accessed June 2, 2017).

Bakshy, Eytan, Solomon Messing and Lada A. Adamic (2015). "Exposure to ideologically diverse news and opinion on Facebook." *Science* (348) 6239. 1130–1132. https://doi.org/10.1126/science.aaa1160 (accessed June 2, 2017).

Beirat Integration (Advisory Council on Integration) (2013). ‚Soziale Teilhabe' Handlungsempfehlungen des Beirats der Integrationsbeauftragten. Die Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration, ed. Feb. 22, 2013. http://www.bagiv.de/pdf/soziale-teilhabe-empfehlungen-beirat.pdf (accessed June 2, 2017).

Benkler, Yochai, Robert Faris, Hal Roberts and Ethan Zuckerman (2017). "Breitbart-led right-wing media ecosystem altered broader media agenda." *Columbia Journalism Review*. http://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php (accessed June 2, 2017)

Bentele, Günter (2016). "Wahrheit." *Handbuch Medien-und Informationsethik*. Jessica Heesen, ed. Stuttgart. 59–66. http://link.springer.com/chapter/10.1007/978-3-476-05394-7_3 (accessed June 2, 2017).

Bertelsmann Stiftung (2011). *Soziale Gerechtigkeit in der OECD – Wo steht Deutschland? Sustainable Governance Indicators 2011*. Gütersloh. http://news.sgi-network.org/uploads/tx_amsgistudies/SGI11_Social_Justice_DE.pdf (accessed June 2, 2017).

Bessi, Alessandro and Emilio Ferrara (2016). “Social bots distort the 2016 US Presidential election online discussion.” First Monday, 21(11). http://firstmonday.org/ojs/index.php/fm/article/view/7090

Boland, Brian (2014). “Organic reach on Facebook: Your questions answered.” June 5, 2014. https://www.facebook.com/business/news/Organic-Reach-on-Facebook (accessed April 6, 2017).

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle and James H. Fowler (2012). “A 61-million-person experiment in social influence and political mobilization.” *Nature* (489) 7415. 295–298. https://doi.org/10.1038/nature11421 (accessed June 2, 2017).

Borgesius, Frederik J. Zuiderveen, Damian Trilling, Judith Moeller, Balázs Bodó, Cales H. de Vreese and Natali Helberger (2016). Should we worry about filter bubbles? (SSRN Scholarly Paper No. ID 2758126). Rochester NY: Social Science Research Network. https://papers.ssrn.com/abstract=2758126 (accessed June 2, 2017).

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). Greater Internet use is not associated with faster growth in political polarization among US demographic groups. Proceedings of the National Academy of Sciences, 114(40), 10612–10617.

Bundesgesetzblatt (Federal Legal Gazette) (1966). “Internationaler Pakt über wirtschaftliche, soziale und kulturelle Rechte.” Bundesgesetzblatt (BGBl) 1976 II, 428. 19.12.1966. http://www.institut-fuer-menschenrechte.de/fileadmin/user_upload/PDF-Dateien/Pakte_Konventionen/ICESCR/icescr_de.pdf (accessed May 8, 2017).

Bundesprüfstelle für jugendgefährdende Medien (BPjM, Federal Review Board for Media Harmful to Minors) (n.d.). BPjM-Modul. http://www.bundespruefstelle.de/bpjm/Aufgaben/Listenfuehrung/bpjm-modul.html (accessed June 2, 2017).

Bundesverfassungsgericht (German Constitutional Court) (1981). BVerfGE 57, 295 - 3. Rundfunkentscheidung. June 16, 1981. http://sorminiserv.unibe.ch:8080/tools/ainfo.exe?Command=ShowPrintVersion&Name=bv057295 (accessed June 2, 2017).

Burri, Mira (2013). “Contemplating a ‘public service navigator’: In search of new (and better) functioning public service media.” May 5, 2015. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2364951 (accessed June 2, 2017).

Cheng, Justin, Michael Bernstein, Cristian Danescu-Niculescu-Mizil and Jure Leskovec (2017). “Anyone can become a troll: Causes of trolling behavior in online discussions.” https://pdfs.semanticscholar.org/3427/1ca4d5c91e258c2e5bf67a5f4d4698dd1885.pdf (accessed June 2, 2017).

Churchill, Winston (1943). “House of Commons Rebuilding.” Hansard, Oct. 28, 1943. *HC* (393). 403–473. http://hansard.millbanksystems.com/commons/1943/oct/28/house-of-commons-rebuilding (accessed June 2, 2017).

Clark, Jack (2015). “Google turning its lucrative web search over to AI machines.” *Bloomberg.com*. Oct. 26, 2015. https://www.bloomberg.com/news/articles/2015-10-26/google-turning-its-lucrative-web-search-over-to-ai-machines (accessed June 2, 2017).

Constine, Josh (2016). “How Facebook News Feed works.” Sept. 6, 2016. http://social.techcrunch.com/2016/09/06/ultimate-guide-to-the-news-feed/ (accessed June 2, 2017).

Costa, Albert, Alice Foucart, Sayun Hayakawa, Malina Aparici, Jose Apesteguia, Joy Heafner and Boaz Keysar (2014). “Your morals depend on language.” *PLOS ONE* (9) 4, e94842. https://doi.org/10.1371/journal.pone.0094842 (accessed June 2, 2017).

Council of Europe (2012a). Recommendation CM/Rec(2012)3 of the Committee of Ministers to member states on the protection of human rights with regard to search engines. April 4, 2012. https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805caa87 (accessed June 2, 2017).

Council of Europe (2012b). Recommendation CM/Rec(2012)4 of the Committee of Ministers to member states on the protection of human rights with regard to social networking services. https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805caa9b (accessed June 2, 2017).

Cutts, Matt (2011). “Finding more high-quality sites in search (2011).” Feb. 24, 2011. https://googleblog.blogspot.de/2011/02/finding-more-high-quality-sites-in.html (accessed June 2, 2017).

Dearringer, Jeremy (2011). “Mission ImposSERPble: Establishing click-through rates.” July 25, 2011. https://moz.com/blog/mission-imposserpble-establishing-clickthrough-rates (accessed June 2, 2017).

Del Vicario, Michaela, Allesandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Staley and Walter Quattrociocchi (2016). “The spreading of misinformation online.” *Proceedings of the National Academy of Sciences* (113) 3. 554–559. https://doi.org/10.1073/pnas.1517441113 (accessed June 2, 2017).

Deutscher Presserat (German Press Council) (2017). German Press Code. Guidelines for journalistic work as recommended by the German Press Council. Complaints Procedure. Berlin. http://www.presserat.de/fileadmin/user_upload/Downloads_Dateien/Pressekodex2017english.pdf (accessed Oct. 20, 2017).

Diakopoulos, Nicholas (2016a). “Accountability in algorithmic decision-making.” *Communications of the ACM* (59) 2. 56–62. https://doi.org/10.1145/2844110 (accessed June 2, 2017).

