



# Public Discourse in Online Environments

Contributions to description, measurement and communication  
behavior change in political online ecosystems

Berlin, May 2023

**Lisa Oswald**

Dissertation submitted to the Hertie School  
in partial fulfilment of the requirements for the degree of

**Doctor rerum politicarum (Dr. rer. pol.)**

in the

**Doctoral Program in Governance**

## Advisors

### First advisor

Prof. Dr. Simon Munzert  
Professor of Data Science and Public Policy  
Hertie School, Berlin, Germany

### Second advisor

Prof. Dr. Ralph Hertwig  
Professor of Psychology  
Director Max Planck Institute for Human Development, Berlin, Germany

### Third advisor

Prof. Andrew Guess, PhD  
Assistant Professor of Politics and Public Affairs  
Princeton University, New Jersey, USA

## Evaluators

### First Evaluator

Prof. Dr. Fabrizio Gilardi  
Professor of Policy Analysis  
University of Zurich, Zurich, Switzerland

### Second evaluator

Prof. Dr. Yannis Theocharis  
Professor of Digital Governance  
Technical University of Munich, Munich, Germany

## Summary

The public discourse is an essential element of democracy. Today, political information consumption is largely mediated through digital technologies and important political discussions have moved to online platforms. This dissertation examines different elements and dynamics of the online public discourse within five research articles. First, a systematic review synthesizes evidence on the broader relationship between digital media and democracy, which highly depends on political context. While most beneficial effects of information access and political participation can be found in autocratic regimes and emerging democracies, there is more reason for concern in established democracies where negative effects on political trust, populism and polarization accumulate.

Next, using digital trace data combined with survey data to zoom into the political information ecosystem in Germany, I find that it consists of far more than news media, but that in-depth political discussions also take place in seemingly unpolitical niches of the internet. While the user base of online public broadcasting consists of a comparably narrow audience, most citizens consume political content via mainstream information hubs that can provide a common basis for public discourse. The most predictive variable for selection into political information *consumption* is political knowledge, not partisanship. However, people who actively *write* comments online show more extreme political opinions, highlighting differences between active and passive political engagement online. This differential selection pattern is further mirrored in the final project, where high levels of social media activity also relate to stronger comment *toxicity*.

An agenda to reduce online toxicity requires an efficient method to measure the quality of online discussions. Using discussion data from Reddit, I compare simple computational measures to manual content coding and find that an automated toxicity classifier combined with the length of comments corresponds to established indicators of deliberative quality. I then develop a preemptive strategy with the aim to reduce toxicity in political discussions among highly active social media users. Testing various behavioral interventions, I find that simple ‘nudges’ in the user interface do not reduce the toxicity of comments. However, mixed findings on a more complex ‘boosting’ intervention, explaining the role of empathy and perspective taking for political discussions, evince opportunities to open up online public discourse for those who may otherwise be silenced by high levels of toxicity.

Especially when combined with established approaches of survey methodology and content analysis, computational social science commands many tools to map elements and dynamics of the online public discourse from a holistic perspective. This does not only bear potential to reconnect theoretical ideals of the public sphere with real life online, but it also enables the generation of evidence for policymakers and civil society to keep up with the pace of the fast moving target of digital media.

## Contents

<b>Introduction</b>	<b>1</b>
Understanding the online public discourse . . . . .	4
The online public discourse from different theoretical angles . . . . .	8
Political online ecosystems . . . . .	14
Selection and representation in the online public discourse . . . . .	16
Discussion quality and toxicity in the online public discourse . . . . .	18
Computational assessment of deliberative quality . . . . .	19
Testing preemptive empathy interventions to reduce toxicity . . . . .	20
<b>Outlook</b>	<b>24</b>
Consequences for representation . . . . .	27
Potentials of computational social science . . . . .	29
<b>Publications</b>	<b>45</b>
P1. A systematic review of worldwide causal and correlational evidence on digital media and democracy . . . . .	45
P2. More than News! Mapping the Deliberative Potential of Political Online Ecosystems with Digital Trace Data. . . . .	121
P3. Beyond the tip of the iceberg? Exploring Characteristics of the Online Public with Digital Trace Data. . . . .	170
P4. Automating the Analysis of Online Deliberation? A Comparison of Manual and Computational Measures Applied to Climate Change Discussions. . . . .	199
P5. Effects of Preemptive Empathy Interventions on Reply Toxicity among Highly Active Social Media Users . . . . .	233
<b>List of included publications</b>	<b>266</b>

## Introduction

Public discourse is essential to public opinion formation and eventually translates into political judgement and action. Today, political information consumption is largely mediated through digital technologies (Newman et al., 2019) and important political discussions moved to online environments. Life without digital media is hard to imagine and, at the same time, established democracies are increasingly under pressure (Alizada et al., 2022; Lorenz-Spreen, Oswald, et al., 2023).

Based on a dataset that includes worldwide observations between 1789 and 2021, the most recent V-Dem report expresses deep concern about a decline of democracy across the past decade. “This intensifying wave of autocratization around the world highlights the need for new initiatives to defend democracy.” (Alizada et al., 2022, p. 5). Regarding public discourse, they further argue that “A signal of toxic polarization, respect for counter arguments and associated aspects of the deliberative component of democracy got worse in more than 32 countries – another increase from only 5 nations in 2011” (Alizada et al., 2022, p. 6).

Public discourse and the underlying ecosystem of political information has a central position in the logic of any democratic system and is, at the same time, the democratic element strongest affected by the digital transformation. The plethora of interlinked arenas of public discourse online and offline can be understood broadly as ‘public sphere’ (Habermas et al., 1974; Jungherr & Schroeder, 2021) or when tied to recent developments in deliberation research, more specifically as ‘deliberative system’ (Bächtiger & Parkinson, 2019). The public sphere is conceptualized as mediating space between civil society and politics that includes the multifaceted media system as set of infrastructures, also including digital media, that enable the publication, distribution, consumption and discussion of political information that allow people to exercise their rights and duties as citizens (Jungherr & Schroeder, 2021).

The role of digital media in democratic regression is a highly contested issue that attracts both academic and public interest. While some draw darkest scenarios predominantly blaming social media (Haidt, 2022), others argue that the link between digital media and democracy is largely unclear (Weiss-Blatt, 2022). It is crucial to distinguish correlation from causation when examining such complex phenomena linking society, technology and politics. Empirical evidence on the link between digital media and democracy is much more complex than some may portray it; however, the field is far from empty.

In 2021, together with Philipp Lorenz-Spreen, Stephan Lewandowsky and Ralph Hertwig, I conducted a systematic review of causal and correlational evidence from  $N = 496$  articles on the link between digital media use and different political variables (Lorenz-Spreen et al., 2021, first publication included in this dissertation, see p. 45).

From a citizen-centered perspective on a deliberative conception of democracy, the political variables under review can be broadly categorized into different stages of the political process. At the information stage as foundation or input for public discourse, we examined the diversity of exposure to political information and the level of misinformation. At the communication stage as core of public discourse, we considered political expression, the level of hate speech and the tendency to connect to similar others in social networks. At the participation stage, as result of opinions formed within public discourse, we examined political participation, support for populists, opinion polarization and political extremism. Political trust and knowledge could be regarded as subordinate variables as they are relevant in any stage of the process.

Overall, we find that there is no binary answer to the question of whether digital media is good or bad for democracy but that the effect highly depends on the political context, the specific political variable, the type of digital media as well as measurement choices. For example, we find largely positive associations between

digital media use and political participation, as well as between digital media use and information consumption, relationships that are likely beneficial for democracy. These beneficial relationships are most pronounced in autocratic regimes and emerging democracies in South America, Africa and Asia. However, other consistently reported associations, such as declining political trust, advantages for populists, and growing polarization, links to digital media that are likely detrimental to democracy, are more pronounced in established democracies such as Europe and the United States.

The fact that, for example, a loss of trust in an autocratic regime might be a necessary first step towards democratization, highlights the importance of political context for the interpretation of effects. Heterogeneous findings for different types of digital media were, for example, found for political knowledge: while overall internet access was largely positively associated with political knowledge, social media use also came with negative effects for political knowledge, likely through mechanisms of distraction and the misperception that ‘news finds me’ online, implying that active information search is unnecessary.

Another contribution of our review is that it describes the state of the research field: little causal evidence, a heavy US focus, mostly survey-based correlational evidence on media use and self-reported political outcomes. This makes it often hard to identify the reasons *why* digital media use is, for example, associated with decreasing trust in politics, or how observations of online hate speech and misinformation relates to the broader online public discourse.

While causal evidence largely mirrored the much greater body of correlational evidence, it became clear that many findings highly depend on how variables were measured. For example, the phenomenon of the much discussed ‘echo chambers’ (Sunstein, 2002), suggesting that digital media users are siloed in attitude reinforcing online environments, appears more pronounced in studies considering social network data in which people are predominantly connected to

like-minded others, whereas, when examining people's media diets through web browsing and survey data, they appear quite diverse with people being regularly exposed to cross cutting content. This highlights the importance to explore and combine various types of data to get a nuanced understanding of phenomena that are crucially relevant for democracy and public discourse.

Particularly the methodological observations of the systematic review highlight two key foci for the following projects of this dissertation, (1) issues of measurement and (2) the combination of descriptive analyses of existing structures with experiments that allow causal conclusions about specific phenomena of the online public discourse. In the following section, I motivate and outline the core empirical contributions of this dissertation which, among the complex interplay between digital media and democracy, focuses on elements and dynamics within the online public discourse.

## **Understanding the online public discourse**

While offline and online life are getting increasingly inseparable, engagement with political issues online has direct implications for offline politics. In the wake of this, political discussions between citizens in online spaces are by no means just a nice-to-have extension of the offline public sphere with little implications for 'real life'. For example, journalists as the fourth democratic power regularly pick up political discussions on digital platforms and amplify them via news media. Weblog discussions can influence the agendas of mass media and vice versa (Messner & Distaso, 2008; Wallsten, 2007) and Twitter networks of journalists are important links between digital and mass media (Neuberger et al., 2019). More directly, hashtag campaigns initiated decentrally by citizens can impact social and legal change (Nenoff, 2020; Sawyer & Gampa, 2018), whereas online hate speech can translate into severe offline hate crime (Müller & Schwarz, 2018).

Overall, the connection between political information consumption and



political participation in the form of, for example, voting or protesting is long established in political science literature (Galston, 2001; Persson, 2015). For example, this link can be established through a direct cognitive pathway in which greater political knowledge translates into political action, or through an indirect pathway via increased political efficacy. More recent literature regarding political information consumption via digital media reports largely positive associations with political participation (Lorenz-Spreen, Oswald, et al., 2023).

Facing considerable links between digital media and democracy and taking into account that the infrastructure of the online public discourse is operated by fast moving actors without political legitimization, we need to better understand arenas, actors and discussions themselves. Where are political information presented online, which platforms are relevant? Who selects into the online public discourse and can we speak of equal representation? How can we assess the quality of the online public discourse and is there a way to improve it?

Fig. 1 summarizes this dissertation's conceptualization of different elements and dynamics of the online public discourse. The upper section is inspired by the idea of a 'deliberative system' (Bächtiger & Parkinson, 2019) in which the degree of public 'deliberation' is an emergent property of potentially complementary discussions in different interlinked arenas with different degrees of formality and deliberative quality. In my conceptualization, the political online ecosystem of websites and the online public are two sub-systems interlinked through mechanisms of selection.

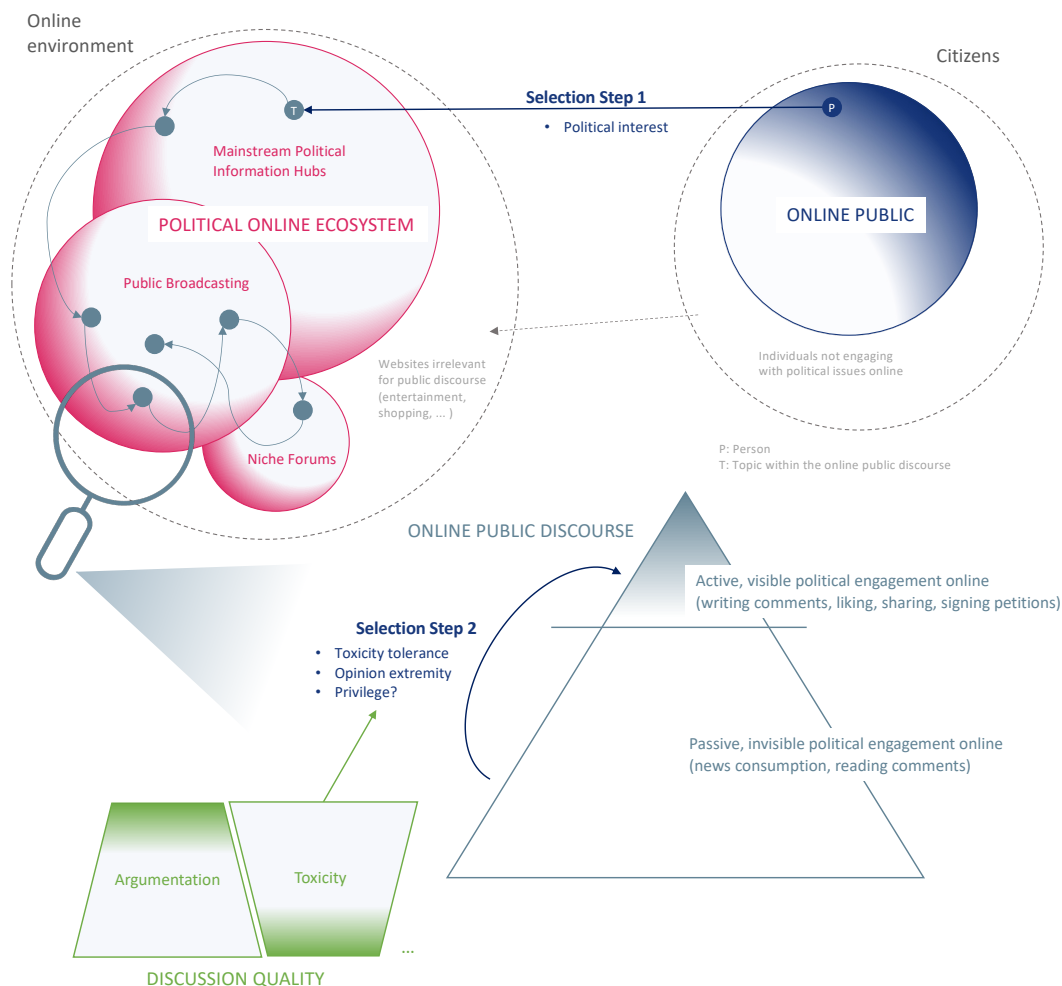
On the one hand, we have the overall online environment, consisting of a variety of interlinked websites. While the majority of websites is irrelevant for online public discourse (e.g. online shopping, banking and various forms of digital entertainment), a subset of the web hosts political information and enables political expression and discussion, while a small fraction of sites even enables political participation (e.g. online petitioning platforms or municipal sites that connect

citizens to relevant political actors). I consider the collection of websites that do play a role in the online public discourse as the *political online ecosystem*.

On the other hand, we have the *online public*, a subset of the overall citizenry that selects into engaging with issues of the online public discourse through sites of the political online ecosystem (Oswald et al., 2022). Just as members of the online public are highly heterogeneous in terms of demographic characteristics, political attitudes etc., websites constituting the political online ecosystem are also characterized by varying degrees of ‘deliberative potential’ (Oswald, 2022b).

Finally, the *online public discourse* includes the political issues featured by the political online ecosystem and discussed by the online public, as well as the links and flows between issues and discussions in the sense of agenda setting and discourse dynamics. Besides the substance of political issues and the semantic links and temporal flows between them, the quality of communication, including the level of argumentation and toxicity, can be characterizing variables of the online public discourse (Oswald, 2022a). The final project of this dissertation targets the toxicity element of the online public discourse (Oswald, 2023) using empathy and perspective taking based interventions with the idea to open up the second selection step to the extent that fewer politically interested individuals are kept from active participation through high levels of toxicity in political discussions online.

## Elements and Dynamics in the Online Public Discourse



**Figure 1**

Graphical summary of this dissertation's main contribution. Conceptualization including (1) the online political ecosystem of politically relevant websites, (2) the online public as politically engaged subset of citizens, and (3) the online public discourse including the discussions of salient political issues. The sites of the online political ecosystem can be characterized by varying, potentially complementary, degrees of information density, discussion quality, and possibilities for political participation; the online public is characterized by varying demographic profiles as well as diverse degrees of political interest, political attitudes etc.; the online public discourse is characterized by varying levels of deliberative quality, including the levels of toxicity and argumentation. Different elements of the online public discourse are connected through links of a two-staged selection process.

## **The online public discourse from different theoretical angles**

The online public discourse can be conceptualized from different theoretical angles that are characterized by a continuum of normative premises, conceptual complexity and empirical operationalizability. This dissertation is influenced by deliberative democratic theory but often takes a more pragmatic perspective of the online public discourse, focusing on issues of information consumption and selection.

According to Habermas et al. (1974), the public sphere is a social space in which citizens can gather to freely discuss and form public opinion. Political communication among citizens within the public sphere is often referred to as 'deliberation' - that, however, only deserves its name when fulfilling a set of discourse quality criteria (Steenbergen et al., 2003). These criteria include, for example, the justification of arguments, reciprocity between discussants and a certain level of respect.

Decades after Jürgen Habermas established the political concept of the public sphere that shaped research agendas from various disciplines of the social sciences, he expressed his thoughts on the digital transformation of the public sphere (Habermas, 2022). His concerns about a fragmentation of the online public sphere through mighty private platforms, the spread of misinformation and hate speech, bearing the danger of political polarization mostly echoes the pessimist wing of current debates (Haidt, 2022).

Throughout his line of argument, he puts particular emphasis on the concept of deliberation, which he regards as the core of a pluralistic democracy. According to Habermas, the more heterogeneous a society, the more crucial is a deliberative process to establish public opinion. In this process, the media have the important role to select the issues that are important, and to guide public attention towards a shared informational baseline. Based on these common sources of information, every citizen builds their individual opinion, which can then be aggregated to inform political decisions. According to (Habermas, 2022), the core to the deliberative

process of public opinion formation are conflicting positions that eventually give power to the best rationales.

His diagnosis of a severe regression of this deliberative foundation for all established democracies is in line with the V-Dem report (Alizada et al., 2022). Whether this regression can be causally linked to the rise of digital media, he notes, is an open empirical question—a question we partly address with our systematic review (Lorenz-Spreen, Oswald, et al., 2023). Furthermore, while changes in media use that mirror a structural change of the public sphere have been observed in past decades, he notes that changes of the deliberative quality of discussion are again an open empirical question. To answer this question in the future, in this dissertation, I provide a systematic comparison of established deliberative quality indicators and different automated measures that allow the assessment of communication quality at larger scale. Despite Habermas’ criticism of various current developments within the political and public culture that he traces back to digital media, he acknowledges their empowering benefits for citizens, especially in oppressive regimes which aligns with our review (Lorenz-Spreen, Oswald, et al., 2023).

In recent years, the classical Habermasian definition of deliberation, strongly focusing on the sound justification of arguments and the civility of a discussion, has been criticized for representing an elitist perspective that places unrealistic demands on citizens. Some voice deep skepticism about the potentials of deliberative democracy: “Oriented by its roots in classical liberalism, deliberative democratic theory is premised on felicitous but unrealistic assumptions regarding the capacities of individual citizens and their communicative competence.” (Rosenberg, 2014, p. 98).

This attitude was however countered in a prominent paper in *Science*, arguing that “It is rare that deliberative development happens spontaneously. [...] The citizenry is quite capable of sound deliberation. But deliberative democratization will not just happen” (Dryzek et al., 2019, p. 1146). The authors

argue that social media plays a major role in deliberative systems, and often amplifies uncivil politics and pathological communication. “Yet the problem is not social media per se but how it is implemented and organized” (Dryzek et al., 2019, p. 1145).

While early research on deliberation proposed a normatively optimistic but theoretically rather abstract understanding of deliberation, including high demands on citizens (Chambers, 1996; Cohen, 1989; Dryzek, 1987; Fishkin, 1991; Habermas, 1995; Rawls, 1995), deliberation research developed (Esau, 2022). New concepts such as humor, emotion and empathy were introduced and the justification of arguments was opened to descriptions of personal experience. Furthermore, the field acknowledged that relevant political discussions are taking place in online environments and that this might require a more systemic perspective (Bächtiger & Parkinson, 2019; Strandberg & Grönlund, 2018).

The shift in measurement decisions also considerably changed the evaluation of the state of democratic discourse in the digital domain. When using a narrow Habermasian definition, most researchers come to largely negative conclusions about the deliberative quality of the online public discourse (Strandberg, 2008; Wilhelm, 1998). When incorporating more inclusive concepts like empathy, humor and emotion, researchers come to more optimistic conclusions (Graham, 2012).

The reality might lie somewhere in the middle, especially when taking into account that most people are largely disengaged with politics. For example, when considering desktop browsing histories, only 1% of content visited online is political in nature (Oswald, 2022b). Considering predictors for online political engagement, there is heavy selection along political interest and knowledge (Oswald et al., 2022). Furthermore, this finding implies the motivation of a subgroup of citizens to engage with political content online and speaks against a uniform distribution of preconditions for deliberation across the public—a finding that bridges the premises of pessimists and optimists, since the motivation to engage with political issues is a

fundamental component for the development of civic competencies that are necessary to constructively engage in public discourse. Motivation is also fundamentally necessary to ‘hear the other side’, show respect to others and tolerate opposing views in democratic societies (Mutz, 2006).

While the distinction between the classical Habermasian and a modern, more inclusive definition of deliberation is now widely accepted in contemporary literature (Bächtiger et al., 2010), the more inclusive tools are not the empirical mainstream yet (Strandberg & Grönlund, 2018), and there is great room for methodological innovations to map and measure deliberation in the online public sphere, an endeavor that I dedicate substantive parts of this dissertation to.