Diakopoulos, Nicholas (2016b). “Artificial moderation: A reading list.” March 29, 2017. https://blog.coralproject.net/artificial-moderation-a-reading-list/ (accessed Jan. 5, 2017).

die medienanstalten (2016). Staatsvertrag für Rundfunk und Telemedien (Rundfunkstaatsvertrag - RStV) vom 31. August 1991. http://www.die-medienanstalten.de/fileadmin/Download/Rechtsgrundlagen/Gesetze_aktuell/19_RfAendStV_medienanstalten_Layout_final.pdf (accessed June 2, 2017).

Dobusch, Leonhard (2015). “Projekt ‚aula‘ sucht Schulen, die mit Liquid Democracy experimentieren wollen.” Nov. 23, 2017. https://netzpolitik.org/2015/projekt-aula-sucht-schulen-die-mit-liquid-democracy-experimentieren-wollen/ (accessed April 27, 2017).

Dobusch, Leonhard (2017). “Die Organisation der Digitalität: Zwischen grenzenloser Offenheit und offener Exklusion.” Feb. 1, 2017. https://netzpolitik.org/2017/die-organisation-der-digitalitaet-zwischen-grenzenloser-offenheit-und-offener-exklusion/ (accessed Feb. 9, 2017).

Donges, Patrick (2016). “Funktionsaufträge des Rundfunks.” *Handbuch Medien- und Informationsethik.* Jessica Heesen, ed. Stuttgart. 89–104. http://link.springer.com/chapter/10.1007/978-3-476-05394-7_4 (accessed June 2, 2017).

Dörr, Dieter and Richard Deicke (2015). *Positive Vielfaltsicherung. Bedeutung und zukünftige Entwicklung der Fensterprogramme für die Meinungsvielfalt in den privaten Fernsehprogrammen*. Mainz. http://www.mainzer-medieninstitut.de/dokumente/Studie%20-%20Positive%20Vielfaltsicherung.pdf (accessed June 2, 2017).

Ecke, Oliver (2016). *Wie häufig und wofür werden Intermediäre genutzt?* Berlin. http://www.die-medienanstalten.de/fileadmin/Download/Veranstaltungen/Pr%C3%A4sentation_Intermedi%C3%A4re/TNS_Intermedi%C3%A4re_und_Meinungsbildung_Pr%C3%A4si_Web_Mappe_final.pdf (accessed June 2, 2017).

El-Arini, Khalid and Joyce Tang (2014). “Click-baiting.” Aug. 25, 2014. https://newsroom.fb.com/news/2014/08/news-feed-fyi-click-baiting/ (accessed March 2, 2017).

Epstein, Robert and Ronald E. Robertson (2015). “The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections.” *Proceedings of the National Academy of Sciences of the United States of America* (112) 33. 4512–4521. https://doi.org/10.1073/pnas.1419828112 (accessed June 2, 2017).

Eslami, Montahhare, Amirhossein Aleyasen, Karrie Karahalios, Kevin Hamilton and Christian Sandvig (2015a). “FeedVis: A path for exploring news feed curation algorithms.” *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*. New York NY: ACM. 65–68. https://doi.org/10.1145/2685553.2702690 (accessed June 2, 2017).

Eslami, Motahhare, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton and Christian Sandvig (2015b). “‘I always assumed that I wasn’t really that close to [her]’: Reasoning about invisible algorithms in news feeds.” Presented at CHI 2015, Crossings, Seoul, Korea. 153–162. https://doi.org/10.1145/2702123.2702556 (accessed June 2, 2017).

Eslami, Motahhare, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton and Alex Kirl (2016). “First I ‘like’ it, then I hide it: Folk theories of social feeds.” *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2371–2382). New York, NY, USA: ACM. https://doi.org/10.1145/2858036.2858494

Etzioni, Amitai and Oren Etzioni (2017). “Incorporating ethics into Artificial Intelligence.” *The Journal of Ethics* March 2017. 1–16.

Eyal, Nir (2014). *Hooked: How to build habit-forming products*. New York NY: Portfolio/Penguin.

Facebook (2006). “Facebook gets a facelift.” Sept. 5, 2006. https://www.facebook.com/notes/facebook/facebook-gets-a-facelift/2207967130/ (accessed March 2, 2017).

Facebook (2016a). “How News Feed works.” June 17, 2017. http://nonprofits.fb.com/2016/06/17/how-news-feed-works/ (accessed Feb. 21, 2017).

Facebook (2016b). “News Feed values.” June 28, 2017. https://newsroom.fb.com/news/2016/06/building-a-better-news-feed-for-you7 (accessed Oct. 20, 2017).

Facebook (n.d.). “Community standards.” https://www.facebook.com/communitystandards (accessed April 8, 2017).

Fairness, Accountability, and Transparency in Machine Learning (FAT/ML) (2016). “Principles for Accountable Algorithms and a Social Impact Statement for Algorithms.” http://www.fatml.org/resources/principles-for-accountable-algorithms (accessed June 2, 2017).

Festinger, Leon (2001). *A theory of cognitive dissonance.* (Reissued by Stanford Univ. Press in 1962, renewed 1985 by author.) Stanford: Stanford Univ. Press.

Fetzer, Thomas (2015). *Effektive Vielfaltssicherung im Internet*. Düsseldorf. https://mbem.nrw/sites/default/files/asset/document/effektive_vielfaltssicherung_fetzer_finalcc_by_nd_30_de.pdf (accessed June 2, 2017).

Filloux, Frederic (2017). "What web page structure reveals on news quality – Monday Note." April 10, 2017. https://mondaynote.com/what-web-page-structure-reveals-on-news-quality-67b845e230db (accessed April 13, 2017).

Flaxman, Seth, Sharad Goel and Justin M. Rao (2016). "Filter bubbles, echo chambers, and online news consumption." *Public Opinion Quarterly* (80) S1. 298–320.

Fogg, B. J. (2003). *Persuasive technology: Using computers to change what we think and do*. Amsterdam, Boston MA: Morgan Kaufmann Publishers.

Future of Life Institute (2017). "Asilomar AI Principles." https://futureoflife.org/ai-principles/ (accessed April 15, 2017).

Gabielkov, Maksym, Arthi Ramachandran, Augustin Chaintreau and Arnaud Legout (2016). "Social clicks: What and who gets read on Twitter?" *ACM SIGMETRICS/IFIP Performance 2016*. https://hal.inria.fr/hal-01281190/file/sigm095-gabielkov.pdf (accessed June 2, 2017).

Galtung, Johan, and Mari Holmboe Ruge (1965). "The structure of foreign news." *Journal of Peace Research* (2) 1. 64–91.

Garb, Rachel (2008). "More transparency in customized search results." June 30, 2008. https://googleblog.blogspot.de/2008/07/more-transparency-in-customized-search.html (accessed June 2, 2017).

Geipel, Janet, Constantinos Hadjichristidis and Luca Surian (2015). "How foreign language shapes moral judgment." *Journal of Experimental Social Psychology* 59. 8–17. https://doi.org/10.1016/j.jesp.2015.02.001 (accessed June 2, 2017).

Gigerenzer, Gerd and Wolfgang Gaissmaier (2011). "Heuristic decision-making." *Annual Review of Psychology* (62) 1. 451–482. https://doi.org/10.1146/annurev-psych-120709-145346 (accessed June 2, 2017).

Gillespie, Tarleton (2014). "The Relevance of Algorithms." In Tareleton Gillespie, Pablo Boczkowski and Kirsten Foot, eds., *Essays on Communication, Materiality, and Society.* (pp. 167–194). Cambridge.

Goodman, Bryce and Seth Flaxman (2016). "EU regulations on algorithmic decision-making and a 'right to explanation.'" *WHI* 2016. 26–30. http://arxiv.org/abs/1606.08813 (accessed June 2, 2017).

Google (2016). “Blitzschnelle Suche.” https://static.googleusercontent.com/media/www.google.com/de//insidesearch/howsearchworks/assets/searchInfographic.pdf (accessed June 2, 2017).

Google (2017). General Guidelines. March 14, 2017. https://static.googleusercontent.com/media/www.google.com/de//insidesearch/howsearchworks/assets/searchqualityevaluatorguidelines.pdf (accessed June 2, 2017).

Gottfried, Jeffrey and Elisa Shearer (2016). “News use across social media platforms 2016.” Pew Research Center. May 26, 2016. http://assets.pewresearch.org/wp-content/uploads/sites/13/2016/05/PJ_2016.05.26_social-media-and-news_FINAL-1.pdf

Habermas, Jürgen (1981). *Theorie des kommunikativen Handelns*. Frankfurt am Main.

Habermas, Jürgen (2008). *Ach, Europa: Kleine Politische Schriften XI* (Originalausgabe). Frankfurt am Main.

Hasebrink, Uwe, Jan-Hinrik Schmidt and Lisa Merten (2016). *Wie fließen Intermediäre in die Meinungsbildung ein? Die qualitative Perspektive der Vertiefungsstudie.* Berlin. http://www.die-medienanstalten.de/fileadmin/Download/Veranstaltungen/Pr%C3%A4sentation_Intermedi%C3%A4re/HBI_Intermedi%C3%A4re_und_Meinungsbildung_Pr%C3%A4si_Web_Mappe_final.pdf (accessed June 2, 2017).

Heesen, Jessica (2016). “Freiheit.” *Handbuch Medien-und Informationsethik* . Jessica Heesen, ed. Stuttgart. 51–58. http://link.springer.com/chapter/10.1007/978-3-476-05394-7_3 (accessed June 2, 2017).

Hegelich, Simon and Dietmar Janetzko (2016). “Are social bots on Twitter political actors? Empirical evidence from a Ukrainian social botnet.” http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/viewPDFInterstitial/13015/12793 (accessed June 2, 2017).

Herbold, Astrid (2016). “Hatespeech: Argumente sind kein Allheilmittel.” *Die Zeit.* Feb. 4, 2014. http://www.zeit.de/digital/internet/2016-02/hatespeech-counterspeech-facebook-forschung/komplettansicht (accessed June 2, 2017).

Hern, Alex (2014). “Why Google has 200m reasons to put engineers over designers.” *The Guardian.* Feb. 5, 2014. https://www.theguardian.com/technology/2014/feb/05/why-google-engineers-designers (accessed June 2, 2017).

Horling, Bryan and Robby Bryant (2009). “Personalized search for everyone.” Dec. 4, 2009. https://googleblog.blogspot.com/2009/12/personalized-search-for-everyone.html (accessed April 6, 2017).

Howard, Philip N. and Bence Kollanyi (2016). “Bots, #StrongerIn, and #Brexit: Computational propaganda during the UK-EU referendum.” *arXiv:1606.06356 [physics]*. http://arxiv.org/abs/1606.06356 (accessed June 2, 2017).

Iredale, Sophie and Aleksej Heinze (2016). “Ethics and professional intimacy within the search engine optimisation (SEO) industry.” *Technology and Intimacy: Choice or Coercion,* vol. 474. David Kreps, Gordon Fletcher and Marie Griffiths, eds. Cham. 106–115. https://doi.org/10.1007/978-3-319-44805-3_9 (accessed June 2, 2017).

Iyengar, Shanto and Kyu S. Hahn (2009). “Red media, blue media: Evidence of ideological selectivity in media use.” *Journal of Communication* (59) 1. 19–39.

Iyengar, Shanto and Sean J. Westwood (2015). “Fear and Loathing across party lines: New evidence on group polarization.” *American Journal of Political Science* (59) 3. 690–707. https://doi.org/10.1111/ajps.12152 (accessed June 2, 2017).

Jenyns, David (2013). “How to recover from a Google Panda/Penguin spanking.” Jan. 6, 2013 http://www.melbourneseoservices.com/how-to-recover-from-google-panda/ (accessed April 11, 2017).

Justitia US Supreme Court (1927). Whitney v. California 274 U.S. 357. https://supreme.justia.com/cases/federal/us/274/357/case.html (accessed June 2, 2017).

Kacholia, Varun (2013). “Showing more high quality content.” Aug. 23, 2013. https://newsroom.fb.com/news/2013/08/news-feed-fyi-showing-more-high-quality-content/ (accessed March 2, 2017).

Kahan, Dan M., Asheley Landrum, Katie Carpenter, Laura Helft and Kathleen Hall Jamieson (2016). *Science Curiosity and Political Information Processing.* (SSRN Scholarly Paper No. ID 2816803). Rochester NY: Social Science Research Network. https://papers.ssrn.com/abstract=2816803 (accessed June 2, 2017).

Kahneman, Daniel (2012a). *Schnelles Denken, langsames Denken*. (T. Schmidt, trans.). Munich.

Kahneman, Daniel (2012b). *Thinking, fast and slow*. London: Penguin Books.

Keegan, Jon (2016). “Blue feed, red feed: See liberal Facebook and conservative Facebook, side by side.” *The Wall Street Journal online.* May 18, 2016. http://graphics.wsj.com/blue-feed-red-feed/ (accessed April 15, 2017).

Klayman, Joshua (1995). “Varieties of Confirmation Bias.” *Psychology of Learning and Motivation,* vol. 32. 385–418). https://doi.org/10.1016/S0079-7421(08)60315-1 (accessed June 2, 2017).

Kollanyi, Bence, Philip N. Howard and Samuel C. Woolley (2016). “Bots and automation over Twitter during the U.S. election.” http://politicalbots.org/wp-content/uploads/2016/10/Data-Memo-First-Presidential-Debate.pdf (accessed June 2, 2017).

Kosinski, Michal, Sandra C. Matz, Samuel D. Gosling, Vesselin Popov and David Stillwell (2015). “Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines.” *American Psychologist* (70) 6. 543.

Kramer, Adam D., Jamie E. Guillory and Jeffrey T. Hancock (2014). “Experimental evidence of massive-scale emotional contagion through social networks.” *Proceedings of the National Academy of Sciences* (111) 24. 8788–8790. https://doi.org/10.1073/pnas.1320040111 (accessed June 2, 2017).