Despite the holistic and constructive perspective that a public sphere angle provides to research on online political communication, the perception of usefulness of the deliberation concept is not shared by everyone (Achen & Bartels, 2017).

Committed proponents and opponents of the concept of public sphere deliberation appear to be split along lines that go far beyond political communication research but may even reflect broader differences in *Menschenbild*. While some claim the death of deliberation in the digital era, I argue that a modern conceptualization of deliberation that moves beyond the classical Habermasian framework is a concept worth considering for both normative and pragmatic reasons.

One obvious factor that may drive differences in values attached to the concept of deliberation may be differences between political systems, in particular, differences between the US political system and, for example, European democracies. While also European democracies suffer from the rise of populists, multi-party systems that involve coalition formation require, by definition, more deliberation on the the level of political elites<sup>1</sup>. Considering the citizens level, while partisan identity signaling is highlighted as important driver of affective political

---

<sup>1</sup> However, it is argued that in principle, also the US constitutional system builds on a deliberative conception of democracy (Sunstein, 2018).

polarization in the US (Bail et al., 2018), party preferences appear more volatile in many European countries, which is reflected in the increasing dealignment between voters and established mainstream parties (Spoon & Klüver, 2019). Furthermore, implied by the popularity of voting advice applications (Munzert & Ramirez-Ruiz, 2021), the alignment of issue attitudes with concrete policy proposals does seem to matter for the vote choices of many Europeans.

However, there may also be a deeper reason, related to differences in the perception of citizens' agency and competences. An interesting parallel can be observed among behavioral science perspectives, a field that is closely linked to political science in the way that it also examines human behavior in response to certain information in certain contexts.

Undoubtedly, the work on heuristics and biases by Kahnemann and Tversky has had enormous influence on the understanding of human cognition and rationality that transcended into the social sciences and even the wider public. Centering around the concept of flawed cognition, they argue that “people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simple judgmental operations” (Tversky & Kahneman, 1974, p.1124).

Starting out from the homo oeconomicus which, obviously, does not accurately reflect human decision making, Kahnemann's notions of flawed decision making can help to make sense of mismatches between preferences and choices. In the political domain, such mismatches are omnipresent, for example, considering the mismatch between economic self-interest and trade preferences (Kertzer et al., 2022) or the gap between environmental attitudes and behavior (Kollmuss & Agyeman, 2002).

However, Kahnemann and Tversky's work is not the first attempt to understand human cognition and decision making and remains by far not the only conception. Already in 1967, Peterson and Beach concluded the much more



optimist metaphor of “the mind as an intuitive statistician” from a review of 160 experiments (Peterson & Beach, 1967). A recent meta study traced this fundamental shift in *Menschenbild*, back to a shift in experimental paradigms (Ralph Hertwig & Tomás Lejarraga, 2021). While the Bayesian reasoning literature relied on scenarios that involve learning and experience, the heuristics-and-biases program relied heavily on described scenarios and one-shot decisions, resulting in fundamental conclusions about every-day cognition. Certainly both types of scenarios find their application in real-life, but one might argue that the overwhelming dominance of the pessimist heuristics-and-biases perspective might not be justified given that learning is fundamental to human cognition.

A *Menschenbild* with a focus on cognitive flaws, incompetences and self-interest, obviously resonates a lot with criticisms of the deliberative ideal of public discourse, simply because the broad public seems incapable to deliberate. On the other hand, consistent empirical findings on transferable competences, cooperation, empathy and adaptive learning challenge the pessimist assumption of citizens’ incapacity and unwillingness to navigate conflicting arguments in political discussions.

While opponents argue that deliberative ideals run past real life political communication, proponents emphasize the need for theoretical anchors to situate online discussions in the larger democratic discourse.

Today, taking a less normative perspective, the public sphere is often simply understood as mediating arena between politics and civil society that includes the media system—infrastructures that enable the publication, distribution, consumption and discussion of information that allow people to exercise their rights and duties as citizens (Jungherr & Schroeder, 2021).

A reduction in conceptual complexity and a focus on empirical pragmatism certainly benefit the analysis of the status quo. While more complex conceptualizations of deliberation and the public sphere that involve heavy

normative premises risk reaching the limits of operationalizability, they may however help to develop future scenarios of a social media environment that serves democratic discourse.

The following sections summarize the different empirical contributions of this dissertation regarding different elements and dynamics of the online public discourse (for an overview, see Fig. 1; for a list of publications, see section B).

## **Political online ecosystems**

Political information consumption is a key democratic element if democracy is understood as ‘rule by the people’. Dahl (2020) defines the three-step process of (1) citizens developing attitudes, values and preferences, (2) citizens choosing the political party that best represents their preferences and concerns and (3) the elected government developing policies to address the preferences expressed by the electorate. Without citizens’ access to and engagement with political information, steps one and two are impossible.

Even though voters may prioritize efficiency in their decision strategies by using heuristics or consuming information selectively based on their interests (Lau & Redlawsk, 2001; Tóth & Chytilék, 2018), citizens’ attitudes and decisions are not just a product of competence, motivation and experience, but they are also influenced considerably by the information ecology around them (Boudreau & MacKenzie, 2014; Henderson et al., 2022; Singh & Roy, 2014).

When regarding the big question of whether the internet facilitates or harms public discourse, most of the previous empirical research on this question has focused on specific elements, such as online news media diets or political discussions on social media platforms. However, increasingly complex media environments are composed of different arenas with different respective potential for democracy.

Using web tracking data in combination with survey data, including full browser histories of  $N = 1,282$  citizens over a six-month period together with

detailed information on demographics and attitudes, I map the complex online political ecosystem in Germany (Oswald, 2022b, see full text of publication from p. 121).

I start by identifying which websites were relevant for the online public discourse in 2017 by matching political key words to the URL strings of the websites visited by participants. I then develop a multidimensional measure of ‘deliberative potential’ along six criteria, as indicator of the extent to which a website may fulfill different functions of a deliberative system.

The first three criteria regard the infrastructure of a website, namely whether the website provides political information, whether the site offers spaces for political discussion, and opportunities for political participation. Three further criteria capture whether a website is connected to other politically relevant websites through network measures of links and flows between websites, whether a website is visited by demographically diverse citizens and whether a website is visited by individuals with heterogeneous political views. Finally, using latent class analysis, I explore different types of websites to structure the political online ecosystem in Germany.

Besides a class of mainstream informational hubs, websites that are widely used across German speaking media audiences, I find a class of quality information providers, including the whole spectrum of online public broadcasting services in Germany. However, this class of high quality journalistic media is not characterized by a diverse audience. A third class of niche online forums hosts political discussions among more tightly-knit online communities, supporting previous findings of incidental exposure to political content online.

Furthermore, I find that while the mainstream information hubs attract a much larger number of website visits, this sample spends relatively more time consuming political information on quality information sites as well as engaging in political discussions within niche online forums. This divergence further highlights a key measurement point for conclusions drawn from web tracking data which may

differ substantially between website visit counts and visit duration.

## **Selection and representation in the online public discourse**

Moving the focus beyond the political online ecosystem (Fig. 1), the next important question for the assessment of the online public discourse is who constitutes the online public, the citizens engaging with political content online?

Public discourse is not only shaped by the settings and discussed issues but also considerably by the individuals involved (Bächtiger & Hangartner, 2010; Esterling, 2018). For example, Beauchamp (2020) describes the ‘deliberativeness’ of online discussions as a function of structures and membership. Online political deliberation, which can narrowly be defined as an ideal form of political discussion in which participants weight conflicting arguments based on high quality information, appears to be a rather rare phenomenon in relation to the overall volume of online communication. To circumvent issues of selection into public discourse, experimental deliberation research uses random sampling and/or random assignment for the benefit of internal validity. However, while experimental evidence can clarify fundamental principles of political discussions under ideal conditions, this evidence hardly generalizes to real-life online public discourse that underlies inherent selection processes.

Using web tracking data in combination with survey data again, in a paper coauthored with Simon Munzert, Pablo Barberá, Andrew Guess and JungHwan Yang, I map out who engages with political content across various arenas of the complex political online ecosystem in Germany (Oswald et al., 2022, see full text of publication from p. 170). We find that especially people with high political knowledge engage with political information online. Unlike previous research that largely focused on digital traces of *active* political engagement, such as comments written or posts shared on social media, we do not find evidence for strong selection along other demographic, attitudinal, and political variables but a large overlap

among the online public.

Selection into political engagement does not necessarily imply a threat to democratic discourse, especially not when occurring along variables like political knowledge or interest. Deliberative processes require motivation to consume relevant information, to process and communicate information, and to tolerate and engage with opposing views. In turn, a lack of motivation should also guard public discourse from those with the lowest knowledge of day-to-day political issues. However, from a broader societal perspective, increasing disengagement with politics can also be a threat to democracy as civil society may fail to hold their representatives accountable.

Following a first selection step along political knowledge, we find that people with rather extreme political views tend to engage more with platforms that enable political expression and discussion than people with more mainstream political views. This finding highlights differences between active and passive political internet use and has implications for the inferences drawn from social media data. While social media data usually captures the visible traces of active engagement in the online public discourse such as writing comments, ‘liking’ or sharing posts, our findings imply the existence of a second selection step into this behavior.

In other words, individuals select into passive engagement first, for example, in the form of political information consumption or discussion reading along their overall political interest or knowledge. Second, among those individuals, only a certain group of individuals, likely those with higher toxicity tolerance (Bor & Petersen, 2022; Kim et al., 2021) and those with more extreme political opinions (Barberá et al., 2015; Bright et al., 2019) select into active engagement in the online public discourse that is visible to others<sup>2</sup>. Therefore, inferences drawn from social

---

<sup>2</sup> A parallel discrepancy between active and passive engagement was also reflected in a recent study on information sharing across 43 countries, finding that those who shared links on Facebook, especially political information, were considerably older than those who saw the links (Moretto et al., 2022).

media data can only reflect the opinions and communication norms of this subgroup of individuals who represent a likely skewed tip of the iceberg.

One important lever for more democratic discourse online is the construction of inclusive online environments that are attractive for everyone who is motivated to contribute, usually implying a certain degree of political knowledge and/or interest. However, considering the secondary selection variables identified, this, for example, implies effective action against online toxicity.

### **Discussion quality and toxicity in the online public discourse**

Toxicity directed against groups can increase safety concerns (Henson et al., 2013) which prevents particularly minorities and traditionally marginalized individuals from participating in the online public discourse (Lenhart et al., 2016). One promising example is that displaying community rules in a Reddit forum slightly increased rule compliance but more importantly, it increased the participation rate of newcomers (Matias, 2019). The authors traced this back to a selection process in online discussions where people make frequent, low-cost decisions on whether to join a discussion or not. Displayed community rules may reduce people's concerns about harassment that would otherwise prevent them from participating (Matias, 2019).

The above described project (Oswald et al., 2022) hints towards a secondary selection step into active contribution to the online public discourse along variables such as political opinion extremity. This phenomenon, fostering extreme voices while keeping marginalized groups largely out of the discourse, aligns with findings on selection into toxic communication behavior (Bor & Petersen, 2022; Kim et al., 2021). Therefore, an assessment of the discussion quality and the development of a strategy to reduce online toxicity to make online environments more inclusive and public discourse more representative are the objectives of the following two projects.

## **Computational assessment of deliberative quality**

Moving beyond the mapping of the political online infrastructure and the individuals involved, a core question regarding the state of public discourse is the deliberative quality of communicative acts. In theory, high-quality discussions can help people acquire an adequate understanding of issues and alleviate mechanisms of opinion polarization (Grönlund et al., 2015; Mutz, 2006). However, the quality of online political discussions in the digital sphere is mourned by researchers and the public (Coe et al., 2014; Frimer et al., 2022; Hmielowski et al., 2014). The state of the online public discourse has given reason for concern in recent years, including phenomena of affective polarization, discussion toxicity and online hate speech (Lewandowsky et al., 2020; Lorenz-Spreen, Oswald, et al., 2023). The prevalence and severity of toxicity in online discussions varies between platforms, contexts and measurement approaches (Coe et al., 2014) but there is large consensus that toxicity, the inverse of the deliberative quality criterion of respect, is an important concept to research in online environments (Chen, 2017; Coe et al., 2014; Kim et al., 2021; Mutz, 2016; Rossini, 2019).

Most prior deliberation research required enormous amounts of time and human effort to assess the deliberative quality of discussions with in-depth manual content coding (Esau et al., 2017; Friess et al., 2020; Steenbergen et al., 2003). Facing the importance and the sheer volume of online discussions, reliable computational approaches to assess the deliberative quality of online discussions at scale could open a new era of deliberation research. Furthermore, regulators as well as platforms who might decide to optimize online environments for deliberation instead of pure engagement, cannot rely on manual content coding. Therefore, this project focuses on the question of whether simple computational measures can contribute to the assessment of deliberative quality at scale (Oswald, 2022a, see full text of publication from p. 199).

More specifically, I compared structural features of discussion threads and

simple text-based measures to established manual content analysis. I applied all measures to online discussions on the social news aggregation and content rating platform ‘Reddit’ that dealt with the 2020 wildfires in Australia and California (Oswald, 2022a). I further compared discussions between two ideologically opposite online communities, one featuring discussions in line with the scientific consensus on climate change and one featuring climate change skepticism.

Obviously, no single computational measure could capture the multidimensional concept of deliberative quality. However, I find that measures of structural complexity, namely the width and depth of comment cascades (Gonzalez-Bailon et al., 2010), capture engagement and participation as important preconditions for deliberation. Furthermore, automated toxicity scores (Votta, 2019) predict manual measures of respect, and the length of comments in combination with the absence of toxicity predicts manual measures of argumentation. While the presented computational approaches could not replace in-depth content coding, the findings imply that selected automated measures can be useful scalable additions to the measurement repertoire for two specific dimensions of online deliberation, argumentation and toxicity.

From the perspective of regulators or ‘socially benevolent’ social media platforms who may eventually prioritize democratic discourse over revenues from user attention, the assessment of dimensions such as the diversity of users participating in discussions, the number of replies a discussion cascades through, the question whether comments go beyond very short exclamations or pure emoji, as well as the level of toxicity, could be a valuable and scalable starting point to shape online environments to better serve democratic discourse.

### **Testing preemptive empathy interventions to reduce toxicity**

In response to findings on a second selection step into active, visible engagement in the online public discourse (Bor & Petersen, 2022; Kim et al., 2021;



Oswald et al., 2022), the objective of the final project of this dissertation was to develop a preemptive strategy to reduce online toxicity before it may silence important voices and trigger subsequent toxicity (Oswald, 2023, see full text of publication from p. 233). To that end, I conducted a preregistered survey experiment that tested multiple behavioral interventions to reduce the toxicity of online communication, three simple prompts or ‘nudges’ (Thaler & Sunstein, 2009) in the user interface and an educative ‘boosting’ intervention targeting users’ motivation (Hertwig & Grüne-Yanoff, 2017).

Following the theoretical turn towards the role of empathy for democratic deliberation (M. E. Morrell, 2010), this study focused on affective empathy and cognitive perspective taking (Stietz et al., 2019). Already Hannah Arendt argued that perspective-taking is essential for democratic opinion formation (Arendt, 1968) while theorists of deliberative democracy further highlight the relevance of a reflective citizenry, able to navigate conflicting arguments in political discussions (Bächtiger et al., 2018; Chambers, 2003; Dryzek, 2002; Muradova, 2020).

While social media platforms are often, and often accurately, portrayed as adversaries of deliberative democracy and gatekeepers of data in an attention economy, researchers can use social media as tools to gather important target groups for online interventions, individuals that are usually difficult to draw into university labs as participants for experiments.

In this project, I collected data using Facebook Ads, headed by the statement “We value your opinion!” with the aim to attract those who are active on the platform and happy to express their political opinions online. The final sample consisted of  $N = 2,154$  highly active social media users from the US and Germany, with high political interest but from both political camps, who opted into the survey without external incentives.

First, participants’ attitudes towards different salient and conflict-prone political issues were assessed. Participants were then presented with seed statements

that were designed like typical social media posts. The topics of the seed statements ranged from climate change, over migration, to the feminist issues of gender-neutral language in Germany and abortion in the US. By deliberately allocating participants to statements that oppose their opinion, the aim was to maximize disagreement between the seed statement and the opinion of the participant.

Before participants commented on the statements, one group was prompted (‘nudged’, Thaler and Sunstein (2009) to take the perspective of the author and to think about possible reasons why the author thinks this way (cognitive perspective taking) and one group was prompted to try to create a sense of empathy with the author and to reflect on how the author might feel (affective empathy). The first control group did not receive a prompt, whereas the second control group received a friction placebo, a prompt to take a deep breath before answering. Prior to reading the provocative posts as a form of educative ‘boosting’ intervention (Hertwig & Grüne-Yanoff, 2017), a fifth group was briefed on the role of empathy and perspective taking for constructive discussions and the distinction between legitimate opinion and harmful content online.

All participants were then asked to comment on the controversial statements. Those free text answers were analyzed using the established toxicity classifier (Votta, 2019) that was tested in the previous project (Oswald, 2022a). Furthermore, as a second outcome, also the length of comments was considered as approximate measure for argumentation, following the findings presented in Oswald (2022a).

Compared to the control group, neither the empathy and perspective taking nudges nor the friction placebo reduced reply toxicity. The educative boosting intervention decreased reply toxicity to some degree but the effect was not robust against the inclusion of covariates. However, all nudging interventions significantly increased the length of replies. Aligning with previous findings on selection into commenting online (Bor & Petersen, 2022; Kim et al., 2021), toxicity was highest among people with high online activity.

The goal of this study was to construct interventions and to test them on the eventual target group. The envisioned pipeline for regulators, (1) to monitor the nature of political discussions based on the findings in Oswald (2022a) and (2) to implement selected interventions tested in Oswald (2023) to mitigate destructive escalation on controversial issues, however appears more complex. Reducing toxicity online did not seem to work via simple changes in the user interface, as I find only weak effects of the experimental treatments, especially the prompt-based ‘nudging’ interventions.

However, this study does not allow concluding judgements on the conceptual role of empathy for democracy. Despite efforts to increase external validity through recruitment via Facebook Ads and the construction of organically appearing social media statements, the study still operates with a hypothetical communication scenario built into a survey experiment and effects highly depend on the design of the treatments.

The results for the boosting condition give reason for hope in democratic education that emphasizes the role of empathy and perspective taking for public discourse. This study only provided a very short and abstract text, mainly targeting participants’ motivation to engage in empathy but did not address the development of the cognitive-emotional competences necessary for effective perspective taking.

While incivility can even be a welcome stimulant to public debate in some specific cases (Masullo Chen et al., 2019; Mutz, 2016), I regard the development of interventions to reduce toxicity in online discussions as important endeavor. Therefore, more complex interventions conveying an understanding of empathy and practicing perspective taking in political discussions that involve conflicting arguments—which are essentially the core of democratic deliberation, should be explored by future research.

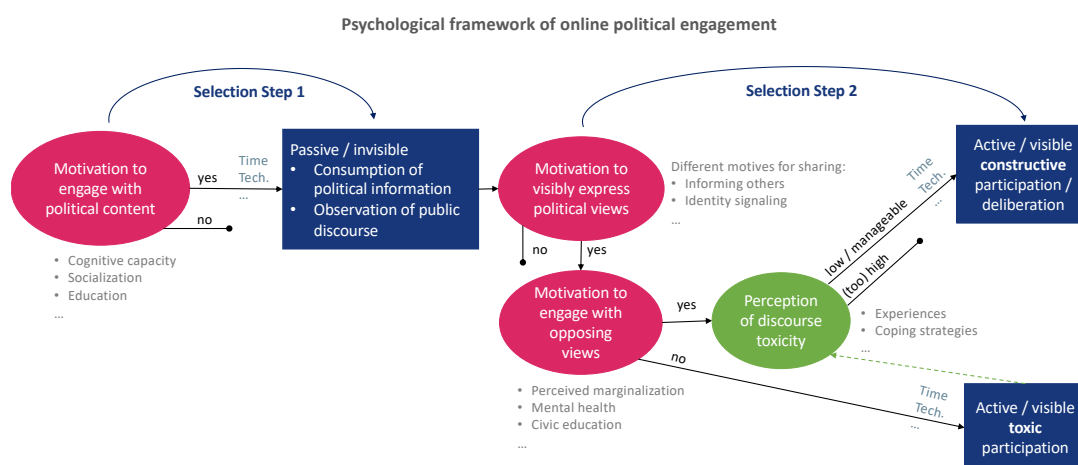
## Outlook

Considering the findings on the two-staged selection process presented in this dissertation, an open question remains. When do factors like political interest translate into deliberation and when into toxic behavior online? This relates back to the fundamental debate on the nature of deliberation as either elitist, a notion stated by pessimists regarding citizens' incapacity to deliberate, or egalitarian, a notion stated by optimists who see potential for constructive public discourse online. What are the underlying factors bridging the fundamentally different notions of political online engagement? In other words, what are the factors driving the divides between constructive participation, toxic participation and non-participation?

Throughout this dissertation, interindividual differences across different types of preconditions for deliberation became visible. The specifics of these preconditions differ substantially between passive and active political engagement online but the path seems to split again for different qualities of active engagement: toxic versus constructive political engagement.

An empirical examination of the connection between political resources and different forms of political engagement is one way to address the question. For example, Rasmussen et al. (2022) found evidence for the paradox that classical resources such as education and internal political efficacy can translate into toxic behavior online.

In a conceptual attempt to disentangle possible mechanisms underlying this paradox and to address the open question above, I suggest a psychological framework that includes various types of *motivation*, each involving different types of resources (see Fig. 2).