Lada, Akos, James Li and Shilin Ding (2017, January 31). “New signals to show you more authentic and timely stories.” https://newsroom.fb.com/news/2017/01/news-feed-fyi-new-signals-to-show-you-more-authentic-and-timely-stories/ (accessed June 15, 2017).

Lazer, David (2015). “The rise of the social algorithm.” *Science* (348) 6239. 1090–1091. https://doi.org/10.1126/science.aab1422 (accessed June 2, 2017).

Leslie, Ian (2016). “The scientists who make apps addictive.” *The Economist / 1843* Oct./Nov. 2016. https://www.1843magazine.com/features/the-scientists-who-make-apps-addictive (accessed June 2, 2017).

Levinovitz, Alan (2017). “William Shatner’s tweets are a classic case of misinformation spread.” April 6, 2017. http://www.slate.com/articles/health_and_science/science/2017/04/what_we_can_learn_from_william_shatner_s_twitter_meltdown.html (accessed April 8, 2017).

Levy, Steven (2016). “How Google is remaking itself as a ‘machine learning first’ company.” *Backchannel.* June 22, 2016. https://backchannel.com/how-google-is-remaking-itself-as-a-machine-learning-first-company-ada63defcb70 (accessed April 7, 2017).

Lichterman, Joseph (2017). “This site is ‘taking the edge off rant mode’ by making readers pass a quiz before commenting.” *NiemanLab.* March 1, 2017*.* http://www.niemanlab.org/2017/03/this-site-is-taking-the-edge-off-rant-mode-by-making-readers-pass-a-quiz-before-commenting/ (accessed June 2, 2017).

Lill, Tobias, Martin U. Müller, Felix Scheidl and Hilmar Schmundt (2012). “Falsche Fans.” *Der Spiegel* 30. http://www.spiegel.de/spiegel/print/d-87482751.html (accessed June 2, 2017).

Lippmann, Walter and Elisabeth Noelle-Neumann (1990). *Die öffentliche Meinung: Reprint des Publizistik-Klassikers*. Bochum.

Lischka, Konrad and Anita Klingel (2017). *Wenn Maschinen Menschen bewerten*. Bertelsmann Stiftung. Gütersloh. https://doi.org/10.11586/2017025 (accessed June 2, 2017).

Lobo, Sascha and Christopher Lauer (2015). *Aufstieg und Niedergang der Piratenpartei: Versuch der Erklärung eines politischen Internetphänomens* (1st ed.). Hamburg.

Lomborg, Stine and Anja Bechmann (2014). "Using APIs for data collection on social media." *The Information Society* (30) 4. 256–265. https://doi.org/10.1080/01972243.2014.915276 (accessed June 2, 2017).

Lumb, David (2015). "Why scientists are upset about the Facebook filter bubble study." *Fast Company.* Aug. 5, 2015. https://www.fastcompany.com/3046111/fast-feed/why-scientists-are-upset-over-the-facebook-filter-bubble-study (accessed June 2, 2017).

Malik, Momin. M. and Jürgen Pfeffer (2016). "Identifying platform effects in social media data." http://mominmalik.com/malik_icwsm2016.pdf (accessed June 2, 2017).

Manjoo, Farhad (2013). "You won't finish this article." *Slate.* June 6, 2013. http://www.slate.com/articles/technology/technology/2013/06/how_people_read_online_why_you_won_t_finish_this_article.html (accessed June 2, 2017).

Meyer, Thomas (2016). "Gleichheit – warum, von was und wie viel?" *Neue Gesellschaft/Frankfurter Hefte* 11. 42–46.

Mierau, Caspar Clemens (2016). "Fake News zum Holocaust sind noch immer Top-Treffer auf Google." *Motherboard.* Dec. 15, 2016. https://motherboard.vice.com/de/article/holocaust-leugnungen-google (accessed April 6, 2017).

Mittelstadt, Brent (2016a). "Auditing for transparency in content personalization systems." *International Journal of Communication* (10). 4991–5002.

Mittelstadt, Brent (2016b). "Auditing for transparency in content personalization systems." *International Journal of Communication* (10). 4991–5002.

Morganti, Luciano, Kristina Irion, Natali Helberger, Katharina Kleinen-von Königslöw and Rob van der Noll (2015). "Regulating the new information intermediaries as gatekeepers of information diversity." *info* (17) 6. 50–71.

Morris, Meredith Ringel (2014). "Social networking site use by mothers of young children." *ACM Digital Library*. 1272–1282. https://doi.org/10.1145/2531602.2531603 (accessed June 2, 2017).

Napoli, Philip M. (2014a). "Automated media: An institutional theory perspective on algorithmic media production and consumption: automated media." *Communication Theory* (24) 3. 340–360. https://doi.org/10.1111/comt.12039 (accessed June 2, 2017).

Napoli, Philip M. (2014b). “On automation in media industries: Integrating algorithmic media production into media industries scholarship.” *Media Industries* (1) 1. 33–38. http://www.mediaindustriesjournal.org/index.php/mij/article/view/14 (accessed June 2, 2017).

Oremus, Will (2016). “Who controls your Facebook feed.” *Slate.* Jan. 3, 2016. http://www.slate.com/articles/technology/cover_story/2016/01/how_facebook_s_news_feed_algorithm_works.html (accessed June 2, 2017).

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

Paus-Hasebrink, Ingrid, Jan-Hinrik Schmidt and Uwe Hasebrink (2009). “Zur Erforschung der Rolle des Social Web im Alltag von Heranwachsenden.” *Heranwachsen mit dem Social Web. Zur Rolle von Web 2.0-Angeboten im Alltag von Jugendlichen und jungen Erwachsenen.* Jan-Hinrik Schmidt, Ingrid Paus-Hasebrink and Uwe Hasebrink, eds. Berlin. 13–40.

Perset, Karine (2010). “The economic and social role of internet intermediaries.” *OECD Digital Economy Papers* No. 171. Paris: OECD Publishing. http://www.oecd-ilibrary.org/science-and-technology/the-economic-and-social-role-of-internet-intermediaries_5kmh79zzs8vb-en (accessed June 2, 2017).

Peters, Bernhard (2002). “Die Leistungsfähigkeit heutiger Öffentlichkeiten – einige theoretische Kontroversen.” *Integration und Medien*. Kurt Imhof, Ottfried Jarren and Roger Blum, eds. Wiesbaden. 23–35. http://link.springer.com/chapter/10.1007/978-3-322-97101-2_3 (accessed June 2, 2017).

Puschmann, Cornelius and Jean Burgess (2013). The politics of Twitter data (SSRN Scholarly Paper No. ID 2206225). Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2206225 (accessed June 17, 2017).

Rainie, Lee, Janna Anderson and Jonathan Albright (2017). “The future of free speech, trolls, anonymity and fake news online.” *PewResearchCenter.* March 29, 2017. http://www.pewinternet.org/2017/03/29/the-future-of-free-speech-trolls-anonymity-and-fake-news-online/ (accessed June 2, 2017).

Ratcliff, Christopher (2016). “WebPromo’s Q&A with Google’s Andrey Lipattsev.” April 6, 2016. https://searchenginewatch.com/2016/04/06/webpromos-qa-with-googles-andrey-lipattsev-transcript/ (accessed April 7, 2017).

Reinbold, Fabian (2017). “Propaganda in der Ära Trump: Die Wahrheit über Fake News.” *Spiegel Online*. Jan. 12, 2017. http://www.spiegel.de/netzwelt/web/donald-trump-die-wahrheit-ueber-fake-news-a-1129628.html (accessed April 13, 2017).

Reinbold, Fabian and Marcel Rosenbach (2016). “Hetze im Netz. Facebook löscht Hass-Kommentare von Berlin aus.” *Spiegel Online.* Jan. 15, 2016. http://www.spiegel.de/netzwelt/web/facebook-neues-loesch-team-geht-gegen-hasskommentare-vor-a-1072175.html (accessed May 26, 2017).

Rosania, Paul (2015). “While you were away... ” Jan. 21, 2017. https://blog.twitter.com/2015/while-you-were-away-0 (accessed March 2, 2017).

Sandvig, Christian, Kevin Hamilton, Karrie Karahalios and Cedric Langbort (2014). “Auditing algorithms: Research methods for detecting discrimination on internet platforms.” Paper presented to *Data and discrimination: Converting critical concerns into productive inquiry*, a preconference at the 64th Annual Meeting of the International Communication Association. May 22, 2014. Seattle WA, USA. https://pdfs.semanticscholar.org/b722/7cbd34766655dea10d0437ab10df3a127396.pdf (accessed June 2, 2017).

Schäfer, Mirko Tobias and Karin van Es (2017). *The datafied society: Studying culture through data*. Amsterdam: University Press Amsterdam. http://dare.uva.nl/aup/en/record/624657 (accessed June 2, 2017).

Schmidt, Jan-Hinrik (2016). “Ethik des Internets.” *Handbuch Medien-und Informationsethik.* Jessica Heesen, ed. Stuttgart. 283–292. http://link.springer.com/chapter/10.1007/978-3-476-05394-7_8 (accessed June 2, 2017).

Schmidt, Jan.-Hinrik, Isabelle Petric, Amelie Rolfs, Uwe Hasebrink and Lisa Merten (2017). *Zur Relevanz von Online-Intermediären für die Meinungsbildung.* Arbeitspapiere des Hans-Bredow-Instituts Nr. 40. Hamburg. http://www.hans-bredow-institut.de/webfm_send/1172 (accessed June 2, 2017).

Schraven, David (2017). “Warum wollt ihr auf Facebook Fakten checken?” *Correktiv.org*. Jan. 20, 2017. https://correctiv.org/blog/2017/01/20/warum-wollt-ihr-fuer-facebook-fakten-checken/ (accessed June 2, 2017).

Schulz, Winfried (1976). *Die Konstruktion von Realität in den Nachrichtenmedien: e. Analyse der aktuellen Berichterstattung.* 1st ed. Freiburg i. Br., Munich.

Schulz, Winfried and Kevin Dankert (2016). “Die Macht der Informationsintermediäre. Erscheinungsformen, Strukturen und Regulierungsoptionen.” Friedrich-Ebert-Stiftung, ed. http://library.fes.de/pdf-files/akademie/12408.pdf (accessed June 2, 2017).

Silverman, Craig (2016). “Here’s how fake election news outperformed real election news on Facebook.” *BuzzFeed News.* Nov. 16, 2016. https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook (accessed Dec. 14, 2016).

Silverman, Craig and Lawrence Alexander (2016). “How teens in the Balkans are duping Trump Supporters with fake news.” *Buzzfeed News*. Nov. 4, 2016. https://www.buzzfeed.com/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo?utm_term=.ckNZqrx7Ln#.vjYnOeyGbZ (accessed June 2, 2017).

Simon, Herbert A. (1979). “Rational decision-making in business organizations.” *The American Economic Review* (69) 4. 493–513.

Spiegel Online (2017). “Google will Fake News in Suchergebnissen bekämpfen.” April 25, 2017. http://www.spiegel.de/netzwelt/web/project-owl-google-will-fake-news-in-suchergebnissen-bekaempfen-a-1144797.html (accessed April 27, 2017).

Stieglitz, Stefan and Linh Dang-Xuan (2012). “Impact and diffusion of sentiment in public communication on Facebook.” *ECIS 2012 Proceedings.* 98. http://aisel.aisnet.org/ecis2012/98 (accessed June 2, 2017).

Stieglitz, Stefan and Linh Dang-Xuan (2013). “Emotions and information diffusion in social media – Sentiment of microblogs and sharing behavior.” *Journal of Management Information Systems* (29) 4. 217–248. https://doi.org/10.2753/MIS0742-1222290408 (accessed June 2, 2017).

Su, Ao-Jan, Y. Charlie Hu, Aleksandar Kuzmanovic and Cheng-Kok Koh (2014). “How to improve your search engine ranking: Myths and reality.” *ACM Transactions on the Web (TWEB)* (8) 2. 8.

Sullivan, Danny (2017). “Google launches new effort to flag upsetting or offensive content in search.” *Search Engine Land*. March 14, 2017. http://searchengineland.com/google-flag-upsetting-offensive-content-271119 (accessed June 2, 2017).

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

Sunstein, Cass R. (2008). *Infotopia: How many minds produce knowledge*. New York NY, Oxford: Oxford University Press.

Taber, Charles. S., Damin Cann and Simona Kucsova (2009). “The motivated processing of political arguments.” *Political Behavior* (31) 2. 137–155. https://doi.org/10.1007/s11109-008-9075-8 (accessed June 2, 2017).

The Coral Project (2016). “About the Coral Project.” http://coralproject.net/about (accessed April 24, 2017).

Trilling, Damian, Marijn van Klingeren and Yarif Tsfati (2016). “Selective exposure, political polarization, and possible mediators: Evidence from the Netherlands.” *International Journal of Public Opinion Research* edw003. https://doi.org/10.1093/ijpor/edw003 (accessed June 2, 2017).

Tufekci, Zeynep (2015). “How Facebook’s algorithm suppresses content diversity (modestly) and how the newsfeed rules your clicks.” May 5, 2015. https://medium.com/message/how-facebook-s-algorithm-suppresses-content-diversity-modestly-how-the-newsfeed-rules-the-clicks-b5f8a4bb7bab (accessed April 14, 2017).

Tutt, Andrew (2016). An FDA for algorithms. (SSRN Scholarly Paper No. ID 2747994). Rochester NY: Social Science Research Network. https://papers.ssrn.com/abstract=2747994 (accessed June 2, 2017).

Tversky, Amos and Daniel Kahneman (1974). “Judgment under uncertainty: Heuristics and biases.” *Science* (185) 4157. 1124–1131.