**Figure 2**

*Suggested psychological framework of online political engagement. Motivation to engage with politics as precondition for the consumption of political content online. This motivation can relate to general cognitive capacity, socialization and education. Motivation to visibly express political views as precondition for active behavior. Additional motivation to engage with opposing views as necessary for constructive participation and deliberation. A lack of this openness or capacity to engage with opposing views may translate into toxic behavior online. If discourse toxicity is perceived as low or manageable, motivation to express political views combined with the willingness to tolerate opposing views may translate into constructive political participation or deliberation online. The toxic behavior of those lacking the secondary motivation feeds back into discourse toxicity and interrupts the path to participation for some who otherwise fulfill all preconditions for deliberation. Resources such as time and technological infrastructure are necessary at all stages to let motivation translate into behavior.*

Within the first selection step into passive political engagement online, motivation to engage with politics is the necessary precondition for the consumption of political content. This motivation could, for example, be traced back to general cognitive capacity, socialization and education. To translate into, at that stage, for other citizens invisible behavior, further resources such as time and technical infrastructure are necessary. Empirically, the first and rather intuitive selection step could, for example, be reflected by high political knowledge and interest of those consuming political information online (Oswald, 2022b).

Within the second selection step, the apparent paradox becomes visible as empirically, often those who are most active online are also those who engage in toxic behavior. Motivation to visibly express political views is the obvious necessary

condition for active behavior. Here, individuals can have different motives for sharing political content online, such as the desire to inform, entertain or connect with others, but also to surprise and provoke (Lorenz-Spreen, Kozyreva, et al., 2023).

However, I argue that a second type of motivation is necessary for constructive participation which may eventually translate into online deliberation. This specific motivation as potential explanatory factor for conflicting findings on the quality of political online engagement has previously been conceptualized in many ways: as motivation to engage with opposing views (Fig. 2), as willingness to hear the other side or political tolerance (Mutz, 2006), as receptiveness (Minson & Dorison, 2022), as deliberative listening (M. Morrell, 2018), as ‘democratic enlightenment’ (Nie et al., 1996) or simply as the opposite of affective polarization (Rasmussen et al., 2022).

A lack of this openness or capacity to engage with opposing views may originate from various factors ranging from economic deprivation and perceptions of marginalization to overall mental health and can, given time and technical infrastructure, eventually translate into toxic behavior online. Supportive factors for the development of motivation to engage with opposing views could, for example, be a combination of civic education and digital literacy (Guess & Munger, 2022; Kahne et al., 2012). Under the premise that an individual is motivated to express political views *and* accepts fundamental democratic principles such as basic tolerance for opposing views, this may translate into active constructive political participation and eventually deliberation online.

However, in line with previous findings on online toxicity (Henson et al., 2013; Lenhart et al., 2016), the model includes another filtering step, namely the premise that discourse toxicity is perceived as low or manageable in the face of, for example, prior experiences and available coping strategies. In turn, the toxic behavior of those unwilling or unable to tolerate opposing views feeds back into

discourse toxicity and consequentially, may push some of those who otherwise fulfill all preconditions for deliberation out of public discourse online.

The suggested framework not only separates different motivations as preconditions for two selection steps into political engagement online, it also distinguishes between different resources which may otherwise blur the picture. There are classical cognitive and educational resources in the first selection step that remain relevant until behavior becomes visible, constructive or toxic in nature (Rasmussen et al., 2022). However, other additional resources are necessary for constructive public discourse which may, I hypothesize, be factors like socio-economic security, mental well-being, civic education, and an internalization of democratic norms that open up peoples' minds for opposing views. Very pragmatic resources such as time, (digital) literacy and technological infrastructure are obviously necessary at all three stages to let motivation translate into behavior.

While the model builds upon existing findings and prior theoretical notions, the suggested structure of underlying psychological mechanisms remains to be tested empirically.

## **Consequences for representation**

Considering political expression online, previous research has shown that a large majority of activity online originates from a small but highly active minority of users that does not represent society at large (Barberá et al., 2015; Bright et al., 2019; Guess, 2021). Therefore, I agree that “online democratic mechanisms need to take steps to promote the views of the majority of users who are less active, or at least limit the impact of the highly active.” Bright et al. (2019, p. 20).

For example, irrespective of political views, when considering URL tracking data, there is a consistent pattern that compared to men, women tend to engage less with political content online (Oswald et al., 2022). This discrepancy might transcend from structural inequalities and path dependencies regarding overall

political participation but also from inhibited motivation to visibly participate in online environments due to expectations of being harassed (see Fig. 2).

In addition, the spiral of silence theory states that people's willingness to express their political opinions online depends on the perceived popularity of their opinions (Matthes et al., 2018; Scheufle & Moy, 2000), implying that toxic experiences after crossing the threshold of visibly expressing political views online likely impacts future motivation.

In contrast to public content moderation preferences (Kozyreva et al., 2023), current developments of weaponized free speech ideals that are propagated by influential actors (Duffield, 2022) strengthen platforms for the voices of traditionally privileged individuals with extreme viewpoints and toxic behavior. This may pose non-linear threats to democratic discourse online through the creation of negative feedback loops.

In other words, toxicity is not randomly distributed online but silences relevant voices in the discourse by predominantly forcing already marginalized groups offline. Therefore, effective responses to online toxicity, with intolerable hate speech at the end of the continuum, are no nice-to-have online well-being boosters but necessary conditions for the future of a healthy online public discourse.

However, online deliberation is more than the absence of hate speech. Eventually, the construction of online environments that are functional, attractive and inclusive for all those motivated to engage in public discourse, combined with a proactive approach to rebalance existing inequalities that suppressed structurally underrepresented but in principle motivated and well informed voices, would be one way to imagine a positive scenario for the role of digital media in democratic societies.

Facing measures of online toxicity, it becomes clear that the normative ideals presented in deliberative theory may not offer an accurate description of the present. However, the established theoretical concept of the public sphere offers a



constructive perspective of what aspects empirical researchers might focus on to provide regulators with relevant evidence to inform policies to reshape the link between digital media and democracy.

## **Potentials of computational social science**

Major hurdles in the endeavor to better understand the state and role of the online public discourse in today's society are issues of measurement and empirical investigation. The emergence of the internet and social media, both glorified, then demonized in the face of unfulfilled hopes for widespread democratization, in principle, provides researchers with unique opportunities to observe individual political engagement in the wild.

The field of computational social science (CSS) bears great potential for the study of political communication (Theocharis & Jungherr, 2021). Social interactions can be monitored without inducing observer effects, artificial lab settings can be circumvented to the benefit of external validity, and the seemingly unlimited amount of data opens opportunities for research groups with limited financial resources (Salganik, 2019). Assuming access to data, digital traces provide social scientists with unprecedented information on communication, social networks and information consumption - optimal conditions for the study of public discourse and political opinion formation.

One challenge identified for the field of CSS is the connection to established theories in the social sciences (Jungherr & Theocharis, 2017), a challenge that I try to address in this dissertation by embedding both experimental and descriptive work in democratic theory.

Furthermore, while truly interdisciplinary work appears most promising for CSS, research groups (especially outside the US), institutional infrastructures and practices still predominantly follow disciplinary boundaries (Theocharis & Jungherr, 2021). In this dissertation, I link the fields of political science and psychology using

computational social science methods but I would warmly welcome greater institutional opening to true interdisciplinarity.

However, at the time of writing this dissertation, the major obstacle to meaningful CSS research appears to become data access. With Elon Musk's takeover of Twitter and his decision to eliminate the academic Twitter API<sup>3</sup> and replace it with a restricted commercial model (Meehan, 2023), an entire research field has been experiencing a major turning point. While some may have argued that Twitter was overresearched in the first place, the Twitter API was not only a resource of free communication data for teaching and experimenting, it was the major data provider for countless research projects across the world, many of which undergo major disruptions at the moment.

While other platforms are announcing the opening of APIs, their terms of service appear largely incompatible with transparent, open science research (Bak-Coleman, 2023). These developments allude to the consequential question of whether we are observing the start of a downward trend for data availability from social media platforms. The protection of researchers from litigation from digital media companies, granting specialized access to non-public data for research projects of crucial societal importance and the mediation of the relationship between platforms and researchers regarding data access by a third party are only some suggested measures that may have the potential alleviate some issues (Brown, 2023).

Research requires resources, and those are not evenly distributed across countries, institutions and career stages. Within the empirical projects of my dissertation, I relied on broadly three different data sources: (1) communication data from Reddit accessed through free and unrestricted APIs, (2) secondary web-tracking data with linked survey data from an existing research project, and (3) original survey experiment data with participants recruited via Facebook Ads. These data sources differ substantially regarding their accessibility which may have

---

<sup>3</sup> Application Programming Interface

led to substantially different research output given a different set of resources.

Online information environments are constantly moving targets that underlie the decisions of powerful actors that are not democratically legitimized (Conger et al., 2022). This dynamic needs regulation and regulation needs evidence. Scientists must keep up with the pace of digital media developments, for example, to better understand the effects of digital media on democracy, so that platforms cannot use empirical darkness as argument for unregulated development (Clegg, 2021).

In this power struggle, it may appear ironic to base research on data provided by key actors that are, with their design decisions, fundamentally responsible for current issues within the online public discourse. However, in this race for evidence, collaborations between researchers and digital media companies that are based on well negotiated, transparent research agreements are one option to move forward.

Finally, the Digital Services Act (DSA) provides researchers with the unique opportunity to access formerly restricted data without having to rely on the good will of platforms. Therefore, if researchers use their levers as provided by the DSA effectively, I am optimistic that the field can use its potential to answer relevant research questions regarding the online public discourse if working collaboratively on high-quality description, large-scale field experiments and with fruitful connections to social science theory.

In this dissertation, I provide various examples of how combinations of different data sources can be used to map complex theoretical concepts, how online data can contribute to innovations in measurement, and how social media platforms can be used as research tools to gather hard to reach target groups for online experiments.

Computational social science commands many tools to map various elements of the online public discourse from a holistic perspective, which may eventually reconnect public sphere theory with real life online.

**Table 1***Potentials of online data to operationalize elements of the online public discourse*

	POLITICAL ONLINE ECOSYSTEM		ONLINE PUBLIC DISCOURSE		ONLINE PUBLIC
Concepts	Infrastructure, Political information	Political issues	Discussion of political issues	Connections between issues	Politically (dis)engaged citizens Power users
Measures	Information providers Discussion forums Petitioning platforms Political information on social media	Political topics Framing	Deliberative quality Toxicity Argumentation	Temporal flow of discussion Semantic links between issues	Demographics Attitudes Knowledge and competencies Motivation Social network
Potential data	<b>Web tracking data</b> Social media data	News articles Social media posts	<b>Social media posts</b> <b>Comments</b> <b>Online experiments</b>	<b>Network data</b> Time series data	<b>Survey responses</b> Social media data

*Note:* Non exhaustive list of concepts, measures and potential data sources. Boldface indicates use in the course of this dissertation.

## References

- Achen, C. H., & Bartels, L. M. (2017). *Democracy for realists: Why elections do not produce responsive government* (Vol. 4). Princeton University Press.
- Alizada, N., Boese, V. A., Lundstedt, M., Morrison, K., Natsika, N., Sato, Y., Tai, H., & Lindberg, S. I. (2022). *Autocratization changing nature?* (Democracy Report). Varieties of Democracy Institute (V-Dem). University of Gothenburg.
- Arendt, H. (1968). *Between Past and Future*. Penguin Classics.
- Bächtiger, A., Dryzek, J. S., Mansbridge, J., & Warren, M. E. (2018). *The Oxford handbook of deliberative democracy*. Oxford University Press.
- Bächtiger, A., & Hangartner, D. (2010). When deliberative theory meets empirical political science: Theoretical and methodological challenges in political deliberation. *Political Studies*, *58*(4), 609–629.
- Bächtiger, A., Niemeyer, S., Neblo, M., Steenbergen, M. R., & Steiner, J. (2010). Disentangling diversity in deliberative democracy: Competing theories, their blind spots and complementarities.
- Bächtiger, A., & Parkinson, J. (2019). *Mapping and measuring deliberation: Towards a new deliberative quality*. Oxford University Press.
- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221.
- Bak-Coleman, J. (2023). TikTok’s API Guidelines Are a Minefield for Researchers. Retrieved March 8, 2023, from <https://techpolicy.press/tiktoks-api-guidelines-are-a-minefield-for-researchers/>
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science*, *26*(10), 1531–1542.

- Beauchamp, N. (2020). Modeling and Measuring Deliberation Online. In *The Oxford Handbook of Networked Communication*.
- Bor, A., & Petersen, M. B. (2022). The Psychology of Online Political Hostility: A Comprehensive, Cross-National Test of the Mismatch Hypothesis [Publisher: Cambridge University Press]. *American Political Science Review*, 116(1), 1–18. <https://doi.org/10.1017/S0003055421000885>
- Boudreau, C., & MacKenzie, S. A. (2014). Informing the electorate? How party cues and policy information affect public opinion about initiatives [Publisher: Wiley Online Library]. *American Journal of Political Science*, 58(1), 48–62.
- Bright, J., Bermudez, S., Pilet, J.-B., & Soubiran, T. (2019). Power users in online democracy: Their origins and impact. *Information, communication & society*, 1–16.
- Brown, M. A. (2023). The Problem with TikTok’s New Researcher API is Not TikTok. Retrieved March 8, 2023, from <https://techpolicy.press/the-problem-with-tiktoks-new-researcher-api-is-not-tiktok/>
- Chambers, S. (1996). *Reasonable Democracy: Listening, Conflict, and Citizenship: Jürgen Habermas and the Politics of Discourse/Simone Chambers*. Ithaca: Cornell University Press.
- Chambers, S. (2003). Deliberative democratic theory [Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA]. *Annual review of political science*, 6(1), 307–326.
- Chen, G. M. (2017). *Online Incivility and Public Debate*. Springer. Retrieved November 22, 2022, from <https://link.springer.com/book/10.1007/978-3-319-56273-5>
- Clegg, N. (2021). What the Wall Street Journal Got Wrong. Retrieved December 23, 2022, from <https://about.fb.com/news/2021/09/what-the-wall-street-journal-got-wrong/>

- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658–679.
- Cohen, J. (1989). *Deliberation and Democratic Legitimacy* 'in *The Good Polity: Normative Analysis of the State*, eds. Alan Hamlin and Philip Pettit. Oxford: Basil Blackwell.
- Conger, K., Isaac, M., Mac, R., & Hsu, T. (2022). Two Weeks of Chaos: Inside Elon Musk's Takeover of Twitter. *The New York Times*. Retrieved February 3, 2023, from <https://www.nytimes.com/2022/11/11/technology/elon-musk-twitter-takeover.html>
- Dahl, R. A. (2020). *On democracy*. Yale university press.
- Dryzek, J. S. (1987). Discursive designs: Critical theory and political institutions. *American Journal of Political Science*, 656–679.
- Dryzek, J. S. (2002). *Deliberative democracy and beyond: Liberals, critics, contestations*. Oxford University Press on Demand.
- Dryzek, J. S., Bächtiger, A., Chambers, S., Cohen, J., Druckman, J. N., Felicetti, A., Fishkin, J. S., Farrell, D. M., Fung, A., & Gutmann, A. (2019). The crisis of democracy and the science of deliberation [Publisher: American Association for the Advancement of Science]. *Science*, *363*(6432), 1144–1146.
- Duffield, W. (2022). What Can Elon Musk Do with Twitter? [Publisher: Cato Institute]. Retrieved August 14, 2022, from <https://policycommons.net/artifacts/2390234/what-can-elon-musk-do-with-twitter/3411430/>
- Esau, K. (2022). Decluttering the Black Box: A Critical Review and Four-Phase Model of Deliberation Research.
- Esau, K., Friess, D., & Eilders, C. (2017). Design Matters! An Empirical Analysis of Online Deliberation on Different News Platforms. *Policy & Internet*, *9*(3), 321–342. <https://doi.org/https://doi.org/10.1002/poi3.154>

- Esterling, K. (2018). Deliberation and Experimental Design. In *The Oxford Handbook of Deliberative Democracy*. Retrieved October 22, 2021, from <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780198747369.001.0001/oxfordhb-9780198747369-e-13>
- Fishkin, J. S. (1991). *Democracy and deliberation: New directions for democratic reform* (Vol. 217). Yale University Press New Haven, CT.
- Friess, D., Ziegele, M., & Heinbach, D. (2020). Collective Civic Moderation for Deliberation? Exploring the Links between Citizens' Organized Engagement in Comment Sections and the Deliberative Quality of Online Discussions. *Political Communication*, *0*(0), 1–23.  
<https://doi.org/10.1080/10584609.2020.1830322>
- Frimer, J., Aujla, H., Feinberg, M., Skitka, L., Aquino, K., & Willer, R. (2022). Incivility is rising among American politicians on Twitter.
- Galston, W. A. (2001). Political knowledge, political engagement, and civic education [Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA]. *Annual review of political science*, *4*(1), 217–234.
- Gonzalez-Bailon, S., Kaltenbrunner, A., & Banchs, R. E. (2010). The structure of political discussion networks: A model for the analysis of online deliberation. *Journal of Information Technology*, *25*(2), 230–243.
- Graham, T. (2012). Beyond “political” communicative spaces: Talking politics on the Wife Swap discussion forum. *Journal of Information Technology & Politics*, *9*(1), 31–45.
- Grönlund, K., Herne, K., & Setälä, M. (2015). Does enclave deliberation polarize opinions? *Political Behavior*, *37*(4), 995–1020.
- Guess, A. M. (2021). (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets. *American Journal of Political Science*.



- Guess, A. M., & Munger, K. (2022). Digital literacy and online political behavior. *Political Science Research and Methods*, 1–19.  
<https://doi.org/10.1017/psrm.2022.17>
- Habermas, J. (2022). Reflections and Hypotheses on a Further Structural Transformation of the Political Public Sphere [Publisher: SAGE Publications Ltd]. *Theory, Culture & Society*, 39(4), 145–171.  
<https://doi.org/10.1177/02632764221112341>
- Habermas, J. (1995). Reconciliation through the public use of reason: Remarks on John Rawls's political liberalism. *The journal of philosophy*, 92(3), 109–131.
- Habermas, J., Lennox, S., & Lennox, F. (1974). The public sphere: An encyclopedia article (1964). *New German Critique*, (3), 49–55.
- Haidt, J. (2022). Why the Past 10 Years of American Life Have Been Uniquely Stupid. Retrieved April 25, 2022, from  
<https://www.theatlantic.com/magazine/archive/2022/05/social-media-democracy-trust-babel/629369/>
- Henderson, J. A., Sheagley, G., Goggin, S. N., Dancey, L., & Theodoridis, A. G. (2022). Primary Divisions: How Voters Evaluate Policy and Group Differences in Intraparty Contests [Publisher: The University of Chicago Press]. *The Journal of Politics*, 84(3), 1760–1776.  
<https://doi.org/10.1086/718208>
- Henson, B., Reynolds, B. W., & Fisher, B. S. (2013). Fear of crime online? Examining the effect of risk, previous victimization, and exposure on fear of online interpersonal victimization [Publisher: Sage Publications Sage CA: Los Angeles, CA]. *Journal of Contemporary Criminal Justice*, 29(4), 475–497.
- Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, 12(6), 973–986.

- Hmielowski, J. D., Hutchens, M. J., & Cicchirillo, V. J. (2014). Living in an age of online incivility: Examining the conditional indirect effects of online discussion on political flaming. *Information, Communication & Society*, *17*(10), 1196–1211.
- Jungherr, A., & Schroeder, R. (2021). Digital Transformations of the Public Arena [ISBN: 9781009064484 9781009065542 Publisher: Cambridge University Press]. *Elements in Politics and Communication*.  
<https://doi.org/10.1017/9781009064484>
- Jungherr, A., & Theocharis, Y. (2017). The empiricist's challenge: Asking meaningful questions in political science in the age of big data [Publisher: Routledge \_eprint: <https://doi.org/10.1080/19331681.2017.1312187>]. *Journal of Information Technology & Politics*, *14*(2), 97–109.  
<https://doi.org/10.1080/19331681.2017.1312187>
- Kahne, J., Lee, N.-J., & Feezell, J. T. (2012). Digital media literacy education and online civic and political participation. *International Journal of Communication*, *6*, 24.
- Kertzer, J. D., Barbera, P., Guess, A., Munzert, S., Yang, J., & Zhou, A. (2022). Trade Attitudes In The Wild, 51.
- Kim, J. W., Guess, A., Nyhan, B., & Reifler, J. (2021). The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity [Publisher: Oxford University Press]. *Journal of Communication*, *71*(6), 922–946.
- Kollmuss, A., & Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental education research*, *8*(3), 239–260.
- Kozyreva, A., Herzog, S. M., Lewandowsky, S., Hertwig, R., Lorenz-Spreen, P., Leiser, M., & Reifler, J. (2023). Resolving content moderation dilemmas between free speech and harmful misinformation [Publisher: Proceedings of

- the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, 120(7), e2210666120. <https://doi.org/10.1073/pnas.2210666120>
- Lau, R. R., & Redlawsk, D. P. (2001). Advantages and disadvantages of cognitive heuristics in political decision making [Publisher: JSTOR]. *American journal of political science*, 951–971.
- Lenhart, A., Ybarra, M., Zickuhr, K., & Price-Feeney, M. (2016). Online harassment, digital abuse, and cyberstalking in America. Data & Society Research Institute. *Center for Innovative Public Health Research*. Retrieved from: [https://datasociety.net/pubs/oh/Online\\_Harassment\\_2016.pdf](https://datasociety.net/pubs/oh/Online_Harassment_2016.pdf).
- Lewandowsky, S., Smillie, L., Garcia, D., Hertwig, R., Weatherall, J., Egidy, S., Robertson, R. E., O'Connor, C., Kozyreva, A., Lorenz-Spreen, P., et al. (2020). Technology and Democracy: Understanding the influence of online technologies on political behaviour and decision-making.
- Lorenz-Spreen, P., Kozyreva, A., Swire-Thompson, B., Geers, M., & Herzog, S. M. (2023). The various motives for sharing content on Twitter, a field study in real time.
- Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2021). Digital Media and Democracy: A Systematic Review of Causal and Correlational Evidence Worldwide.
- Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2023). A systematic review of worldwide causal and correlational evidence on digital media and democracy [Number: 1 Publisher: Nature Publishing Group]. *Nature Human Behaviour*, 7(1), 74–101. <https://doi.org/10.1038/s41562-022-01460-1>
- Masullo Chen, G., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We Should Not Get Rid of Incivility Online [Publisher: SAGE Publications Ltd]. *Social Media + Society*, 5(3), 2056305119862641. <https://doi.org/10.1177/2056305119862641>

- Matias, J. N. (2019). Preventing harassment and increasing group participation through social norms in 2,190 online science discussions [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *116*(20), 9785–9789.  
<https://doi.org/10.1073/pnas.1813486116>
- Matthes, J., Knoll, J., & von Sikorski, C. (2018). The “spiral of silence” revisited: A meta-analysis on the relationship between perceptions of opinion support and political opinion expression [Publisher: Sage Publications Sage CA: Los Angeles, CA]. *Communication Research*, *45*(1), 3–33.
- Meehan, B. (2023). Letter: Imposing Fees to Access the Twitter API Threatens Public-Interest Research. Retrieved March 8, 2023, from <https://independenttechresearch.org/letter-twitter-api-access-threatens-public-interest-research/>
- Messner, M., & Distaso, M. W. (2008). The source cycle: How traditional media and weblogs use each other as sources. *Journalism studies*, *9*(3), 447–463.
- Minson, J. A., & Dorison, C. A. (2022). Why is exposure to opposing views aversive? Reconciling three theoretical perspectives. *Current Opinion in Psychology*, 101435. <https://doi.org/10.1016/j.copsyc.2022.101435>
- Moretto, M., Ortellado, P., Kessler, G., Vommaro, G., Rodriguez-Raga, J. C., Luna, J. P., Heinen, E., Cely, L. F., & Toro, S. (2022). People are more engaged on Facebook as they get older, especially in politics: Evidence from users in 46 countries. *Journal of Quantitative Description: Digital Media*, *2*.  
<https://doi.org/10.51685/jqd.2022.018>
- Morrell, M. (2018). Listening and deliberation. *The Oxford Handbook of deliberative democracy*, 237–250.
- Morrell, M. E. (2010). *Empathy and democracy: Feeling, thinking, and deliberation*. Penn State Press.

- Müller, K., & Schwarz, C. (2018). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association*.
- Munzert, S., & Ramirez-Ruiz, S. (2021). Meta-Analysis of the effects of voting advice applications. *Political Communication*, 1–16.
- Muradova, L. (2020). Seeing the other Side? Perspective-taking and reflective political judgements in interpersonal deliberation. *Political Studies*, 0032321720916605.
- Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press.
- Mutz, D. C. (2016). *In-your-face politics: The consequences of uncivil media*. Princeton University Press. Retrieved November 22, 2022, from <https://press.princeton.edu/books/hardcover/9780691165110/in-your-face-politics>
- Nenoff, A. (2020). #MeToo: A Look at the Influence and Limits of 'Hashtag Activism' to Effectuate Legal Change. *University of Illinois Law Review*, 2020, 1327. <https://heinonline.org/HOL/Page?handle=hein.journals/unillr2020&id=1349&div=&collection=>
- Neuberger, C., Nuernbergk, C., & Langenohl, S. (2019). Journalism as multichannel communication: A newsroom survey on the multiple uses of social media. *Journalism Studies*, 20(9), 1260–1280.
- Newman, N., Fletcher, R., Kalogeropoulos, A., & Nielsen, R. (2019). *Reuters Institute digital news report 2019* (Vol. 2019). Reuters Institute for the Study of Journalism.
- Nie, N. H., Junn, J., & Stehlik-Barry, K. (1996). *Education and democratic citizenship in America*. University of Chicago Press.
- Oswald, L. (2022a). Automating the Analysis of Online Deliberation? A comparison of manual and computational measures applied to climate change discussions. <https://doi.org/10.31235/osf.io/qmzwx>

- Oswald, L. (2022b). More than News! Mapping the Deliberative Potential of a Political Online Ecosystem with Digital Trace Data.  
<https://doi.org/10.31235/osf.io/68udg>
- Oswald, L. (2023). Effects of Preemptive Empathy Interventions on Reply Toxicity among Highly Active Social Media Users.  
<https://doi.org/10.31235/osf.io/evdxy>
- Oswald, L., Munzert, S., Barbera, P., Guess, A., & Yang, J. (2022). Beyond the tip of the iceberg? Exploring Characteristics of the Online Public with Digital Trace Data. <https://doi.org/10.31235/osf.io/yfmzh>
- Persson, M. (2015). Education and political participation [Publisher: Cambridge University Press]. *British Journal of Political Science*, 45(3), 689–703.
- Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. [Publisher: American Psychological Association]. *Psychological bulletin*, 68(1), 29.
- Ralph Hertwig & Tomás Lejarraga. (2021). How Experimental Methods Shaped Views on Human Competence and Rationality. *Psychological Bulletin*.
- Rasmussen, S. H. R., Osmundsen, M., & Petersen, M. B. (2022). Political Resources and Online Political Hostility How and Why Hostility Is More Prevalent Among the Resourceful. <https://doi.org/10.31234/osf.io/tp93r>
- Rawls, J. (1995). Political liberalism: Reply to Habermas. *The journal of philosophy*, 92(3), 132–180.
- Rosenberg, S. W. (2014). Citizen competence and the psychology of deliberation. *Deliberative democracy: Issues and cases*, 98–117.
- Rossini, P. (2019). Disentangling uncivil and intolerant discourse in online political talk [Num Pages: 238]. In R. G. Boatright, T. J. Shaffer, S. Sobieraj, & D. G. Young (Eds.). Routledge. Retrieved November 22, 2022, from <https://eprints.gla.ac.uk/272713/>
- Salganik, M. J. (2019). *Bit by Bit: Social Research in the Digital Age* [Google-Books-ID: 58iXDwAAQBAJ]. Princeton University Press.

- Sawyer, J., & Gampa, A. (2018). Implicit and Explicit Racial Attitudes Changed During Black Lives Matter [Publisher: SAGE Publications Inc]. *Personality and Social Psychology Bulletin*, *44*(7), 1039–1059.  
<https://doi.org/10.1177/0146167218757454>
- Scheufle, D. A., & Moy, P. (2000). Twenty-five years of the spiral of silence: A conceptual review and empirical outlook [Publisher: Oxford University Press]. *International journal of public opinion research*, *12*(1), 3–28.
- Singh, S. P., & Roy, J. (2014). Political knowledge, the decision calculus, and proximity voting [Publisher: Elsevier]. *Electoral Studies*, *34*, 89–99.
- Spoon, J.-J., & Klüver, H. (2019). Party convergence and vote switching: Explaining mainstream party decline across Europe. *European Journal of Political Research*, *58*(4), 1021–1042. <https://doi.org/10.1111/1475-6765.12331>
- Steenbergen, M. R., Bächtiger, A., Spörndli, M., & Steiner, J. (2003). Measuring political deliberation: A discourse quality index. *Comparative European Politics*, *1*(1), 21–48.
- Stietz, J., Jauk, E., Krach, S., & Kanske, P. (2019). Dissociating Empathy From Perspective-Taking: Evidence From Intra- and Inter-Individual Differences Research. *Frontiers in Psychiatry*, *10*, 126.  
<https://doi.org/10.3389/fpsyt.2019.00126>
- Strandberg, K. (2008). Public deliberation goes on-line? An analysis of citizens' political discussions on the Internet prior to the Finnish parliamentary elections in 2007. *Javnost-the public*, *15*(1), 71–89.
- Strandberg, K., & Grönlund, K. (2018). Online Deliberation. In *The Oxford Handbook of Deliberative Democracy*. Retrieved October 22, 2021, from <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780198747369.001.0001/oxfordhb-9780198747369-e-28>
- Sunstein, C. R. (2002). The law of group polarization. *Journal of political philosophy*, *10*(2), 175–195.

- Sunstein, C. R. (2018). *# Republic: Divided democracy in the age of social media*. Princeton University Press.
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving Decisions About Health, Wealth, and Happiness* (rev. a. exp. ed. edition). Penguin Books.
- Theocharis, Y., & Jungherr, A. (2021). Computational Social Science and the Study of Political Communication. *Political Communication*, *38*(1-2).  
<https://doi.org/10.1080/10584609.2020.1833121>
- Tóth, M., & Chytilék, R. (2018). Fast, frugal and correct? An experimental study on the influence of time scarcity and quantity of information on the voter decision making process [Publisher: Springer]. *Public Choice*, *177*(1), 67–86.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124–1131.
- Votta, F. (2019). peRspective: A wrapper for the Perspective API. *R*.  
<https://github.com/favstats>
- Wallsten, K. (2007). Agenda setting and the blogosphere: An analysis of the relationship between mainstream media and political blogs. *Review of policy research*, *24*(6), 567–587.
- Weiss-Blatt, N. (2022). Don't Be So Certain That Social Media Is Undermining Democracy. *The Daily Beast*. Retrieved August 12, 2022, from <https://www.thedailybeast.com/dont-be-so-certain-that-social-media-is-undermining-democracy>
- Wilhelm, A. G. (1998). Virtual sounding boards: How deliberative is on-line political discussion? *Information Communication & Society*, *1*(3), 313–338.



## Appendix A

### Publications

#### **P1. A systematic review of worldwide causal and correlational evidence on digital media and democracy**

##### **Journal Publication**

Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., and Hertwig, R. (2023). A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nature Human Behaviour*, 7(1), Article 1.

<https://doi.org/10.1038/s41562-022-01460-1>

##### **Supplementary Material**

Supplementary information including data, code, preregistration protocol, MOOSE checklist and supplementary figures available under <https://osf.io/7ry4a/>.

Live version of the dataset available under

<https://reclaimingautonomyonline.notion.site/reclaimingautonomyonline/b07fcb5727f94ea3b213e7cf6bdc190b?v=de4064525248412d9a1730066353285c>.

# A Systematic Review of Worldwide Causal and Correlational Evidence on Digital Media and Democracy

now published at:

<https://www.nature.com/articles/s41562-022-01460-1>

Philipp Lorenz-Spreen<sup>1,\*,\dagger</sup>      Lisa Oswald<sup>2,\dagger</sup>      Stephan Lewandowsky<sup>3,4</sup>  
Ralph Hertwig<sup>1</sup>

<sup>1</sup>Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

<sup>2</sup>Hertie School, Berlin, Germany

<sup>3</sup>School of Psychological Science and Cabot Institute, University of Bristol, Bristol, United Kingdom

<sup>4</sup>School of Psychological Science, University of Western Australia, Perth, Australia

<sup>\dagger</sup>these authors contributed equally; \*lorenz-spreen@mpib-berlin.mpg.de

March 14, 2023

## ABSTRACT

One of today's most controversial and consequential issues is whether the global uptake of digital media is causally related to a decline in democracy. We conducted a systematic review of causal and correlational evidence ( $N = 496$  articles) on the link between digital media use and different political variables. Some associations, such as increasing political participation and information consumption, are likely to be beneficial for democracy and were often observed in autocracies and emerging democracies. Other associations, such as declining political trust, increasing populism, and growing polarization, are likely to be detrimental to democracy and were more pronounced in established democracies. While the impact of digital media on political systems depends on the specific variable and system in question, several variables show clear directions of associations. The evidence calls for research efforts and vigilance by governments and civil societies to better understand, design and regulate the interplay of digital media and democracy.

## Introduction

The ongoing heated debate on the opportunities and dangers that digital media pose to democracy has been hampered by disjointed and conflicting results (for recent overviews see [1, 2, 3, 4]). Disagreement about the role of new media is not a novel phenomenon; throughout history, evolving communication technologies have provoked concerns and debates. One likely source of concern is the dual-use dilemma, that is, the inescapable fact that technologies can be used for both noble and malicious aims. For instance, during the Second World War, radio was used as a propaganda tool by Nazi Germany [5], whereas allied radio, such as the BBC, supported resistance against the Nazi regime, for example, by providing tactical information on allied military activities [6, 7]. In the context of the Rwandan genocide, radio was used to incite Rwandan Hutus to massacre the country's Tutsi minority [8]. In the aftermath of the genocide, using the same means to cause different ends, the radio soap opera *Musekweya* successfully reduced intergroup prejudice in a yearlong field experiment [9, 10].

Digital media appears to be another double-edged sword. On the one hand, it can empower citizens, as demonstrated in movements such as the Arab Spring [11], Fridays for Future, and #MeToo [12]. On the other hand, digital media can also be instrumental in inciting destructive behaviours and tendencies such as polarization and populism [13], as well as fatal events such as the attack on the United States Capitol in January 2021. Relatedly, the way political leaders use or avoid digital media can vary greatly depending on the political context. U.S. President Trump used it to spread numerous lies ranging from claims about systematic voter fraud in the 2020 presidential election to claims about the harmlessness of Covid-19. In spring 2022, Russian President Putin has most social media platforms banned, probably to prevent access to information about his army's attack on Ukraine, that would bypass the state-controlled classical media [14]. At the same time, Ukrainian President Zelensky has skilfully used social media to boost Ukrainian morale and win the information war with Russia. Examples of the dual-use dilemma of digital media abound.

Clearly, digital media can foster liberation, democratization, and participation, but can also play an important role in eroding democracy. The role of digital media is further complicated because unlike other communication technologies, it enables individuals to easily produce and disseminate content themselves and offers largely frictionless interaction between users. These properties have not only moved the self-organised political behaviour of citizens into the spotlight [15], they have also shifted power to large digital media platforms. Unlike broadcasters, digital media platforms typically do not create content; instead, their power lies in providing and governing a digital infrastructure. Although that infrastructure could serve as an online public sphere [16], it is the platforms, that exert much control over the dynamics of information flow.

Our goal is to advance the scientific and public debate on the relationship between digital media and democracy by providing an evidence-based picture of this complex constellation. To this end, we comprehensively reviewed and synthesized the available scientific knowledge [17] on the link between digital media and various politically significant variables such as participation, trust and polarization.

We aimed to answer the preregistered question *If, to what degree and in which contexts, do digital media have detrimental effects on democracy?* (preregistered protocol, including research question and search strategy, at <https://osf.io/7ry4a/>). This two staged question encompasses, first, the assessment of the direction of effects, and, second, how these effects play out as a function of political contexts.

A major difficulty facing researchers and policy makers is that most studies relating digital media use to political attitudes and behaviours are correlational. Because it is nearly impossible to simulate democracy in the laboratory, researchers are forced to rely on observational data that typically only provide correlational evidence. We therefore pursued two approaches: First, we collected and synthesized a broad set of articles that examine associations between digital media use and different political variables. We then conducted an in-depth analysis of the small subset of articles reporting causal evidence. This two-step approach permitted us to focus on causal effects while still taking the full spectrum of correlational evidence into account.

For the present purpose, we adopted a broad understanding of digital media, ranging from general internet access to the use of specific social media platforms, including exposure to certain types of content on these platforms. To be considered as valid digital media variable in our review, information or discussion forums have to be hosted via the internet or need to describe specific features of online communication. For example, we considered the online outlets of traditional newspapers or TV channels as digital source of political information but not the original traditional media themselves. We provide an overview of digital media variables present in our review sample in Fig. 1d and discern in our analyses between the two overarching types of digital media: internet, broadly defined, on the one hand and social media in particular on the other hand.

We further aimed to synthesize evidence on a broad spectrum of political attitudes and behaviours that are relevant to basic democratic principles [18]. We therefore grounded our assessment of political variables in the literature that examines elements of modern democracies that are considered essential to their functioning, such as citizens' basic trust in media and institutions [19], a well-informed public [20], an active civil society [21, 22], and exposure to a variety of opinions [23, 24]. We also included phenomena that are considered detrimental to the functioning of democracies, including open discrimination against people [25], political polarization to the advantage of political extremists and populists [26], and social segregation in homogeneous networks [23, 27].

The political variables in focus are themselves multi-dimensional and may be heterogeneous and conflicting. For example, polarization encompasses partisan sorting [28], affective polarization [29], issue alignment [30, 31], and a number of other phenomena (see [32] for an excellent literature review on media effects on variations of ideological and affective polarization). For our purpose, however, we take a broader perspective, examining and comparing across different political variables the directions—beneficial or detrimental to democracy—in which digital media effects play out.

Notwithstanding the nuances within each dimension of political behavior, wherever possible we explicitly interpret each change in a political variable as tending to be either beneficial or detrimental to democracy. Even though we try to

refrain from normative judgements, the nature of our research question required us to interpret the reported evidence with regard to its relation with democracy. For example, an increase in political knowledge is generally considered to be beneficial under the democratic ideal of an informed citizenry [20]. Similarly, a certain level of trust in democratic institutions is crucial for a functioning democracy [33]. By contrast, various forms of polarization (particularly affective polarization) tend to split societies into opposing camps and threaten democratic decision making [34, 35]. Likewise, populist politics, that are often coupled with right-wing nationalist ideologies, artificially divide society into a corrupt “elite” that is opposed by “the people”, which runs counter to the ideals of a pluralistic democracy and undermines citizens’ trust in politics and the media [36, 37]. We therefore consider polarization and populism, for example, to be detrimental for democracy.

There are already some systematic reviews of subsets of associations between political behaviour and media use that fall within the scope of our analysis, including reviews of the association between media and radicalization [38, 39], polarization [32], hate speech [40], participation [41, 42, 43, 44, 45], echo chambers [46], and campaigning on Twitter [47]. These extant reviews did, however, not contrast and integrate the wide range of politically relevant variables into one comprehensive analysis—an objective that we pursue here. For the most relevant review articles, we matched the references provided in them with our reference list (see Materials and Methods section for details). Importantly, and unlike some extant reviews, our focus is not on institutions, the political behaviour of political elites (e.g., their strategic use of social media; see [47, 48]), or higher-level outcomes (e.g., policy innovation in governments [49]). We also do not consider the effects of traditional media (e.g., television or radio) or consumption behaviours that are not specific to digital media (e.g., selective exposure [50]). Furthermore, we do not focus on the microscopic psychological mechanisms that could shape polarization on social media (for a review see [51]). For reasons of external validity, we omitted small scale laboratory-only experiments (e.g., [52]), but included field experiments in our review. We included studies using a variety of methods—from surveys to large-scale analyses of social media data—and across different disciplines, that are relevant to our research question. Details on the inclusion and exclusion criteria are provided in the Materials and Methods section. Our goal with this knowledge synthesis is to provide a nuanced foundation of shared facts for a constructive stage in the academic but also societal debate about the future of digital media and their role in democracy. In our view, this debate and the future design of digital media for democracy requires a comprehensive assessment of its impact. We therefore not only focus on individual dimensions of political behaviour but, thus going beyond the extant reviews, compare these dimensions and the methods by which they have been researched so far. This approach aims to stimulate research that fills evidence gaps and establishes missing links that only become apparent when comparing the dimensions.

## Results

After conducting a preregistered search (most recent update September 15, 2021) and selection process, we arrived at a final sample of  $N = 496$  articles. For further analysis, we classified them by the set of variables between which they report associations: type of digital media (e.g., social media, online news), political variables (e.g., trust, participation),

and characteristics of the information ecology (e.g., misinformation, selective exposure), as depicted in Fig. 1a. Each article was coded according to the combination of these variables as well as method, specific outcome variable, and, if applicable, the direction of association and potential moderator variables (see Materials and Methods for details). The resulting table of the fully coded set of studies is at <https://osf.io/7ry4a/>, alongside the code for the analyses and visualizations offered here.

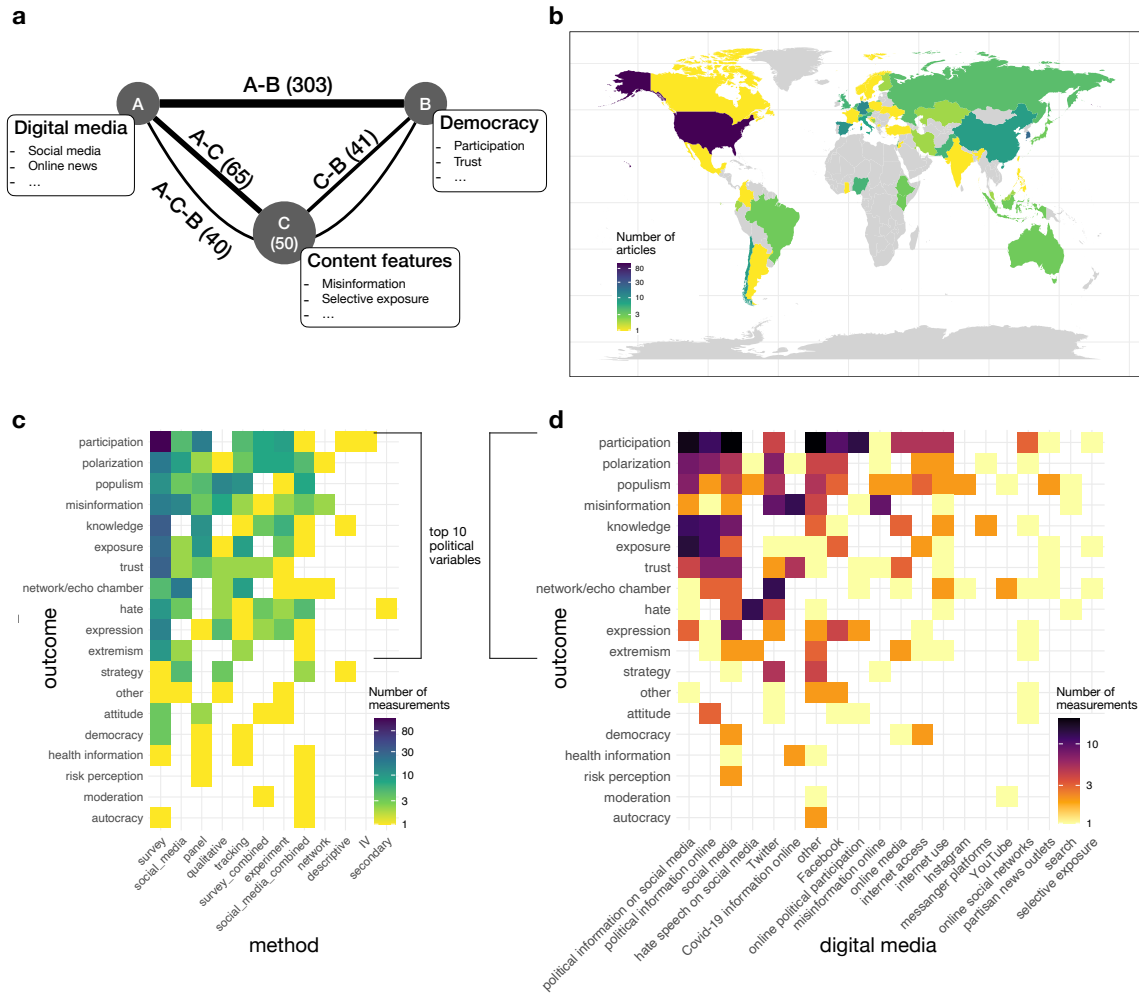


Figure 1: Summary of the reviewed articles. **a** Combinations of variables in the sample: digital media (A), political variables (B) and content features such as selective exposure or misinformation (C). Numbers in brackets count articles in our sample that measure an association between variables. **b** Geographic distribution of articles that reported site of data collection. **c** Distribution of measurements (counted separately whenever one article reported several variables) over combinations of outcome variables and methods and **d** over combinations of outcome variables and digital media variables.