Twitter (2017). “About your Twitter timeline.” https://support.twitter.com/articles/164083#settings (accessed June 2, 2017).

Underwood, Mimi (2015). “Updating our search quality rating guidelines.” *Google Webmaster Central Blog.* Nov. 19, 2015. https://webmasters.googleblog.com/2015/11/updating-our-search-quality-rating.html (accessed April 6, 2017).

US Supreme Court (1969). Brandenburg v. Ohio 395 U.S. 444. June 9, 1969. https://supreme.justia.com/cases/federal/us/395/444/ (accessed June 2, 2017).

van der Linden, Sander, Anthony Leiserowitz, Seth Rosenthal and Edward Maibach (2017). “Inoculating the public against misinformation about climate change.” *Global Challenges* Feb. 27, 2017. https://doi.org/10.1002/gch2.201600008 (accessed June 2, 2017).

van Deursen, Alexander J. A. M. and Jan A. G. M. van Dijk (2009). “Using the Internet: Skill related problems in users’ online behavior.” *Interacting with Computers* (21) 5–6. 393–402. https://doi.org/10.1016/j.intcom.2009.06.005 (accessed June 2, 2017).

Vieth, Kilian and Ben Wagner (2017). *Teilhabe, ausgerechnet*. Hrsg. Bertelsmann Stiftung. Gütersloh. https://doi.org/10.11586/2017027 (accessed June 2, 2017).

Vike-Freiberga, Vaira, Herta Däubler-Gmelin, Ben Hammersley and Luis Miguel Poiares Pessoa Maduro (2013). The report of the High Level Group on Media Freedom and Pluralism. https://ec.europa.eu/digital-single-market/sites/digital-agenda/files/HLG%20Final%20Report.pdf (accessed June 2, 2017).

Volokh, Eugene and Donald M. Falk (2012). *First amendment protection for search engine search results – White paper commissioned by Google.* (SSRN Scholarly Paper No. ID 2055364). Rochester NY: Social Science Research Network. https://papers.ssrn.com/abstract=2055364 (accessed June 2, 2017).

Weischenberg, Siegfried (2001). *Nachrichten-Journalismus: Anleitungen und Qualitäts-Standards für die Medienpraxis*. Wiesbaden. http://link.springer.com/openurl?genre=book&isbn=978-3-322-80408-2 (accessed June 2, 2017).

Williams, Raymond (1974). *Television: Technology and cultural form*. London: Fontana.

Wojcieszak, Magdalena (2010). "'Don't talk to me': Effects of ideologically homogeneous online groups and politically dissimilar offline ties on extremism." *New Media & Society.* Jan. 28, 2010. http://nms.sagepub.com/content/early/2010/01/28/1461444809342775.abstract (accessed June 2, 2017).

Xu, Jie, Akos Lada and Vibhi Kant (2016). "Showing you more personally informative stories." Aug. 11, 2016. http://newsroom.fb.com/news/2016/08/news-feed-fyi-showing-you-more-personally-informative-stories/ (accessed March 1, 2017).

Zarsky, Tal (2016). "The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision-making." *Science, Technology, & Human Values* (41) 1. 118–132. https://doi.org/10.1177/0162243915605575 (accessed June 2, 2017).

Zeifman, Igal (2017). "Bot Traffic Report 2016." Jan. 24, 2017. https://www.incapsula.com/blog/bot-traffic-report-2016.html (accessed June 2, 2017).

Zhang, Cheng and Si Chen (2016). "Using qualitative feedback to show relevant stories." Feb. 1, 2016. https://newsroom.fb.com/news/2016/02/news-feed-fyi-using-qualitative-feedback-to-show-relevant-stories/ (accessed March 2, 2017).

Zollo, Fabiana, Alessandro Bessi, Michaela Del Vicario, Antonio Scala, Guido Caldarelli, Louis Shekhtman, Shlomo Havlin and Walter Quattrociocchi (2015). "Debunking in a world of tribes." *arXiv preprint arXiv:1510.04267*. https://arxiv.org/abs/1510.04267 (accessed June 2, 2017).

Zuckerberg, Mark (2016, November 12). "I want to share some thoughts on Facebook and the election." https://www.facebook.com/zuck/posts/10103253901916271 (accessed June 15, 2017).

Zweig, Katharina Anna (2017a). "Watching the watchers: Epstein and Robertson's 'search engine manipulation effect.'" *Algotithm Watch* April 7, 2017. https://algorithmwatch.org/watching-the-watchers-epstein-and-robertsons-search-engine-manipulation-effect/ (accessed April 7, 2017).

Zweig, Katharina Anna (2017b, forthcoming). *Wo maschinelle Prognosen irren können*. Bertelsmann Stiftung. Gütersloh.

10 Executive Summary

Essentially this paper answers three key questions relating to public discourse and opinion in the digital age:

1. *Media transformation: How is public discourse changing because of the new digital platforms through which many people now receive socially relevant information?*
   When all age groups are considered, intermediaries driven by algorithmic processes, such as Google and Facebook, have had a large but not defining influence on how public opinion is formed, compared to editorially driven media such as television. These intermediaries judge the relevance of content based on the public's immediate reaction to a much greater degree than do traditional media.
2. *Social consequences: In terms of quality and diversity, is the information that reaches people via these new channels suitable for democratic decision-making processes and does it promote participation?*
   Use of these intermediaries for forming public opinion is leading to a structural change in the public sphere. Key factors here are the algorithmic processes used as a basic formational tool and the leading role that user reactions play as input for these processes. Since they are the result of numerous psychological factors, these digitally assessed and, above all, impulsive public reactions are poorly suited to determining relevance as defined by traditional social values. Until now, these values, such as truth, diversity and social integration, have served in Germany as the basis for public opinion as formed by editorial media.
3. *Solutions: How can the new digital platforms be designed to ensure they promote participation?*
   Algorithms that sort content and personalize how it is assembled form the core of the complex, interdependent process underlying public discourse and the formation of public opinion. That is why solutions must be found here first. The most important areas that will need action in the foreseeable future are facilitating external research and evaluation, strengthening the diversity of algorithmic processes, anchoring guiding principles (e.g. by focusing on professional ethics) and increasing awareness among the public.

1. Media transformation: Intermediaries are a relevant but not determining factor for the formation of public opinion.

Numerous studies have shown that so-called *intermediaries* such as Google and Facebook play a role in the formation of public opinion in numerous countries including Germany. For example, 57% of German Internet users receive politically and socially relevant information via search machines or social networks. And although the share of users who say that social networks are their most important news source is still relatively small at 6% of all Internet users, this figure is significantly higher among younger users. It can thus be assumed that these types of platforms will generally increase in importance. The formation of public opinion is "no longer conceivable without intermediaries," as researchers at the Hamburg-based Hans Bredow Institute put it in 2016.