Fig. 1 reports the composition of the set of included articles. Fig. 1a confirms that the search query mainly returned articles concerned with the most relevant associations between digital media and political outcomes (A–B,  $N = 299$ ). Most of the articles were published in the last five years, highlighting the fast growth of interest in the link between digital media and democracy. Articles span a range of disciplines, including political science, psychology, computational science, and communication science. Although a preponderance of articles focused on the United States, there was still a large geographical variation overall (see Fig. 1b).

Fig. 1c shows the distribution of measurements (counted separately when one article reported several outcomes) across methods and political variables. Our search query was designed to capture a broad range of politically relevant variables, which meant that we had to group them into broader categories. The ten most frequently reported categories of variables were trust in institutions, different variants of political participation (e.g., voter turnout or protest participation), exposure to diverse viewpoints in the news, political knowledge, political expression, measures of populism (e.g., support for far-right parties or anti-minority rhetoric), prevalence and spread of misinformation, measures of polarization (e.g., negative attitudes towards political opponents or fragmented and adversarial discourse), homophily in social networks (i.e., social connections between like-minded individuals) and online hate (i.e., hate speech or hate crime). Similarly, the distribution of outcomes and associated digital media variables in Fig. 1d shows that many studies focused on political information online, and specifically political information on social media, in combination with political polarization and participation, while other digital media variables, such as messenger platforms are less explored. The full table, including the reported political variables within each category, can be found at <https://osf.io/7ry4a/>. Fig. 1 also reveals gaps in the literature, such as rarely explored geographical regions (e.g., Africa) and under-studied methods–variable combinations (e.g., involving the combination of data sources such as social media data with survey or secondary data).

Direction of Associations

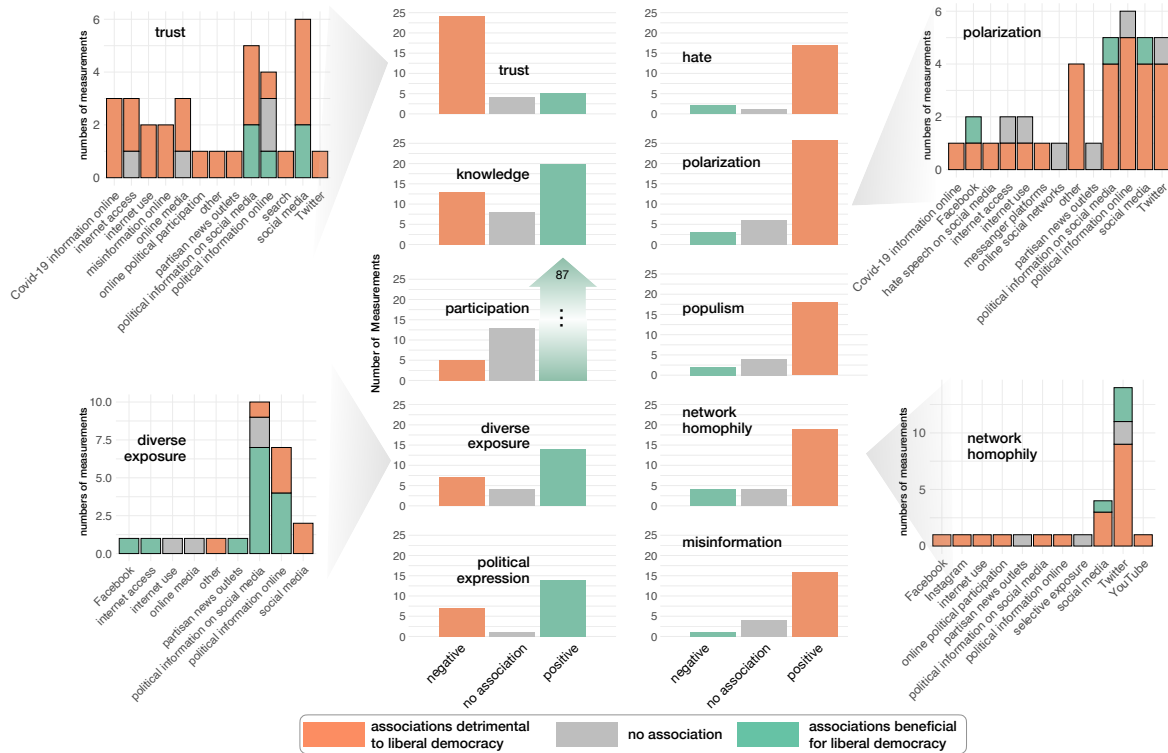


Figure 2: Distribution of directions of associations from the full sample, reported for various political variables (see Fig. 1d for a breakdown). Insets show exemplary the distribution of associations with trust, news exposure, polarization, and network homophily over the different digital media variables with which they were associated.

In the first part of our research question, we ask whether the available evidence suggest that the effects of digital media are predominantly beneficial or detrimental to democracy. To find an answer, we first selected subsets of articles that addressed the ten most frequently studied categories of political variables (from here on simply referred to as political variables). We did not test specific hypotheses in our review. A total of  $N = 355$  associations were reported for these variables (when an article examined two relevant outcome variables, two associations were counted). The independent variable across these articles was always a measure of the usage of some type of digital media, such as online news consumption or social media uptake. Statistically speaking, the independent variables can be positively or negatively associated with the political outcome variable. For instance, more digital media use could be associated with more expression of hate (positive association) or less expression of hate (negative association), or they could be not associated at all. We decided to present relationships not at a statistical level but at a conceptual level. We therefore classified each observed statistical association as beneficial or detrimental depending on whether its direction was aligned or misaligned with democracy. For example, a positive statistical association between digital media use and hate speech



was coded as a detrimental association; by contrast, a positive statistical association between digital media use and participation was coded as beneficial. Throughout, we represent beneficial associations in turquoise and detrimental associations in orange, irrespective of the underlying statistical polarity.

Fig. 2 provides an overview of the 10 most frequently studied political variables and the reported directions—color-coded in terms of whether they are beneficial or detrimental to democracy—of each of their associations with digital media use. This overview encompasses both correlational and causal evidence. Some findings in Fig. 2 suggest that digital media can foster democratic objectives. First, the associations reported for participation point mostly in beneficial directions for democracy (aligned with previous results [45]), including a wide range of political and civic behaviour (see Fig. 1d), from low-effort participation such as liking/sharing political messages on social media to high-cost activities such as protesting in oppressive regimes. Second, measures of political knowledge and diversity of news exposure appeared to be associated with digital media in beneficial ways, but the overall picture was slightly less clear. Third, the literature was also split on how political expression is associated with digital media. Articles reporting beneficial associations between digital media and citizens' political expression were opposed by a number of articles describing detrimental associations. These detrimental associations relate to the "spiral of silence" idea, that is, the notion that people's willingness to express their political opinions online depends on the perceived popularity of their opinions (see relevant overview articles [53, 54]).

Fourth, we observed consistent detrimental associations for a number of variables. Specifically, the associations with trust in institutions were overwhelmingly pointing in directions detrimental to a functioning democracy. Measures of hate, polarization, and populism were also widely reported to have detrimental associations with digital media use in the clear majority of articles. Likewise, increased digital media use was often associated with a greater exposure to misinformation. Finally, we also found that digital media were associated with homophily in social networks in detrimental ways (mostly measured on social media, and here especially Twitter), but the pattern of evidence was a little less consistent. Differences in the consistency of results were also reflected when broken down along associated digital media variables (see the insets in Fig. 2): For instance, both trust and polarization measures were consistently associated with media use across types of digital media, ranging from social media to political information online; in contrast, results for homophily were concentrated on social media and especially Twitter, while measurements of news exposure were mostly concentrated on political information online. This points not only to different operationalizations of related outcome measures, such as diverse information exposure and homophilic network structures, but also to differences between the distinct domains of digital media in which these very related phenomena are measured. Similar observations can be made when separating associations between general types of digital media: social media vs. internet more broadly in Supplementary Figure 1.

Next, we distinguished between articles reporting correlational versus causal evidence and focused on the small subset of articles reporting the latter ( $N = 24$ ). We excluded causal evidence on the effects of voting advice applications from our summary as a very specific form of digital media, explicitly constructed to inform vote choices, and already

extensively discussed in a meta analysis [55].

### **Causal inference**

Usually, the absence of randomized treatment assignment, an inescapable feature of observational data (e.g., survey data), precludes the identification of causal effects because individuals differ systematically on variables other than the treatment (or independent) variable. However, under certain conditions it is possible to rule out non-causal explanations for associations, even in studies without random assignment that are based on observational data (see [56, 57, 58]). For a more detailed explanation of the fundamental principles of causal inference, see the Supplementary Material page 5 and, for example, the work of 2021 laureates of the Nobel Memorial Prize in Economics [56, 57, 58].

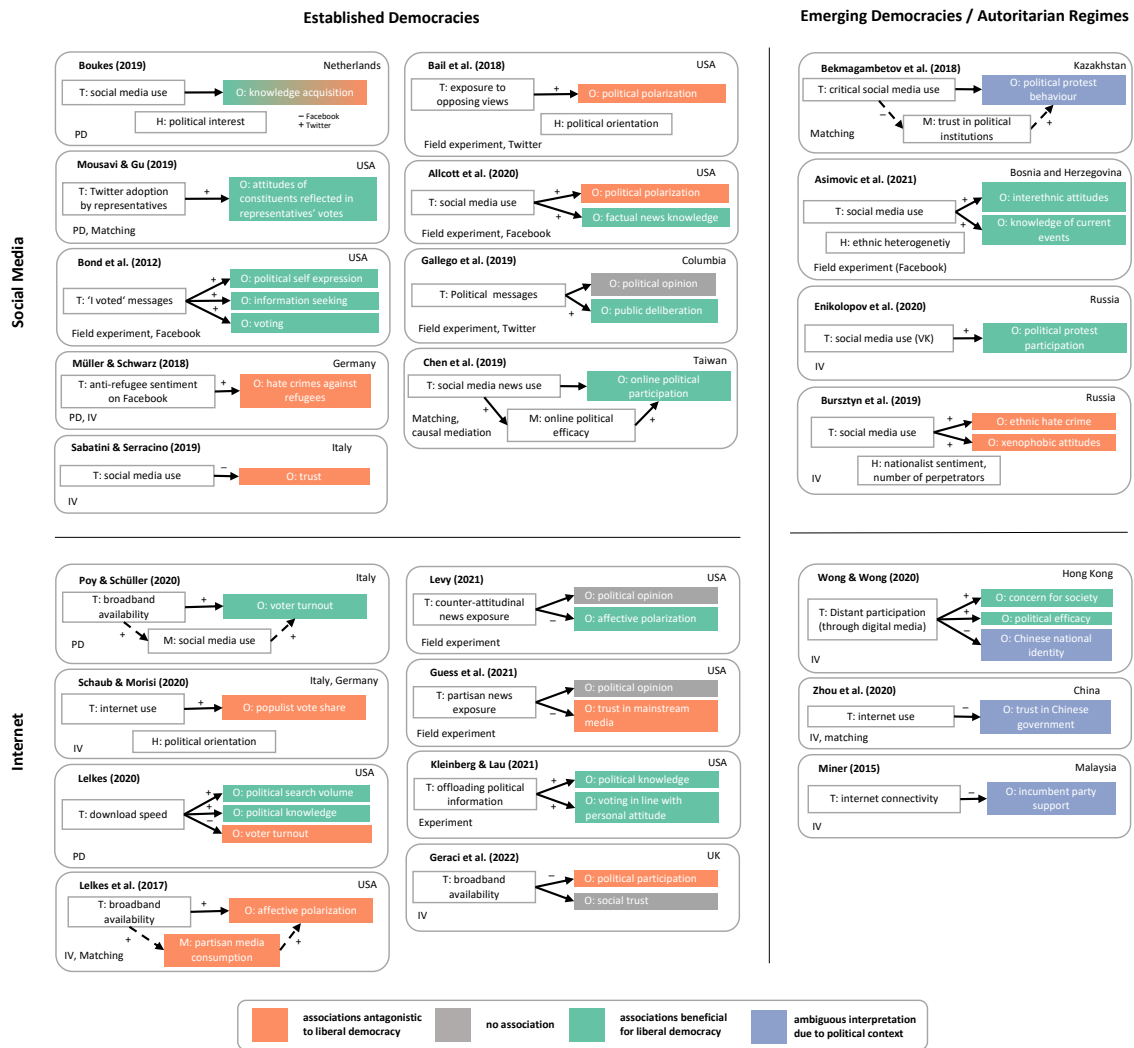


Figure 3: Summary of causal evidence for digital media effects on political variables. Each box represents one article. Treatments are in white boxes on the left, political outcome variables in coloured boxes on the right; M denotes mediators; H represents sources of effect heterogeneity or moderators. Positive (+) and negative (-) signs at paths indicate reported direction of effects. Location of sample indicated in top right corner of boxes, primary causal inference strategy in bottom left. Strategies include statistical estimation strategies such as instrumental variables (IV), matching, and panel designs (PD) that use, for example, fixed effects (FE) or difference in difference (DiD) for causal estimation, as well as lab or field experiments (e.g., field experiments rolled out on various platforms that are often supplemented with IV estimation to account for imperfect compliance). Detrimental effects on liberal democracy are shown in orange, beneficial effects in turquoise, effects open to interpretation in purple, and null effects in grey. Solid arrows represent pathways for which authors provide causal identification strategies, dashed arrows represent descriptive (mediation) pathways.

Common causal inference techniques that were used in our sample include instrumental variable designs that introduce exogenous variation in the treatment variable [59, 60, 61, 62, 63], matching approaches to explicitly balance treatment and control groups [64, 65, 66], and panel designs that account for unobserved confounders with unit and/or time fixed effects [67, 68]. We also found multiple large-scale field experiments conducted on social media platforms [69, 70, 71, 72] as well as various natural experiments [61, 73, 62, 59].

Fig. 3 summarizes the findings and primary causal inference techniques of these articles. Again, causal effects were coded as beneficial for or detrimental to democracy. This figure is structured according to whether evidence stemmed from established democracies or from emerging democracies and authoritarian regimes, adopting classifications from the Liberal Democracy Index provided by the Varieties of Democracy project [18]. In some autocratic regimes (e.g., China), it is particularly difficult to interpret certain effects. For example, a loss of trust in government suggests a precarious development for an established democracy; in authoritarian regimes, however, it may indicate a necessary step toward overcoming an oppressive regime and, eventually, progressing toward a more liberal and democratic system. Instead of simply adopting the authors' interpretation of the effects or imposing our own interpretation of effects in authoritarian contexts, we leave this interpretation to the reader (denoted in purple in the figure). The overall picture converges closely with the one drawn in Fig. 2. We found general trends of digital media use increasing participation and knowledge but also increasing political polarization and decreasing trust that mostly aligned with correlational evidence.

### Effects on key political variables

In the following, we provide a short synopsis of the results, point to conflicting trends, and highlight some examples of the full set of correlational and causal evidence, reported in Fig. 2 and Fig. 3, for six variables that we found to be particularly crucial for democracy: participation, trust, political knowledge, polarisation, populism, network structures and news exposure. The chosen examples are stand-ins and illustrations of the general trends.

**Participation.** Consistent with past meta-analyses [42, 43, 45], the body of correlational evidence supported a beneficial association between digital media use and political participation and mobilization.

Causal analyses of the effects of digital media on political participation in established democracies mostly studied voting and voter turnout [74, 67, 75, 71, 76, 64]; articles concerned with other regions of the world focused rather on political protest behavior [66, 61, 59]. Other articles considered online political participation [71, 65]. One study, applying causal mediation analysis to assess a causal mechanism [77], found that information-oriented social media use affects political participation, mediated or enabled through the user's online political efficacy [65]. Overall, our evidence synthesis found largely beneficial mobilizing effects for political participation across this set of articles. Our search did not identify any studies that examined causal effects of digital media on political participation in authoritarian regimes in Africa or the Middle East.

**Trust.** Many articles in our sample found detrimental associations between digital media and various dimensions of trust (Fig. 2). For example, detrimental associations were found for trust in governments and politics [78, 60, 79, 80, 59, 81, 66, 82], trust in media [83], and social and institutional trust [84]. During the COVID-19 pandemic, digital media use was reported to be negatively associated with trust in vaccines [85, 86]. Yet the results about associations with trust are not entirely homogeneous. One multinational survey found beneficial associations with trust in science [87]; others found increasing trust in democracy with digital media use in Eastern and Central European samples [88, 89]. Nevertheless, the large majority of reported associations between digital media use and trust appear to be detrimental for democracy. While the evidence stems mostly from surveys, results gathered with other methods underpin these findings (see Fig. 2 inset).

The majority of articles identifying causal effects also find predominantly detrimental effects of digital media on trust. A field experiment in the United States that set browser defaults to partisan media outlets [37] found a long-term loss of trust in mainstream media. Studies examining social trust as a central component of social capital find consistent detrimental effects of social media use [84]; in contrast, they find no effects of broadband internet in general on social trust [90]. In authoritarian regimes in Asia, increasing unrestricted internet access decreased levels of trust in the political system [91, 73, 59]. This finding confirms the predominant association observed in most other countries. Yet it also illustrates how digital media is a double-edged sword, depending on the political context: By reducing trust in institutions, digital media can threaten existing democracies as well as foster emerging democratic developments in authoritarian regimes.

**Political knowledge.** The picture was less clear for associations between the consumption of digital media and political knowledge. Still, the majority of associations point in beneficial directions and were found in both cross-sectional surveys [92, 93, 94, 95, 96, 97, 98, 99] and panel surveys [100, 101, 102]. Studies linking web-tracking and survey data showed increased learning about politics [103], but also a turning away from important topics [104], whereas other experiments demonstrated an overall beneficial effect of digital media on issue salience [105]. These findings, however, stand in contrast to other studies that find a detrimental association between political knowledge and digital media use [106, 107, 108, 109, 110].

The body of causal evidence on political knowledge also tends to paint a relatively promising picture. Multiple articles found that engagement with digital media increased political knowledge [67, 74, 70, 72] and that engagement with political content on social media increased political interest among adolescents [111]. In line with these findings, it has been reported that political messages on social media, as well as faster download speed, can increase information-seeking in the political domain [71, 67]. By contrast, there is evidence for a decrease in political knowledge [112], which is mediated through the news-finds-me effect: Social media users' believe that actively seeking out news is no longer required to stay informed, as they expect to be presented with important information.

It is important to note that most of these effects are accompanied by considerable heterogeneity in the population that benefits and the type of digital media. For example, politically interested individuals showed higher knowledge

acquisition when engaging with Twitter, whereas the opposite effects emerged for engagement with Facebook [113]. Furthermore, there is evidence that the news-finds-me effect on social media can be mitigated when users consult alternative news sources [112].

**Polarization.** Most articles found detrimental associations between digital media and different forms of political polarization [114, 115, 116, 117, 118]. Our review obtained evidence for increasing out-group polarization on social media in a range of political contexts and on various platforms [119, 120, 121, 122]. Increasing polarization was also linked to exposure to viewpoints opposed to one's own on social media feed [123, 69]. Articles comparing several political systems found associations that were country-dependent [124], again highlighting the importance of political context [125]. Nevertheless, high digital media use was for the most part linked to higher levels of polarization, although there was some evidence for balanced online discourse without pronounced patterns of polarization [126, 127, 128], as well as evidence for potentially depolarizing tendencies [129].

The body of causal articles largely supported the detrimental associations of digital media that emerged, by and large, in the correlational articles. Among established democracies, both social media use and overall internet use increased political polarization [63, 70]. This was also the case for an experimental treatment that exposed users to opposing views on Twitter [69]. However, some findings run counter to the latter result [130]: in a 2-month field experiment, exposure to counterattitudinal news on Facebook reduced affective polarization (the authors used opposing news outlets as treatment instead of opinions on social media). Furthermore, one other field experiment did not find evidence that exposure to partisan online news substantively shifted political opinions, but found a long-term loss of trust in mainstream media [37]. Still, taking all evidence into account, the overall picture remains largely consistent on the detrimental association between digital media and political polarization, including some but not all causal evidence.

**Populism.** Articles on populism in our review examined either vote share and other popularity indicators for populist parties or the prevalence of populist messages and communication styles on digital media. Overall, articles using panel surveys, tracking data, and methods linking surveys to social media data consistently found that digital media use was associated with higher levels of populism. For example, digital platforms were observed to benefit populist parties more than they benefit established politicians [131, 132, 133, 134]. In a panel survey in Germany, a decline in trust that accompanied increasing digital media consumption was also linked to a turn towards the hard-right populist AfD party [80]. This relationship might be connected to AfD's greater online presence, relative to other German political parties, [132], even though these activities might be partly driven by automated accounts. There is also evidence for an association between increased social media use and online right-wing radicalization in Austria, Sweden, and Australia [135, 136, 137]. Only a minority of articles found no relationship or the reverse relationship between digital media and populism [138, 139, 140]. For instance, in Japan, internet exposure was associated with increased tolerance towards foreigners [141].

Similarly, most causal inference studies linked increased populism to digital media use. For instance, digital media use in Europe led to increased far-right populist support [142, 63], and there was causal evidence that digital media can

propagate ethnic hate crimes in both democratic and authoritarian countries [68, 62]. Leaving the US and European political context, in Malaysia, internet exposure was found to cause decreasing support for the authoritarian, populist government [60].

**Echo chambers and news exposure.** The evidence on echo chambers points in different directions depending on the outcome measure. On the one hand, when looking at news consumption, several articles showed that social media and search engines diversify people's news diets [143, 144, 145, 67, 146]. On the other hand, when considering social networks and the impact of digital media on homophilic structures, the literature contains consistent reports of ideologically homogeneous social clusters [147, 148, 149, 150, 151]. This underscores an important point: Some seemingly paradoxical results can potentially be resolved by looking more closely at context and specific outcome measurement (see also Supplementary Figure 2). The former observation of diverse news exposure might fit with the beneficial relationship between digital media and knowledge reported in [74, 67, 94, 102, 95], and the homophilic social structures could be connected to the prevalence of hate speech and anti-outgroup sentiments [152, 153, 120, 154, 155].

Heterogeneity

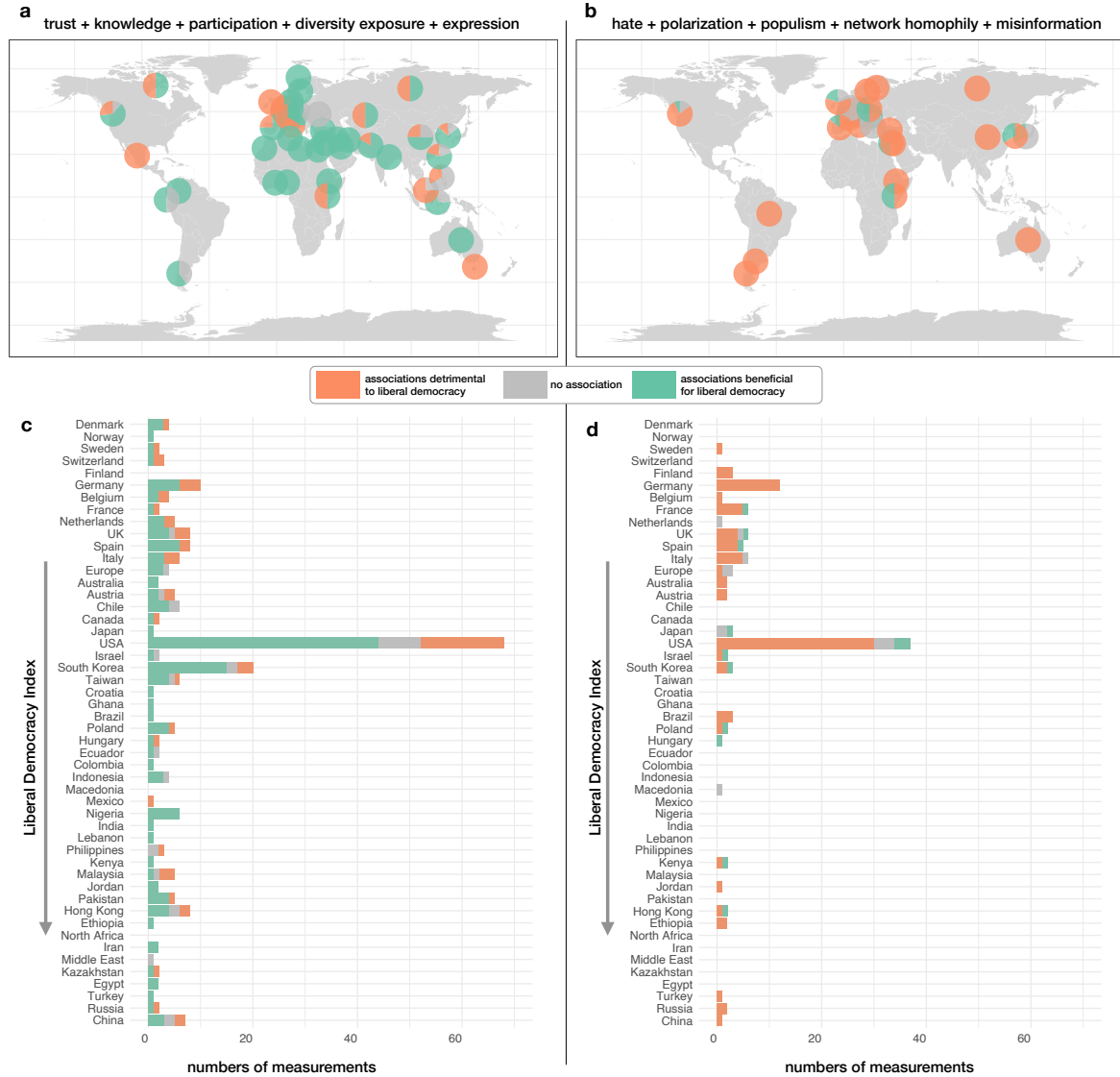


Figure 4: Geographical distribution of associations, columns showing beneficial and detrimental outcomes. **a** Geographical distribution of reported associations for the variables trust, knowledge, participation, exposure, and expression. Pie charts show the composition of directions for each country studied. **b** Geographic representation of reported associations for the variables hate, polarization, populism, homophily, and misinformation. **c** Data and variables in **a**, in absolute numbers of reported associations and sorted along the Liberal Democracy Index [18]. **d** Data and variables in **b**, in absolute numbers of reported associations and sorted along the Liberal Democracy Index.



We now turn to the second part of our research question and analyse the effects of digital media use in light of different political contexts. Fig. 4 shows the geographical distribution of effect directions around the globe. Notably, most beneficial effects on democracy were found in emerging democracies in South America, Africa, and South Asia. Mixed effects, by contrast, were distributed across Europe, the United States, Russia, and China. Similarly, detrimental outcomes were mainly found in Europe, the United States and partly Russia, although this may reflect a lack of studies undertaken in authoritarian contexts. These patterns are also shown in Fig. 4c and d, where countries are listed according to the Liberal Democracy Index. Moderators—variables such as partisanship and news consumption that are sources of effect heterogeneity—displayed in Supplementary Figure 3 also show slight differences between outcomes. Beneficial outcomes seemed to be more often moderated by political interest and news consumption, whereas detrimental outcomes tended to be moderated by political position and partisanship.

Furthermore, many causal articles acknowledge that effects differ between subgroups of their sample by including interaction terms in their statistical models. For example, the polarizing effects of digital media differ between Northern and Southern European media systems [142]: While consumption of right-leaning digital media increased far-right votes, especially in Southern Europe, the consumption of news media and public broadcasting in Northern European media systems with high journalistic standards appears to mitigate these effects. Another example of differential effects between subgroups was found in Russia, where the effects of social media on xenophobic violence were only present in areas with pre-existing nationalist sentiment. This effect was especially pronounced for hate crimes with a larger number of perpetrators, indicating that digital media was serving a coordinating function. In summary, a range of articles found heterogeneity in effects for varying levels of political interest [67, 113], political orientation [63, 70, 69], and different characteristics of online content [111].

Most authors, particularly those of the causal inference articles in our body of evidence, explicitly emphasized the national, cultural, temporal and political boundary conditions for interpreting and generalising their results (see, e.g., [111]). By contrast, especially in articles conducted on US samples, the national context and results' potential conditionality was often not highlighted. We strongly caution against a generalization of findings that are necessarily bound to a specific political setting (e.g. the US) to other contexts.

Sampling methods and risk of bias

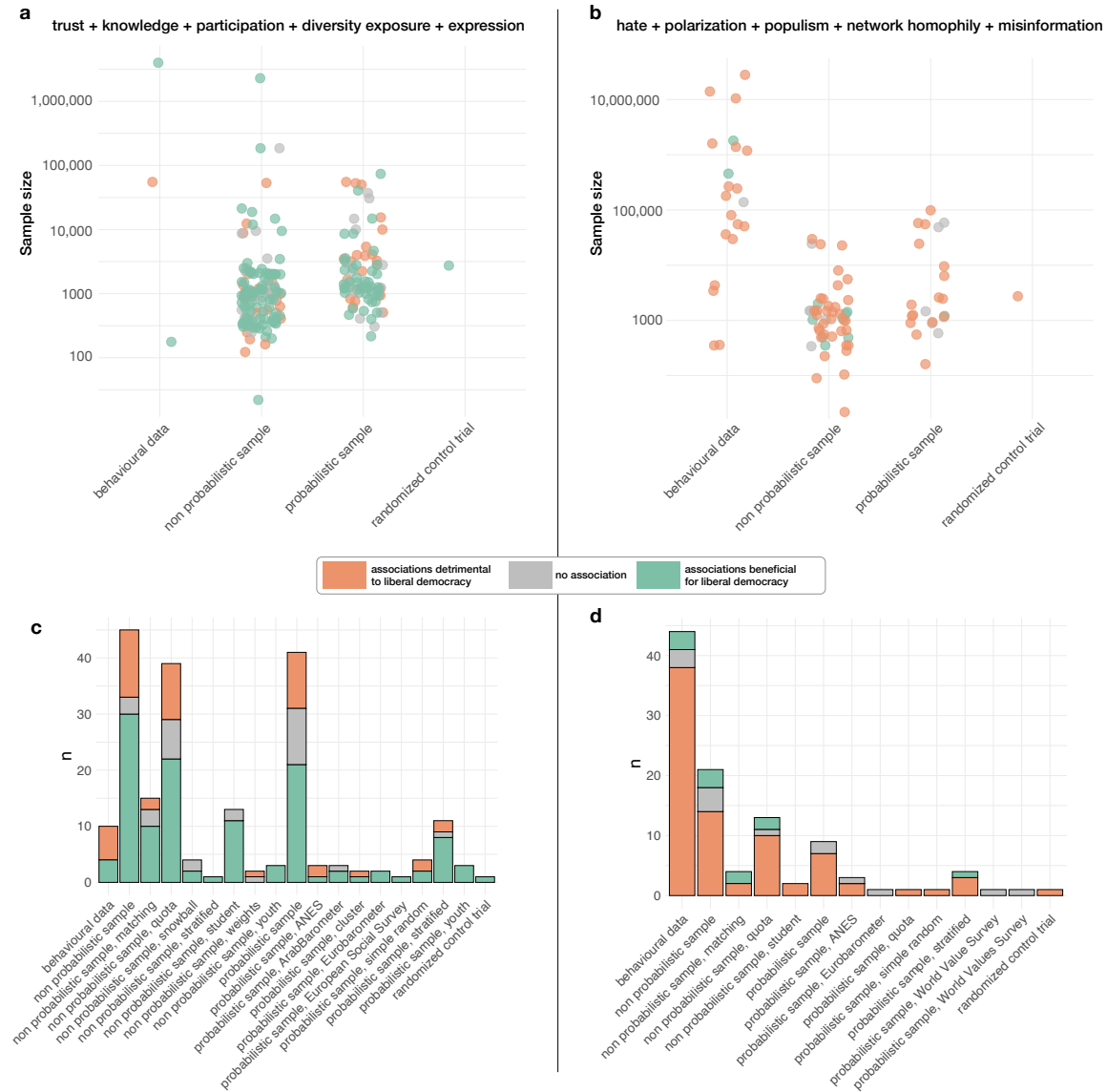


Figure 5: Sample size and sampling strategy for reported associations, columns are split between beneficial and detrimental outcomes. **a** and **b** show sample size vs. sampling methods, each dot represents one measurement, as well as color coded the direction of the reported association. **c** and **d** break down sampling methods in more detail and count the number of reported associations including their direction.

In order to assess study quality and risk of bias, we additionally coded important methodological aspects of the studies, specifically, the sampling method, sample size and transparency indicators, such as competing interest, open data

practices and pre-registrations. In Fig. 5, we show an excerpt from that analysis. Different sampling methods naturally result in different sample sizes as shown in Fig. 5a and b. Furthermore, behavioural data is much more prevalent for studies that look at detrimental outcomes, such as polarization and echo chambers. Classic surveys with probability samples or quota samples, in contrast, are often used to examine beneficial outcome measures like trust and participation (see Fig. 5c and d). Overall, however, no very clear pattern emerges in terms of the reported directions of associations. If anything, large probabilistic samples report relatively less beneficial associations for both types of outcomes (see Fig. 5). Generally, different types of data have different advantages, such as probability and quota samples approximating more closely the ideal of representativeness, whereas the observation of actual behavior on social media escaping the potential downsides of self-report. A potential blind spot in studies working with behavioural data from social media, inaccessible to both us and the original authors of the studies, is the selection of data provided by platforms. Therefore, it is tremendously important for researchers to get unrestricted access or, at least, transparent provision of random samples of data by platforms. The selection of users into the platforms, however, remains an open issue for behavioural data as it is often unclear who the active users are and why they are active online. We find that political outcome measures studied with behavioral data appear to show quite distinct results compared to when studied with large-scale survey data. Combining both data types would likely maximize the chances for reliable conclusions about the impact of digital media on democracy.

We found relatively few null effects for some variables. This could be accurate, but it could also be driven by the file-drawer problem—the failure to publish null results. In an attempt to examine the extent of a potential file-drawer problem, we contacted authors via large mailing lists but did not receive any unpublished work that fitted our study selection criteria. With regard to possible risk of bias, we found that only in 143 out of 355 measurements, authors clearly communicated that no conflict of interest was present (beyond the usual funding statement). However, we did not find a striking imbalance in the distribution of reported associations between those articles in which no competing interest was and was not explicitly stated. The few associations for which a conflict of interests was stated, four pointed in beneficial, three in detrimental and two reported lack of directionality. In only 79 of 355 measurements the researchers did use open data practices. Considering articles that reported detrimental associations, we did not find a clear difference in the directions between those with and without open data. However, considering articles that report beneficial outcomes, the numbers of positive findings in the studies without open data are relatively much larger than for the open science studies. Namely, 103 beneficial and 33 detrimental associations were reported in those without open data, while 19 beneficial versus 14 detrimental were reported in studies with open data practices. This observation might be due to the large number of survey-based studies about participation, which often do not follow open data practices. Even fewer of the studies in our sample were pre-registered, namely, 13 of the 355, where nine of those reported detrimental associations and only three beneficial and one found no direction of association. To shed light on other potential biases, we additionally examined temporal variations in the directions of reported associations and found, besides the general explosive growth of studies in this domain, a slight trend toward an increasing number of both detrimental directions and null effects over time (see Supplementary Figure 4). On the author-level, there was no

clear pattern in the associations reported by those authors who published the greatest number of articles in our sample; several authors variously reported detrimental and beneficial effects as well as null effects, with a few exceptions (see Supplementary Figure 5). Their co-authorship network in Supplementary Figure 6, split for the two types of outcomes measures, some communities of co-authors can be seen, however, no clear pattern of preferred direction of reported association can be spotted. Overall, we did not find evidence of a systematic bias in either direction driven by temporal trends or particular authors.

## Discussion

Regardless of whether they are authoritarian, illiberal, or democratic, governments around the world are concerned with how digital media affects governance and their citizenry's political beliefs and behaviors. A flurry of recent interdisciplinary research, stimulated in part by new methodological possibilities and data sources, has shed light on this potential interplay.

Although classical survey methods are still predominant, novel ways of linking data types, for example linking URL tracking data or social media data with surveys, permit more complex empirical designs and analyses. Furthermore, digital trace data allow an expansion in sample size. The articles we reviewed included surveys with a few hundred, up to a few thousand participants, but also large-scale social media analyses that included behavioral traces of millions. Yet with computational social science in its early days, the amount of evidence supporting and justifying causal conclusions is still limited. Causal effects of digital media on political variables are also hard to pin down empirically due to a plethora of complexities and context factors, as well as the highly dynamic technological developments that make predicting the future difficult. While emergent political phenomena are hard to simulate in the lab, the value of estimation and data collection strategies to draw causal inferences from real-life data is enormous. However, the long established trade-off between internal and external validity still applies which also highlights the value of high quality descriptive work.

Taking into account both, correlational and causal evidence, our review suggests that digital media use is clearly associated with variables such as trust, participation, and polarization. They are critical for the functioning of any political system, in particular democracies. Extant research reports relatively few null effects. However, the trends on each factor mostly converge, both across research methods and across correlative and causal evidence.

Our results also highlight that digital media are a double-edged sword, with both beneficial and detrimental effects on democracy. What is considered beneficial or detrimental will, at least partly, hinge on the political system in question: Intensifying populism and network homophily may benefit a populist regime or a populist politician but undermine a pluralistic democracy. For democratic countries, evidence clearly indicates that digital media increases political participation. Less clear but still suggestive are the findings that digital media has positive effects on political knowledge and exposure to diverse viewpoints in news. On the negative side, however, digital media use is associated with eroding the "glue that keeps democracies together" [33]: trust in political institutions. The results indicating this danger

converge across methods. Furthermore, our results also suggest that digital media use is associated with increases in hate, populism, and polarisation. Again, the findings converge across causal and correlational articles.

Alongside the need for more causal evidence, we found several research gaps, including the relationship between trust and digital media and the seeming contradiction between network homophily and diverse news exposure. Methods that link tracking data for measuring news exposure with behavioural data from social media (e.g., sharing activities or the sentiment of commenting) are crucial to a better understanding of this apparent contradiction.

### **Limitations**

The articles in our sample incorporate a plethora of methods and measures. As a result, it was necessary to classify variables and effects into broad categories. This is a trade-off we had to make in exchange for the breadth of our overview of the landscape of evidence across disciplines. For the same reason, we could not provide a quantitative comparison across the diverse sample of articles. We believe that digital media research would benefit from more unified measures (e.g., for polarization), methods across disciplines to allow for better comparability in the future, and a systematic comparison of different types digital media (i.e., Facebook and Twitter are not of one kind nor, in all likelihood, are their effects), and the extensions of outcome measurements beyond certain types of digital media. This follows other recent calls for commensurate measures of political and affective polarization [156]. The breadth of our review and the large number of political outcome measures in particular, made it necessary to be quite restrictive on other ends (see Fig. 6 for our exclusion process and Table S1 for the detailed criteria). We explicitly decided to prioritize the selection of causal evidence (see Fig. 7 for an overview of the causal inference techniques that we considered) and other large-sample, quantitative, published evidence. However, following this preregistered search strategy led to the selection of unequal numbers of studies for different outcome variables. For example, our search query selected considerably more studies examining political participation than political expression or trust, while at the same time, it did not include all studies that are included in other systematic reviews [45] due to stricter exclusion criteria.

The interpretation of our results was in several cases hampered by a lack of appropriate baseline measures. There is no clear measure of what constitutes a reasonable benchmark of desirable political behaviour in a healthy democracy. In addition, there were no means of quantification of some of these behaviours in the past, outside of digital media. This problem is particularly pronounced for factors such as exposure to diverse news, social network homophily, misinformation, and hate speech. Measuring these phenomena at scale is possible through digital media (e.g., by analysing social network structure); much less is known about their prevalence and dynamics in offline settings. Many articles therefore lacked a baseline. For instance, it is not clear what level of homophily in social networks is desirable or undesirable in a democratic society. Nor is it clear how to interpret the results of certain studies on polarization [69, 130], whose findings depend on whether one assumes that social media has increased or decreased the exposure to opposing views relative to some offline benchmark. For example, if exposure to opposing views is increased on social media, the conclusion of one study [130] would be that it reduces polarization, but if exposure is decreased one would come to the opposite conclusion. Notably, in this study, counter attitudinal exposure was found to be down-ranked by

Facebook's news feed—hence supporting a process that fosters polarization instead of counteracting it. Furthermore, results about populism might be skewed: Descriptive evidence on the relative activity and popularity of right-wing populist parties in Europe suggests their over-representation, as in the case of Germany's AfD, on social media, relative to established democratic parties (see e.g., [132]). Therefore, it is difficult to interpret even causal effects of digital media use on populist support in isolation from the relative preponderance of right-wing content online.

### **Conclusion**

Our results provide grounds for concern. Alongside the positive effects of digital media for democracy, there is clear evidence of serious threats to democracy. In light of the importance of these corrosive and potentially difficult-to-reverse effects for democracy, a better understanding of the diverging effects of digital media in different political contexts (e.g., authoritarian vs. democratic) is urgently needed. To this end, methodological innovation is required. This includes, for instance, more research using causal inference methodologies, as well as research that examines digital media use across multiple and interdependent measures of political behaviour. More research and better study designs will, however, also depend on access to data collected by the platforms. This access has been restricted or foreclosed. Yet without independent research that has unhampered access to all relevant data, the effects of digital media can hardly be understood in time. This is concerning even more so because digital media can implement architectural changes that, even if seemingly small, can scale up to widespread behavioral effects. Regulation may be required to facilitate this access [157]. Most importantly, we suggest that the bulk of empirical findings summarized here can be attributed to the current status quo of an information ecosystem produced and curated by large, commercial platforms. They have succeeded in attracting a vast global audience of users. The sheer size of their audience as well as their power over what content and how content gets the most attention has led, in the words of the philosopher Jürgen Habermas, to a new structural transformation of the public sphere [16]. In this new public sphere, everybody can be a potential author spontaneously producing content, both right-wing radical networks as well as the courageous Belarusian women standing up for human rights and against a repressive regime. One need not share Habermas' conception of "deliberate democracy" to see that current platforms fail to produce an information ecosystem that empowers citizens to make political choices that are as rationally motivated as possible. Our results show how this ecosystem plays out to have important consequences for political behaviours and attitudes. They further underscore that finding out which aspects of this relationship are detrimental to democracy and how they can be contained while actively preserving and fostering the emancipatory potential of digital media is, perhaps, one of the most important global tasks of the present. Our analysis hopes to contribute to the empirical basis of this endeavour.

## Materials and Methods

This systematic review follows the MOOSE Guidelines for Meta-Analyses and Systematic Reviews of Observational Studies [158]. The detailed protocol of the review process was preregistered on OSF at <https://osf.io/7ry4a/>. The repository also contains the completed MOOSE checklist showing where each guideline is addressed in the text.