The principles that these intermediaries use for shaping content are leading to a structural shift in the public sphere. Key aspects are:

- **Decoupling of publication and reach:** Anyone can make information public, but not everyone finds an audience. Attention is only gained through the interaction of people and algorithmic decision-making (ADM) processes.
- **Detachment from publications**: Each story or post has its own reach. Inclusion in a publication reduces the chances of becoming part of the content provided by intermediaries.
- **Personalization**: Users receive more information about their specific areas of interest.
- **Increased influence of public on reach**: User reactions influence ADM processes in general and the reach of each article or post.
- **Centralization of curators**: Intermediaries offer much less diversity than do traditionally curated media.
- **Interplay of human and machine-based curation**: Traditionally curated media disseminate content via intermediaries and use the resulting reactions on intermediary sites as a measure of public interest.

The comparison of processes in the figure below shows the key role now being played by user reactions and algorithmic processes in public communication and the formation of public opinion. User reactions and algorithmic processes determine the attention received via intermediaries. Our hypothesis is that these reactions and processes cannot be definitively included in a linear causal chain.

Figure 1: Organization of an algorithmically structured public sphere

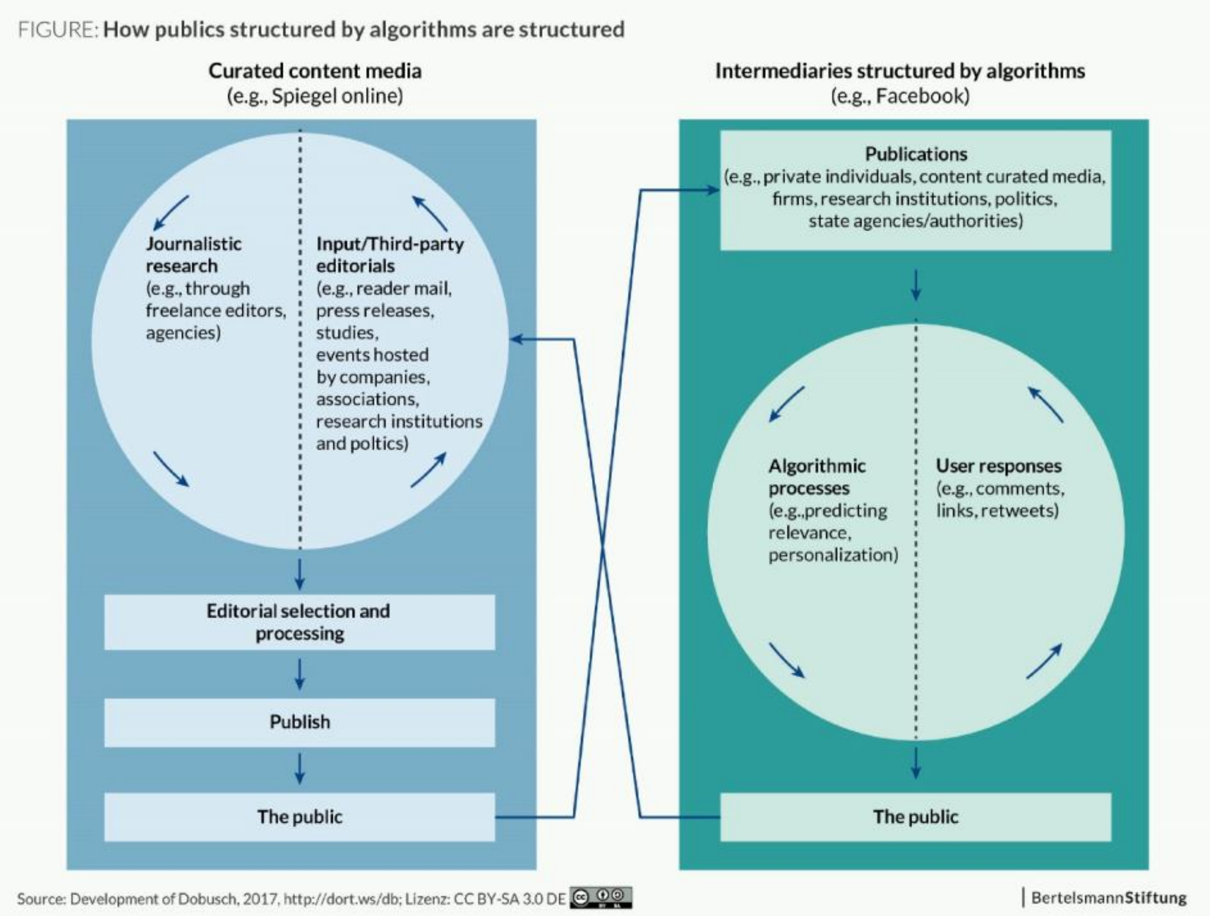


FIGURE: How publics structured by algorithms are structured
Curated content media
(e.g., Spiegel online)
Journalistic research (e.g., through freelance editors, agencies)
Input/Third-party editorials (e.g., reader mail, press releases, studies, events hosted by companies, associations, research institutions and poltics)
Editorial selection and processing
Publish
The public
Intermediaries structured by algorithms
(e.g., Facebook)
Publications (e.g., private individuals, content curated media, firms, research institutions, politics, state agencies/authorities)
Algorithmic processes (e.g.,predicting relevance, personalization)
User responses (e.g., comments, links, retweets)
The public
| BertelsmannStiftung

Source: Development of Dobusch, 2017, http://dort.ws/db; Lizenz: CC BY-SA 3.0 DE

Google, Facebook and other intermediaries are already playing an important role in the public discourse, even though as a rule these platforms were not originally conceived to supply consumers with media content from journalistic organizations. They use technical systems instead to decide whether certain content taken from a huge pool of information could be interesting or relevant to a specific user. These systems were initially intended – in the case of search engines – to identify websites, for example, that contain certain information or – in the case of social networks – to display in a prominent position particularly interesting messages or photos from a user's own circle of friends. In many cases they therefore sort content according to completely different criteria than editors at a daily newspaper or a magazine would. “Relevant” means something different to Google than it does to Facebook, and both understand it differently than the editors at Spiegel Online or Sueddeutsche.de.

The intermediaries use numerous variables to calculate the relevance of individual items. These variables range from basic behavioral metrics such as scrolling speed or how long a page is viewed to the level of interaction among multiple users in a social network. When someone with whom a user has repeatedly communicated on Facebook posts content, the probability is higher that the user will be shown this content than they would if someone posts with whom the user has, theoretically, a digital connection, but with whom the user has never truly had contact. The signals that other users send – often unknowingly – are also included in assessments of relevance, whether they be the insertion of links, clicks on links, clicks on the “like” button, forwards, shares or the number of comments that certain content receives.

2. Social consequences: As used by the most relevant intermediaries currently forming public opinion, algorithmic processes evaluate users' reactions to content. They promote and reinforce a manner of human cognition that is susceptible to distortion, and they themselves are susceptible to technical manipulation.

The metrics signaling relevance – about which platform operators are hesitant to provide details because of competition-related factors or other reasons – are potentially problematic. This is true, first, because the operators themselves are constantly changing the metrics. Systems such as those used by Google and Facebook are being altered on an ongoing basis; the operators experiment with and tweak almost every aspect of the user interface and other platform features in order to achieve specific goals such as increased interactivity. Each of these changes can potentially impact the signals that the platforms themselves capture to measure relevance.