Fig. 6 summarizes the search query that we used on two established academic databases, Scopus and Web of Science (both highly recommended search tools [159]), the resulting number of articles from the query and the subsequent exclusion steps, leading to the final sample size of  $N = 496$  articles under consideration.

**Study selection criteria.** We included only original, empirical work. Conceptual or theoretical work, simulation studies, and evidence synthesizing studies were excluded. Articles had to be published in academic journals in English. Unpublished studies for which only the abstract or a preprinted version was available were excluded from the review. We excluded small- $N$  laboratory experiments and small- $N$  student surveys ( $N < 100$ ) from our body of original work due to validity concerns. Although correlational evidence cannot establish a causal direction, we focused on articles that examined effects of digital media on democracy but not the opposite. We therefore excluded, for example, articles that examined ways to digitize democratic procedures. To be included, articles had to include at least two distinct variables, a digital media variable and a political outcome. Articles measuring a single variable were only included if this variable was a feature of digital media (e.g., hate speech prevalence, homophily in online social networks, prevalence of misinformation in digital media).

**Search strategy, study selection, coding and data extraction.** Articles eligible for our study had to be published before September 15, 2021. We sourced our review database from Scopus and Web of Science, as suggested by [159]. The search query (see Fig. 6) was constructed in consultation with professional librarians and was designed to be as broad as possible in order to pick up any articles containing original empirical evidence of direct or indirect effects of digital media on democracy (including correlational evidence). We further consulted recent, existing review articles in the field [39, 32, 40] to check for important articles that did not appear in the review body. Articles that were included manually are referenced separately in the flow chart (see Fig. 6). In addition, we contacted authors via large mailing lists of researchers working on computational social science and misinformation, but did not receive any unpublished work that fitted our study selection criteria. The query retrieved  $N = 3,512$  articles. Of these, 1,352 were retained after screening the titles for irrelevant topics. This first coding round, whether an article, based on the title, fits the review frame or not was split between two coders. Coders could flag articles that are subject to discussion to let the other coder double check the decision. In this round, only clearly not fitting articles were excluded from the sample. A list of exclusion criteria can be found in the SI.

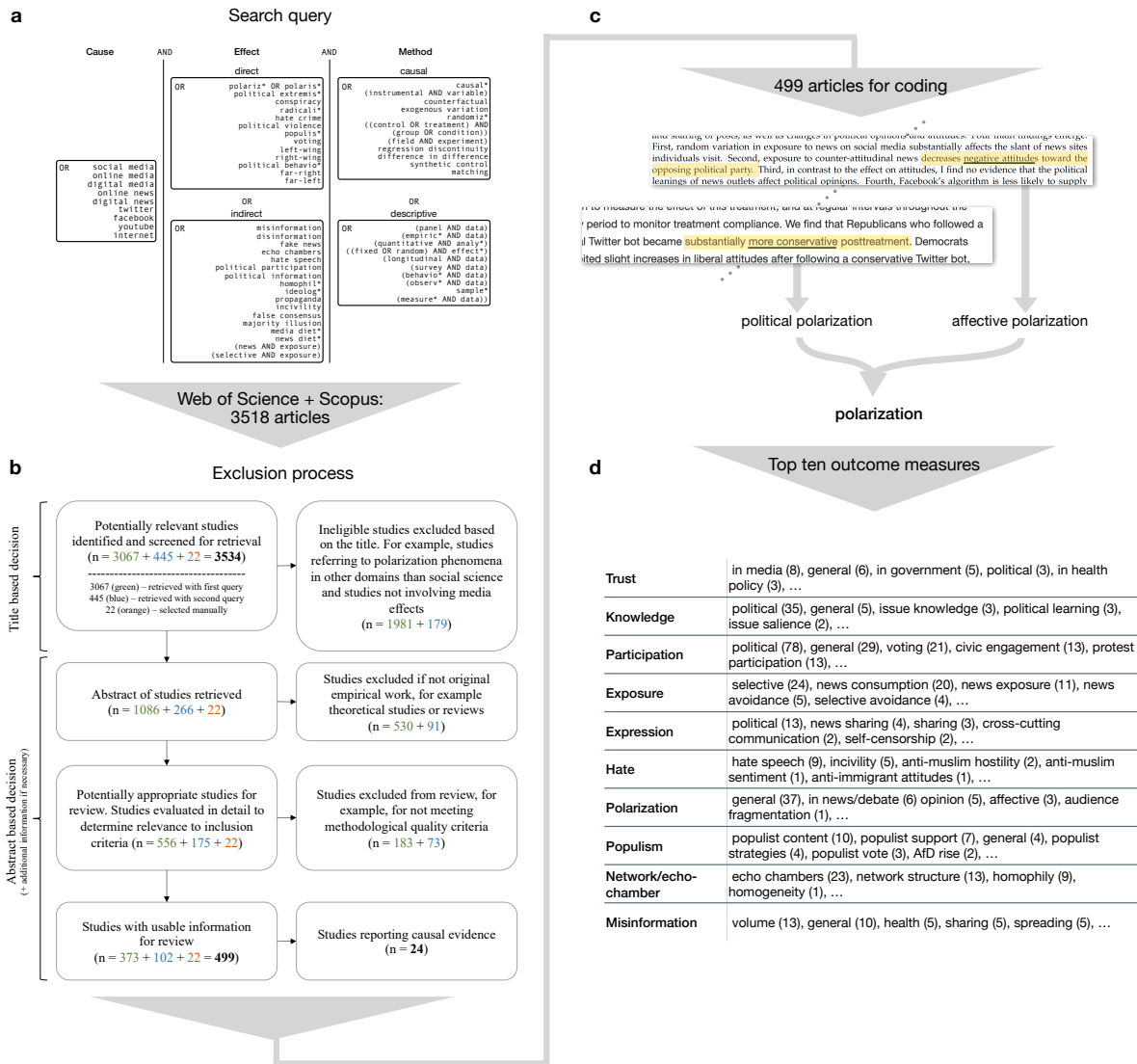


Figure 6: Strategy for curating the sample of relevant articles. **a** Keywords included in our search query, run on Web of Science and Scopus, with logical connectors. Focus on causal inference methods (method column), but also inclusion of descriptive quantitative evidence. Relationships between digital media (cause column) and political outcomes (direct effect box) or content features (indirect effect box). **b** Flow-chart representing the step wise exclusion process, starting with title based exclusion, followed by abstract based exclusion. **c** Exemplary illustration of the outcome variable extraction from the abstracts. **d** Breakdown of the most frequently reported political variables into top 10 categories. Numbers in brackets are counts of measurements in the set.



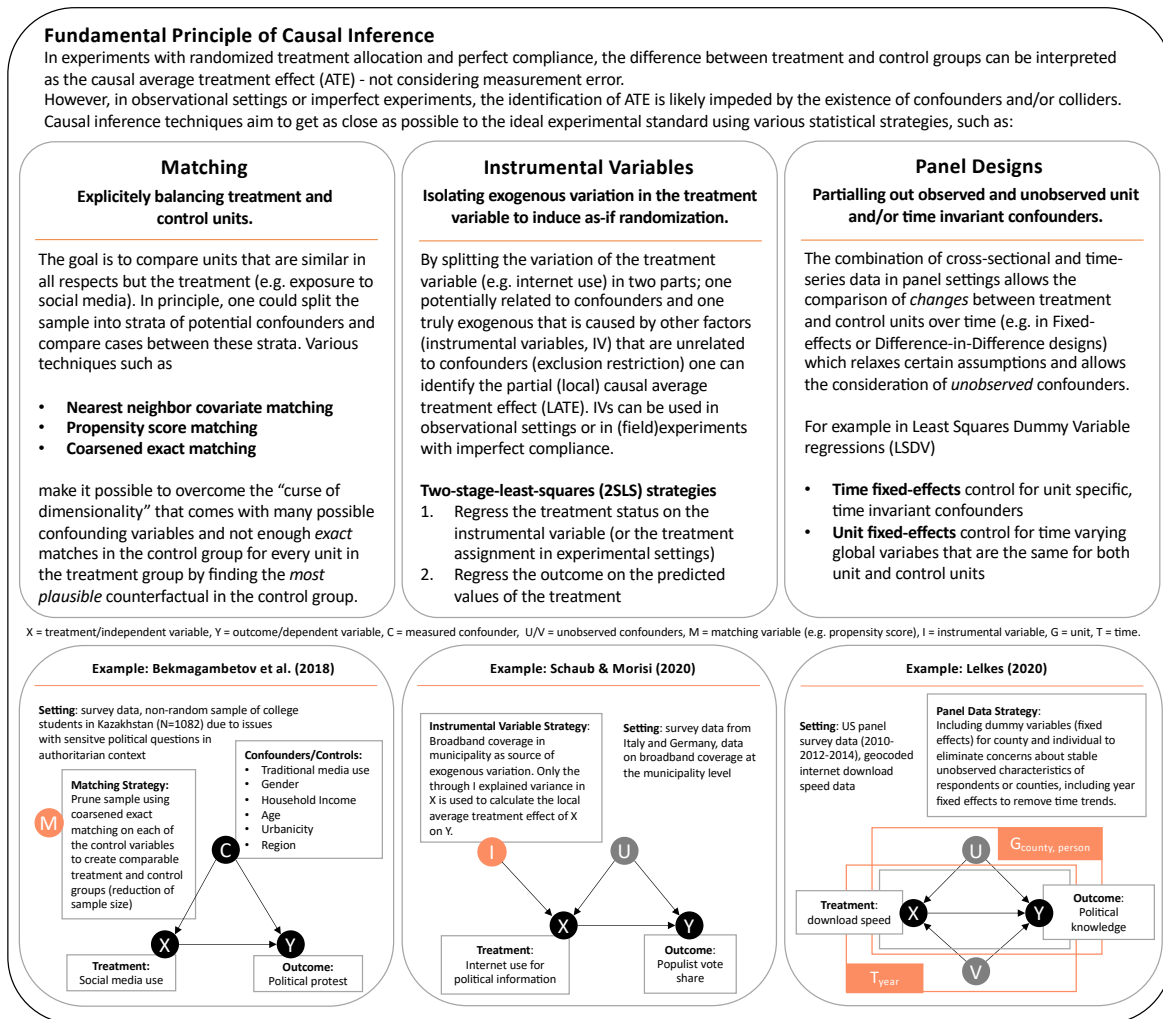


Figure 7: Summary of causal inference techniques used in our sample of causal evidence (excluding field experiments).

The next coding round, whether an article, based on the abstract, fits the review frame was conducted in parallel by two coders. The inter-coder reliability, after this round of article selection, was Krippendorff’s alpha of 0.66 (87% agreement). After calculating this value, disagreement between coders was solved through discussion. At this stage, we excluded all studies that were not original empirical work such as other reviews or conceptual articles, simulation studies and purely methodological articles (e.g., hate speech or misinformation detection approaches). This coding round was followed by a more in-depth coding round. Here we refined our exclusion decisions (e.g. excluded studies that examined the digitisation of government, preprints, small-scale lab experiments, small-scale convenience or student samples and studies that only included one variable (e.g. description of online forums), see Supplementary Table 1 for a

detailed list of criteria. A full-text screen was performed in cases where the relevant information could not be retrieved from the abstract and for all articles implying causal evidence.

After both rounds of abstract screening, 475 articles remained in our sample. After cross-checking the results of our literature search against the references from existing reviews, we found and included further  $N = 24$  articles that met our thematic criteria but were not identified by our search string. Ultimately, a total of 496 articles were selected into the final review sample. Fig. 6b summarizes the selection procedure.

The following information was extracted from each article using a standardized data extraction form: variable groups under research (digital media, features of media and/or political outcome variables), the concrete digital media under research, the explicit political outcome variable, the methods used, the country of origin, causal claims, possible effect heterogeneity (moderation) as well as various potential sources of bias. In order to assess various quality criteria of the studies, the coders had to visit the full text of the articles (e.g. to find the declaration of competing interests, preregistration or data availability statements, or to consider the methods section). Therefore, and facing the large number of articles under consideration, blinding could not be established during this procedure.

When conducting a systematic review with a broad scope, categories of the variables cannot be exhaustively defined prior to coding. Therefore, variable categories, especially for the digital media variables and the political outcome variables, were chosen inductively. In the first extraction step, coders stuck closely to the phrasing of the authors of the respective study. To reduce redundancy and refine the clustering of the variables, we iteratively generated frequency tables and manually sorted single variables to the best fitting categories until a small number of clearly distinct categories was selected. After the categories were defined, both coders re-coded 10% of the sample to calculate inter coder reliabilities for all key variables. We provide a table of inter coder reliabilities (percentage agreements and Krippendorff's Alphas), see Supplementary Table 2.

**Data synthesis and analysis.** Due to considerable heterogeneity in methods in the articles—including self-report surveys through network analysis of social media data, URL tracking data, and field experiments—no calculation of meta-analytic effect sizes was possible. The final table of selected articles with coded variables will be published alongside this article as a major result of this review project. The effect directions of 10 important political outcome variables (four consistent with liberal democracy, four opposing democratic values) are summarized in Fig. 2. For articles dealing with these political variables, we also assessed the country in which the study was conducted, explicit sources of effect heterogeneity such as demographic characteristics of study participants or characteristics of the digital media platform (Fig. 4).

For the overview analysis, which includes both correlational and causal evidence, we mainly restricted ourselves to the evaluation effects reported in the abstracts. Articles making explicit causal claims and/or using causal inference methods (Fig. 7) were examined in-depth and summarized as simplified path diagrams with information on mediators, moderators, country of origin, and method used (Fig. 3).

## 1 Deviations from the protocol

The volume of papers our query returned prevented an in-depth analysis of confounding variables. Instead, our assessment of quality relied on the sampling strategy and sampling strategy and sample size, the method used, sources of heterogeneity and transparency criteria, like open data practices and pre-registration. Furthermore, we were able to construct the co-author network by matching the author's names, but unable to produce a meaningful co-citation network due to the incompleteness and ambiguity of references in the export format that we used.

**Data availability:** The dataset including all originally collected studies with decision stages (N = 3531, "full\_data.xlsx"), the table including all papers within our review sample (N = 496, "data\_review.xlsx"), and the table including all effects reported within papers dealing with the top ten outcome measures (N = 355, "data\_effects.xlsx") are available at <https://osf.io/7ry4a/>.

**Code availability:** R scripts for all analyses and figures are available at <https://osf.io/7ry4a/>.

**Supplementary Information:** Supplementary information, including supplementary figures and tables can be found under [https://static-content.springer.com/esm/art%3A10.1038%2Fs41562-022-01460-1/MediaObjects/41562\\_2022\\_1460\\_MOESM1\\_ESM.pdf](https://static-content.springer.com/esm/art%3A10.1038%2Fs41562-022-01460-1/MediaObjects/41562_2022_1460_MOESM1_ESM.pdf).

**Acknowledgements:** We thank Simon Munzert for providing his perspective on causal inference and issues specific to political science. We thank Deborah Ain for editing the manuscript and Friederike Stock for help in the literature comparison. PLS, SL and RH acknowledge financial support from the Volkswagen Foundation (grant "Reclaiming individual autonomy and democratic discourse online: How to rebalance human and algorithmic decision making"). SL acknowledges support from the Humboldt Foundation through a research award. SL was also in part supported by an ERC Advanced Grant (PRODEMINFO) during completion of this paper. LO acknowledges financial support by the German National Academic Foundation in the form of a PhD scholarship. The authors received no specific funding for this work. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

**Author Contributions:** All authors designed the study. PLS and LO selected and coded the literature. PLS and LO evaluated the coded tables. All authors analyzed the results and wrote the manuscript. PLS and LO contributed equally to this work.

**Competing Interests Statement:** The authors report no competing interests.

## References

- [1] Persily, N. & Tucker, J. A. *Social Media and Democracy: The State of the Field, Prospects for Reform* (Cambridge University Press, 2020).
- [2] Tucker, J. A. *et al.* Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)* (2018).
- [3] Rau, J. P. & Stier, S. Die echokammer-hypothese: Fragmentierung der öffentlichkeit und politische polarisierung durch digitale medien? *Zeitschrift für Vergleichende Politikwissenschaft* **13**, 399–417 (2019).
- [4] Lewandowsky, S. *et al.* Technology and democracy: Understanding the influence of online technologies on political behaviour and decision-making (2020).
- [5] Adena, M., Enikolopov, R., Petrova, M., Santarosa, V. & Zhuravskaya, E. Radio and the Rise of the Nazis in Prewar Germany. *The Quarterly Journal of Economics* **130**, 1885–1939 (2015). Publisher: MIT Press.
- [6] Adena, M., Enikolopov, R., Petrova, M. & Voth, H.-J. Bombs, broadcasts and resistance: Allied intervention and domestic opposition to the Nazi regime during World War II. *Available at SSRN 3661643* (2021).
- [7] Gagliarducci, S., Onorato, M. G., Sobbrino, F. & Tabellini, G. War of the waves: Radio and resistance during World War II. *American Economic Journal: Applied Economics* **12**, 1–38 (2020).
- [8] Li, D. Echoes of violence: Considerations on radio and genocide in Rwanda. *Journal of Genocide Research* **6**, 9–27 (2004). Publisher: Taylor & Francis.
- [9] Paluck, E. L. Reducing intergroup prejudice and conflict using the media: a field experiment in Rwanda. *Journal of personality and social psychology* **96**, 574 (2009). Publisher: American Psychological Association.
- [10] Staub, E. & Pearlman, L. A. Reducing intergroup prejudice and conflict: A commentary. (2009). Publisher: American Psychological Association.
- [11] Howard, P. N. & Hussain, M. M. *Democracy's fourth wave?: digital media and the Arab Spring* (Oxford University Press, 2013).
- [12] Jackson, S. J., Bailey, M. & Welles, B. F. *#HashtagActivism: Networks of race and gender justice* (MIT Press, 2020).
- [13] Engesser, S., Ernst, N., Esser, F. & Büchel, F. Populism and social media: How politicians spread a fragmented ideology. *Information, communication & society* **20**, 1109–1126 (2017).
- [14] Warzel, C. The Information War Isnt Over Yet (2022). URL <https://www.theatlantic.com/technology/archive/2022/03/russia-ukraine-war-propaganda/626975/>.
- [15] Bak-Coleman, J. B. *et al.* Stewardship of global collective behavior. *Proceedings of the National Academy of Sciences* **118** (2021).

- [16] Habermas, J. Überlegungen und hypothesen zu einem erneuten strukturwandel der politischen öffentlichkeit. *LEVIATHAN* 470 (2021).
- [17] The value of evidence synthesis. *Nature Human Behaviour* **5**, 539–539 (2021).
- [18] Coppedge, M. *et al.* *Varieties of Democracy: Measuring Two Centuries of Political Change* (Cambridge University Press, Cambridge, 2020).
- [19] Warren, M. E. What kinds of trust does a democracy need? Trust from the perspective of democratic theory. In *Handbook on political trust* (Edward Elgar Publishing, 2017).
- [20] Milner, H. *Civic literacy: How informed citizens make democracy work* (UPNE, 2002).
- [21] Kohler-Koch, B. & Quittkat, C. *De-mystification of participatory democracy: EU-governance and civil society* (OUP Oxford, 2013).
- [22] O'connell, B. & Gardner, J. W. *Civil society: The underpinnings of American democracy* (UPNE, 1999).
- [23] Sunstein, C. R. The law of group polarization. *Journal of political philosophy* **10**, 175–195 (2002). Publisher: Wiley Online Library.
- [24] Habermas, J., Lennox, S. & Lennox, F. The public sphere: An encyclopedia article (1964). *New German Critique* 49–55 (1974). Publisher: JSTOR.
- [25] Howard, J. W. Free speech and hate speech. *Annual Review of Political Science* **22**, 93–109 (2019). Publisher: Annual Reviews.
- [26] Müller, J.-W. *What is populism?* (University of Pennsylvania Press, 2016).
- [27] Pariser, E. *The filter bubble: What the Internet is hiding from you* (Penguin UK, 2011).
- [28] Davis, N. T. & Dunaway, J. L. Party polarization, media choice, and mass partisan-ideological sorting. *Public Opinion Quarterly* **80**, 272–297 (2016).
- [29] Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. & Westwood, S. J. The origins and consequences of affective polarization in the united states. *Annual Review of Political Science* **22**, 129–146 (2019).
- [30] Kozlowski, A. C. & Murphy, J. P. Issue alignment and partisanship in the american public: Revisiting the partisans without constraint thesis. *Social Science Research* **94**, 102498 (2021).
- [31] Baldassarri, D. & Gelman, A. Partisans without constraint: Political polarization and trends in american public opinion. *American Journal of Sociology* **114**, 408–446 (2008).
- [32] Kubin, E. & von Sikorski, C. The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association* 1–19 (2021).
- [33] Dodsworth, S. & Cheeseman, N. Political trust: The glue that keeps democracies together. *London: Westminster Foundation for Democracy* (2020).