A good example is the People You May Know feature used by Facebook, which provides users with suggestions for additional contacts based on assessments of possible acquaintances within the network. When this function was introduced, the number of links made every day within the Facebook community immediately doubled. The network of relationships displayed on such platforms is thus dependent on the products and services that the operators offer. At the same time, the networks of acquaintances thus captured also become variables in the metrics determining what is relevant. Whoever makes additional "friends" is thus possibly shown different content.

A further problem stemming from the metrics collected by the platform operators is the type of interaction for which such platforms are optimized. A key design maxim is that interactions should be as simple and convenient as possible in order to maximize the probability of their taking place. Clicking on a "like" button or a link demands almost no cognitive effort, and many users are evidently happy to indulge this lack of effort. Empiric studies suggest, for example, that many articles in social networks forwarded with a click to the user's circle of friends could not possibly have been read. Users thus disseminate media content after having seen only the headline and introduction. To some extent they deceive the algorithm and, with it, their "friends and followers" into believing that they have engaged with the text.

The ease of interaction also promotes cognitive distortions that have been known to social psychologists for decades. A prime example is the availability heuristic: If an event or memory can easily be recalled, it is assumed to be particularly probable or common. The consequence is that users frequently encounter unread media content that has been forwarded due to a headline, and the content is thus later remembered as being "true" or "likely." This is also the case when the text itself makes it clear that the headline is a grotesque exaggeration or simply misleading.

Other psychological factors play an important role here, for example the fact that people use social media in particular not only for informational purposes, but also as a tool for identity management, with some media content being forwarded only to demonstrate the user's affiliation with a certain political camp, for example. Moreover, the design of many digital platforms explicitly and intentionally encourages a fleeting, emotional engagement with content. Studies of networking platforms indeed show that *content which rouses emotion* is commented on and shared particularly often – above all when negative emotions are involved.

Such an emotional treatment of news content can lead to increased **societal polarization,** a hypothesis for which initial empirical evidence already exists, especially in the United States. At the same time, however, such polarizing effects seem to be dependent on a number of other factors such as a country's electoral system. Societies with first-past-the-post systems such as the US are potentially more liable to extreme political polarization than those with proportional systems, in which ruling coalitions change and institutionalized multiparty structures tend to balance out competing interests. Existing societal polarization presumably influences and is influenced by the algorithmic ranking of media content. For example, one study shows that Facebook users who believe in conspiracy theories tend over time to turn to the community of conspiracy theorists holding the same views. This process is possibly exacerbated by algorithms that increasingly present them with the relevant content. These systems could in fact result in the creation of so-called **echo chambers,** at least among people with extremist views.

Technical manipulation can also influence the metrics that intermediaries use to ascertain relevance. So-called bots – partially self-acting software applications that can be disguised as real users, for example in social networks – can massively distort the volume of digital communication occurring around certain topics. According to one study, during the recent presidential election in the US, 400,000 such bots were in use on Twitter, accounting for about one-fifth of the entire discussion of the candidates' TV debates. It is not clear to what extent these types of automated systems actually influence how people vote. What is clear is that the responses they produce – clicks, likes, shares – are included in the relevance assessments generated by ADM systems. Bots can thus make an article seem so interesting that an algorithm will then present it to human users.

In sum, it can be said that the relevance assessments that algorithmic systems create for media content do not necessarily reflect criteria that are desirable from a societal perspective. Basic values such as truthfulness or social integration do not play a role. The main goal is to increase the probability of an interaction and the time users spend visiting the relevant platform. Interested parties whose goal is a strategic dissemination of disinformation can use these mechanisms to further their cause: A creative, targeted lie can, on balance, prove more emotionally "inspiring" and more successful within such systems – and thus have a greater reach – than the boring truth.

3. Solutions: Algorithmic sorting of content is at the heart of the complex interdependencies affecting public discourse in the digital context. This is where solutions must be applied.

The last section of this paper contains a series of possible solutions for these challenges. A first goal, one that is comparatively easy to achieve, is making users more aware of the processes and mechanisms described here. Studies show that users of social networking platforms do not even know that such ranking algorithms exist, let alone how they work. Educational responses, including in the area of continuing education, would thus be appropriate, along with efforts to increase awareness of disinformation and to decrease susceptibility to it, for example through the use of fact-based materials.

Platform operators themselves clearly have more effective possibilities for intervention. For example, they could do more to ensure that values such as appropriateness, responsibility and competency are adhered to when the relevant systems are being designed and developed. A medium-term goal could be defining industry-wide professional ethics for developers of ADM systems.

Moreover, researchers who do not work for platform operators should be in a position to examine and evaluate the impact being made by the operators' decisions. Until now it has been difficult if not impossible for external researchers or government authorities to gain access to the required data, of which operators have vast amounts at their disposal. Neither the design decisions made by platform operators nor the impacts of those decisions on individual users are transparent to any significant degree. Systematic distortions, for example in favor of one political viewpoint or another, are difficult to identify using currently available data. More transparency – through a combination of industry self-regulation and, where necessary, legislative measures – would make it possible to gain an unbiased understanding of the actual social consequences of algorithmic ranking and to identify potential dangers early on. Making it easier to conduct research would also stimulate an objective, solution-oriented debate of the issue and could help identify new solutions. Measures like these could also make it easier to design algorithmic systems that increase participation. This would foster a more differentiated view of algorithmic processes and could increase trust in those systems that are designed to benefit all of society.

11 About the authors

Konrad Lischka (born in 1979) has been writing about the digital society since 1999. His publications include books, essays and blog posts. After receiving his degree (Diplom) in journalism and completing his studies at the Deutsche Journalistenschule (German School of Journalism), he served as editor-in-chief of the magazine *bücher* and was deputy director of the Web section at Spiegel Online. He then began addressing issues relating to media and web policy as the Digital Society officer at the state chancellery of North Rhine–Westphalia. Since 2016 he has been a project manager at the Bertelsmann Stiftung for the Participation in a Digitized World project.

Christian Stöcker, PhD, (born in 1973) studied psychology in Würzburg and Bristol, receiving his doctorate in cognitive psychology in 2003. He then studied cultural criticism at the Bavarian Theater Academy in Munich while simultaneously writing for the *Süddeutsche Zeitung* and *Die Zeit* newspapers and for the SPIEGEL ONLINE, Germany's largest non-tabloid newssite. He has been writing for the Science and Web sections at SPIEGEL ONLINE since 2005. In 2011 he became head of the site's Web, Tech and Media section. He has been a professor of digital communication at the Hamburg University of Applied Sciences since 2016.

Address | Contact

Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
33311 Gütersloh
Germany
Telephone +49 5241 81-81216

Konrad Lischka
Digitization Taskforce
Telephone +49 5241 81-81216
konrad.lischka@bertelsmann-stiftung.de

www.bertelsmann-stiftung.de