- [34] McCoy, J. & Somer, M. Toward a theory of pernicious polarization and how it harms democracies: Comparative evidence and possible remedies. *The Annals of the American Academy of Political and Social Science* **681**, 234–271 (2019).
- [35] Danescu, E. Varieties of democracy (v-dem) annual report 2019-" democracy facing global challenges. Tech. Rep., V-Dem Institute, the Department of Political Science at the University of Gothenburg (2019).
- [36] Bächtiger, A., Dryzek, J. S., Mansbridge, J. & Warren, M. Deliberative democracy. *The Oxford handbook of deliberative democracy* 1 (2018).
- [37] Guess, A. M., Barberá, P., Munzert, S. & Yang, J. The consequences of online partisan media. *Proceedings of the National Academy of Sciences* **118** (2021). Publisher: National Acad Sciences.
- [38] OdaĀĀ, Ā., Leiser, A. & Boehnke, K. Reviewing the role of the internet in radicalization processes1. *Journal for Deradicalization* 261–300 (2019).
- [39] Hassan, G. *et al.* Exposure to Extremist Online Content Could Lead to Violent Radicalization: A Systematic Review of Empirical Evidence. *International Journal of Developmental Sciences* **12**, 71–88 (2018).
- [40] Castano-Pulgarín, S. A., Suárez-Betancur, N., Vega, L. M. T. & López, H. M. H. Internet, social media and online hate speech. systematic review. *Aggression and Violent Behavior* 101608 (2021).
- [41] Angyal, E. & Fellner, Z. How are online and offline political activities connected? A comparison of studies. *Intersections East European Journal of Society and Politics* **6**, 81–98 (2020).
- [42] Chae, Y., Lee, S. & Kim, Y. Meta-analysis of the relationship between Internet use and political participation: examining main and moderating effects. *Asian Journal of Communication* **29**, 35–54 (2019).
- [43] Oser, J. & Boulianne, S. Reinforcement effects between digital media use and political participation : a meta-analysis of repeated-wave panel data. *Public Opinion Quarterly* **84**, 355–365 (2020).
- [44] Boulianne, S. Social media use and participation: A meta-analysis of current research. *Information, communication & society* **18**, 524–538 (2015).
- [45] Boulianne, S. Twenty Years of Digital Media Effects on Civic and Political Participation. *Communication Research* **47**, 947–966 (2020).
- [46] Terren, L. & Borge-Bravo, R. Echo chambers on social media: A systematic review of the literature. *Review of Communication Research* **9**, 99–118 (2021).
- [47] Jungherr, A. Twitter use in election campaigns: A systematic literature review. *Journal of information technology & politics* **13**, 72–91 (2016). Publisher: Taylor & Francis.
- [48] Lewandowsky, S., Jetter, M. & Ecker, U. K. Using the presidents tweets to understand political diversion in the age of social media. *Nature communications* **11**, 1–12 (2020).
- [49] LaCombe, S. J., Tolbert, C. & Mossberger, K. Information and policy innovation in us states. *Political Research Quarterly* 10659129211006783 (2021).

- [50] Stroud, N. J. Polarization and partisan selective exposure. *Journal of communication* **60**, 556–576 (2010).
- [51] Van Bavel, J. J., Rathje, S., Harris, E., Robertson, C. & Sternisko, A. How social media shapes polarization. *Trends in Cognitive Sciences* (2021).
- [52] Banks, A., Calvo, E., Karol, D. & Telhami, S. # polarizedfeeds: Three experiments on polarization, framing, and social media. *The International Journal of Press/Politics* **26**, 609–634 (2021).
- [53] Scheufle, D. A. & Moy, P. Twenty-five years of the spiral of silence: A conceptual review and empirical outlook. *International journal of public opinion research* **12**, 3–28 (2000).
- [54] Matthes, J., Knoll, J. & von Sikorski, C. The spiral of silence revisited: A meta-analysis on the relationship between perceptions of opinion support and political opinion expression. *Communication Research* **45**, 3–33 (2018).
- [55] Munzert, S. & Ramirez-Ruiz, S. Meta-Analysis of the effects of voting advice applications. *Political Communication* 1–16 (2021). Publisher: Taylor & Francis.
- [56] of Sciences, T. R. S. A. The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2021 (2021).
- [57] Angrist, J. D. & Pischke, J.-S. *Mostly harmless econometrics* (Princeton university press, 2008).
- [58] Angrist, J. D. & Pischke, J.-S. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of economic perspectives* **24**, 3–30 (2010).
- [59] Zhou, D., Deng, W. & Wu, X. Impacts of Internet Use on Political Trust: New Evidence from China. *Emerging Markets Finance and Trade* **56**, 3235–3251 (2020).
- [60] Miner, L. The unintended consequences of internet diffusion: Evidence from Malaysia. *Journal of Public Economics* **132**, 66–78 (2015).
- [61] Enikolopov, R., Makarin, A. & Petrova, M. Social Media and Protest Participation: Evidence From Russia. *Econometrica* **88**, 1479–1514 (2020).
- [62] Petrova, M., Bursztyn, L., Egorov, G. & Enikolopov, R. Social Media and Xenophobia: Evidence from Russia (2020).
- [63] Schaub, M. & Morisi, D. Voter mobilisation in the echo chamber: Broadband internet and the rise of populism in Europe. *European Journal of Political Research* **59**, 752–773 (2020). Publisher: Wiley Online Library.
- [64] Mousavi, R. & Gu, B. The impact of twitter adoption on lawmakers’ voting orientations. *Information Systems Research* **30**, 133–153 (2019).
- [65] Chen, C., Bai, Y. & Wang, R. Online political efficacy and political participation: A mediation analysis based on the evidence from Taiwan. *New Media and Society* **21**, 1667–1696 (2019).
- [66] Bekmagambetov, A. *et al.* Critical social media information flows: political trust and protest behaviour among Kazakhstani college students. *Central Asian Survey* **37**, 526–545 (2018).

- [67] Lelkes, Y. A Bigger Pie: The Effects of High-Speed Internet on Political Behavior. *Journal of Computer-Mediated Communication* **25**, 199–216 (2020).
- [68] Müller, K. & Schwarz, C. Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association* (2018).
- [69] Bail, C. A. *et al.* Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* **115**, 9216–9221 (2018). Publisher: National Acad Sciences.
- [70] Allcott, H., Braghieri, L., Eichmeyer, S. & Gentzkow, M. The welfare effects of social media. *American Economic Review* **110**, 629–76 (2020).
- [71] Bond, R. M. *et al.* A 61-million-person experiment in social influence and political mobilization. *Nature* **489**, 295–298 (2012).
- [72] Asimovic, N., Nagler, J., Bonneau, R. & Tucker, J. A. Testing the effects of Facebook usage in an ethnically polarized setting. *Proceedings of the National Academy of Sciences* **118** (2021). Publisher: National Academy of Sciences Section: Social Sciences.
- [73] Wong, S. H.-W. & Wong, M. Y. H. "Distant Participation" and Youth Political Attitudes: Evidence from a Natural Experiment. *Social Science Quarterly* **101**, 1489–1512 (2020).
- [74] Kleinberg, M. & Lau, R. Googling Politics: How Offloading Affects Voting and Political Knowledge. *Political Psychology* **42**, 93–110 (2021).
- [75] Poy, S. & Schüller, S. Internet and voting in the social media era: Evidence from a local broadband policy. *Research Policy* **49** (2020).
- [76] Lee, S. & Xenos, M. Incidental news exposure via social media and political participation: Evidence of reciprocal effects. *New Media and Society* (2020).
- [77] Imai, K., Keele, L. & Tingley, D. A general approach to causal mediation analysis. *Psychological methods* **15**, 309 (2010). Publisher: American Psychological Association.
- [78] Porumbescu, G. Not All Bad News After All? Exploring the Relationship Between Citizens' Use of Online Mass Media for Government Information and Trust in Government. *International Public Management Journal* **20**, 409–441 (2017).
- [79] Zhu, Z., Liu, Y., Kapucu, N. & Peng, Z. Online media and trust in government during crisis: The moderating role of sense of security. *International Journal of Disaster Risk Reduction* **50** (2020).
- [80] Zimmermann, F. & Kohring, M. Mistrust, Disinforming News, and Vote Choice: A Panel Survey on the Origins and Consequences of Believing Disinformation in the 2017 German Parliamentary Election. *Political Communication* **37**, 215–237 (2020).
- [81] Bucy, E. P. & Groshek, J. Empirical support for the media participation hypothesis: Trends across presidential elections, 1992–2012. *New Media & Society* **20**, 1889–1909 (2018).



- [82] Arlt, D. Who trusts the news media? Exploring the factors shaping trust in the news media in German-speaking Switzerland. *Studies in Communication Sciences* **18**, 231–245 (2019).
- [83] Park, S., Fisher, C., Flew, T. & Dulleck, U. Global Mistrust in News: The Impact of Social Media on Trust. *Imm-International Journal on Media Management* **22**, 83–96 (2020).
- [84] Sabatini, F. & Sarracino, F. Online Social Networks and Trust. *Social Indicators Research* **142**, 229–260 (2019). URL <https://doi.org/10.1007/s11205-018-1887-2>.
- [85] Carrieri, V., Madio, L. & Principe, F. Vaccine hesitancy and (fake) news: Quasi-experimental evidence from Italy. *Health Economics (United Kingdom)* **28**, 1377–1382 (2019).
- [86] Casara, B., Suitner, C. & Bettinsoli, M. Viral suspicions: Vaccine hesitancy in the web 2.0. *Journal of Experimental Psychology: Applied* **25**, 354–371 (2019).
- [87] Huber, B., Barnidge, M., Gil de Zuniga, H. & Liu, J. Fostering public trust in science: The role of social media. *Public Understanding of Science* **28**, 759–777 (2019).
- [88] Placek, M. A. #Democracy: social media use and democratic legitimacy in Central and Eastern Europe. *Democratization* **24**, 632–650 (2017).
- [89] Placek, M. Can the Internet Aid Democratic Consolidation?: Online News and Legitimacy in Central and Eastern Europe. Placek, M.(2018). *Can the Internet Aid Democratic Consolidation* (2018).
- [90] Geraci, A., Nardotto, M., Reggiani, T. & Sabatini, F. Broadband Internet and social capital. *Journal of Public Economics* **206**, 104578 (2022). URL <https://www.sciencedirect.com/science/article/pii/S0047272721002140>.
- [91] Min, G., Yu, Z. & Li, F. *Analysis of Moral Deviation in Netnews Post-Bumping* (2015).
- [92] Di, C. & Fang, W. New channels, new ways of becoming informed? Examining the acquisition of public affairs knowledge by young people in China. *Information Development* **35**, 688–702 (2019).
- [93] Wei, K., Lin, Y.-R. & Yan, M. Examining Protest as An Intervention to Reduce Online Prejudice: A Case Study of Prejudice Against Immigrants. 2443–2454 (2020).
- [94] Alam, A., Adnan, H. M. & Kotamjani, S. S. Examining the Impact of Using Social Networks on Political Knowledge and Political Attitude by Iranian University Students. *Jurnal Komunikasi-Malaysian Journal of Communication* **35**, 125–140 (2019).
- [95] Beaudoin, C. E. The internet's impact on international knowledge. *New Media & Society* **10**, 455–474 (2008).
- [96] Ida, R., Saud, M. & Mashud, M. An empirical analysis of social media usage, political learning and participation among youth: a comparative study of Indonesia and Pakistan. *Quality and Quantity* **54**, 1285–1297 (2020).
- [97] Salaudeen, M. & Onyechi, N. Digital media vs mainstream media: Exploring the influences of media exposure and information preference as correlates of media credibility. *Cogent Arts and Humanities* **7** (2020).

- [98] Imran, M. & Fatima, M. Connectivism: E-learning of democratic values on social media public spheres. vol. 2017-December, 82–89 (2018).
- [99] Park, C. & Kaye, B. News Engagement on Social Media and Democratic Citizenship: Direct and Moderating Roles of Curatorial News Use in Political Involvement. *Journalism and Mass Communication Quarterly* **95**, 1103–1127 (2018).
- [100] Gottfried, J. A., Hardy, B. W., Holbert, R. L., Winneg, K. M. & Jamieson, K. H. The Changing Nature of Political Debate Consumption: Social Media, Multitasking, and Knowledge Acquisition. *Political Communication* **34**, 172–199 (2017).
- [101] Kim, D. H. & Kwak, N. Media Diversity Policies for the Public: Empirical Evidence Examining Exposure Diversity and Democratic Citizenship. *Journal of Broadcasting & Electronic Media* **61**, 682–702 (2017).
- [102] Beam, M. A., Hutchens, M. J. & Hmielowski, J. D. Clicking vs. sharing: The relationship between online news behaviors and political knowledge. *Computers in Human Behavior* **59**, 215–220 (2016).
- [103] Edgerly, S., Thorson, K. & Wells, C. Young Citizens, Social Media, and the Dynamics of Political Learning in the U.S. Presidential Primary Election. *American Behavioral Scientist* **62**, 1042–1060 (2018).
- [104] Cardenal, A., Galais, C. & Majó-Vázquez, S. Is Facebook Eroding the Public Agenda? Evidence from Survey and Web-Tracking Data. *International Journal of Public Opinion Research* **31**, 589–608 (2019).
- [105] Feezell, J. T. Agenda Setting through Social Media: The Importance of Incidental News Exposure and Social Filtering in the Digital Era. *Political Research Quarterly* **71**, 482–494 (2018).
- [106] Kelly Garrett, R. Social medias contribution to political misperceptions in U.S. Presidential elections. *PLoS ONE* **14** (2019).
- [107] Lee, S. & Xenos, M. Social distraction? Social media use and political knowledge in two U.S. Presidential elections. *Computers in Human Behavior* **90**, 18–25 (2019).
- [108] van Erkel, P. & Van Aelst, P. Why Dont We Learn from Social Media? Studying Effects of and Mechanisms behind Social Media News Use on General Surveillance Political Knowledge. *Political Communication* 1–19 (2020).
- [109] Lee, S. Connecting social media use with gaps in knowledge and participation in a protest context: the case of candle light vigil in South Korea. *Asian Journal of Communication* **29**, 111–127 (2019).
- [110] Cacciatore, M. A. *et al.* Is Facebook Making Us Dumber? Exploring Social Media Use as a Predictor of Political Knowledge. *Journalism & Mass Communication Quarterly* **95**, 404–424 (2018).
- [111] Moeller, J., Shehata, A. & Kruikemeier, S. Internet use and political interest: Growth curves, reinforcing spirals, and causal effects during adolescence. *Journal of Communication* **68**, 1052–1078 (2018).
- [112] Lee, S. Probing the Mechanisms Through Which Social Media Erodes Political Knowledge: The Role of the News-Finds-Me Perception. *Mass Communication and Society* **23**, 810–832 (2020).

- [113] Boukes, M. Social network sites and acquiring current affairs knowledge: The impact of Twitter and Facebook usage on learning about the news. *Journal of Information Technology & Politics* **16**, 36–51 (2019).
- [114] Adam, S., Haussler, T., Schmid-Petri, H. & Reber, U. Coalitions and counter-coalitions in online contestation: An analysis of the German and British climate change debate. *New Media & Society* **21**, 2671–2690 (2019).
- [115] North, S., Piwek, L. & Joinson, A. Battle for Britain: Analyzing Events as Drivers of Political Tribalism in Twitter Discussions of Brexit. *Policy and Internet* (2020).
- [116] Bryson, B. Polarizing the middle: internet exposure and public opinion. *International Journal of Sociology and Social Policy* **40**, 99–113 (2019).
- [117] Lee, C., Shin, J. & Hong, A. Does social media use really make people politically polarized? Direct and indirect effects of social media use on political polarization in South Korea. *Telematics and Informatics* **35**, 245–254 (2018). Publisher: Elsevier.
- [118] Cho, J., Ahmed, S., Keum, H., Choi, Y. & Lee, J. Influencing Myself: Self-Reinforcement Through Online Political Expression. *Communication Research* **45**, 83–111 (2018).
- [119] Yarchi, M., Baden, C. & Kligler-Vilenchik, N. Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication* 1–42 (2020). Publisher: Taylor & Francis.
- [120] Workneh, T. Social media, protest, & outrage communication in Ethiopia: toward fractured publics or pluralistic polity? *Information Communication and Society* **24**, 309–328 (2021).
- [121] Hawdon, J., Ranganathan, S., Leman, S., Bookhultz, S. & Mitra, T. *Social media use, political polarization, and social capital: is social media tearing the u.s. apart?*, vol. 12194 LNCS of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2020).
- [122] Urman, A. News Consumption of Russian Vkontakte Users: Polarization and News Avoidance. *International Journal of Communication* **13**, 5158–5182 (2019).
- [123] Kibet, A. & Ward, S. Socially Networked Heterogeneity: The Influence of WhatsApp as a Social Networking Site on Polarisation in Kenya. *African Journalism Studies* **39**, 42–66 (2018).
- [124] Fletcher, R., Cornia, A. & Nielsen, R. How Polarized Are Online and Offline News Audiences? A Comparative Analysis of Twelve Countries. *International Journal of Press/Politics* **25**, 169–195 (2020).
- [125] Lu, J. & Luo, C. Development consensus in the Internet context: penetration, freedom, and participation in 38 countries. *Information Development* **36**, 288–300 (2020).
- [126] Lai, M., Tambuscio, M., Patti, V., Ruffo, G. & Rosso, P. Stance polarity in political debates: A diachronic perspective of network homophily and conversations on Twitter. *Data and Knowledge Engineering* **124** (2019).
- [127] Kobayashi, T., Ogawa, Y., Suzuki, T. & Yamamoto, H. News audience fragmentation in the Japanese Twitter-sphere. *Asian Journal of Communication* **29**, 274–290 (2019).

- [128] Nguyen, A. & Vu, H. Testing popular news discourse on the echo chamber effect: Does political polarisation occur among those relying on social media as their primary politics news source? *First Monday* **24** (2019).
- [129] Beam, M., Hutchens, M. & Hmielowski, J. Facebook news and (de)polarization: reinforcing spirals in the 2016 US election. *Information Communication and Society* **21**, 940–958 (2018).
- [130] Levy, R. Social media, news consumption, and polarization: Evidence from a field experiment. *American economic review* **111**, 831–70 (2021).
- [131] Carrella, F. #Populism on twitter: Statistical analysis of the correlation between tweet popularity and populist discursive features. *Brno Studies in English* **46**, 5–23 (2020).
- [132] Serrano, J., Shahrezaye, M., Papakyriakopoulos, O. & Hegelich, S. The rise of Germany's AfD: A social media analysis. 214–223 (2019).
- [133] Schumann, S., Boer, D., Hanke, K. & Liu, J. Social media use and support for populist radical right parties: assessing exposure and selection effects in a two-wave panel study. *Information Communication and Society* (2019).
- [134] Schumann, S., Thomas, F., Ehrke, F., Bertlich, T. & Dupont, J. C. Maintenance or change? Examining the reinforcing spiral between social media news use and populist attitudes. *Information, Communication & Society* 1–18 (2021).
- [135] Heiss, R. & Matthes, J. Stuck in a Nativist Spiral: Content, Selection, and Effects of Right-Wing Populists Communication on Facebook. *Political Communication* **37**, 303–328 (2020).
- [136] Bliuc, A.-M. *et al.* The effects of local socio-political events on group cohesion in online far-right communities. *PLoS ONE* **15** (2020).
- [137] Schulze, H. Who uses right-wing alternative online media? An exploration of audience characteristics. *Politics and Governance* **8**, 6–18 (2020).
- [138] Boulianne, S., Koc-Michalska, K. & Bimber, B. Right-wing populism, social media and echo chambers in Western democracies. *New Media and Society* **22**, 683–699 (2020).
- [139] Jeroense, T., Luimers, J., Jacobs, K. & Spierings, N. Political social media use and its linkage to populist and postmaterialist attitudes and vote intention in the Netherlands. *European Political Science* (2021).
- [140] Bosilkov, I. Media populism in Macedonia: Right-wing populist style in the coverage of the migrant crisis. *Central European Journal of Communication* **12**, 206–223 (2019).
- [141] Seebruck, R. Technology and Tolerance in Japan: Internet Use and Positive Attitudes and Behaviors Toward Foreigners. *Social Science Japan Journal* **16**, 279–300 (2013).
- [142] DOROSHENKO, L. Far-Right Parties in the European Union and Media Populism: A Comparative Analysis of 10 Countries During European Parliament Elections. *International Journal of Communication* **12**, 3186–3206 (2018).

- [143] Fletcher, R. & Nielsen, R. Are people incidentally exposed to news on social media? A comparative analysis. *New Media and Society* **20**, 2450–2468 (2018).
- [144] Guess, A. M. (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets. *American Journal of Political Science* (2021).
- [145] Strauss, N., Huber, B. & Gil de Zuniga, H. "Yes, I Saw It - But Didn't Read It ..." A Cross-Country Study, Exploring Relationships between Incidental News Exposure and News Use across Platforms. *Digital Journalism* **8**, 1181–1205 (2020).
- [146] Yang, T., Majó-Vázquez, S., Nielsen, R. & González-Bailón, S. Exposure to news grows less fragmented with an increase in mobile access. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 28678–28683 (2020).
- [147] Rivero, G. Preaching to the choir: ideology and following behaviour in social media. *Contemporary Social Science* **14**, 54–70 (2019).
- [148] Cinelli, M., de Francisci Morales, G., Galeazzi, A., Quattrociocchi, W. & Starnini, M. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences of the United States of America* **118** (2021).
- [149] Cota, W., Ferreira, S., Pastor-Satorras, R. & Starnini, M. Quantifying echo chamber effects in information spreading over political communication networks. *EPJ Data Science* **8** (2019).
- [150] Guerrero-Solá, F. & Lopez-Gonzalez, H. Government Formation and Political Discussions in Twitter: An Extended Model for Quantifying Political Distances in Multiparty Democracies. *Social Science Computer Review* **37**, 3–21 (2019).
- [151] Koiranen, I., Koivula, A., Keipi, T. & Saarinen, A. Shared contexts, shared background, shared values - Homophily in Finnish parliament members social networks on Twitter. *Telematics and Informatics* **36**, 117–131 (2019).
- [152] Barnidge, M., Macafee, T., Alvarez, G. & Rojas, H. Citizenship and Political Participation in Colombia: How Orientations Toward Citizenship Associate with Political and Civic Behaviors. *International Journal of Communication* **8**, 1831–1850 (2014).
- [153] Soral, W., Liu, J. & Bilewicz, M. Media of contempt: Social media consumption predicts normative acceptance of anti-muslim hate speech and islamoprejudice. *International Journal of Conflict and Violence* **14**, 1–13 (2020).
- [154] Tornberg, A. & Wahlstrom, M. Unveiling the radical right online: Exploring framing and identity in an online anti-immigrant discussion group. *Sociologisk Forskning* **55**, 267–292 (2018).
- [155] Rathje, S., Van Bavel, J. J. & van der Linden, S. Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences* **118** (2021). Publisher: National Acad Sciences.
- [156] Kubin, E. & von Sikorski, C. The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association* **45**, 188–206 (2021).

- [157] Pasquetto, I. V. *et al.* Tackling misinformation: What researchers could do with social media data. *The Harvard Kennedy School Misinformation Review* (2020).
- [158] Moher, D. *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine* **6**, e1000097 (2009). Publisher: Public Library of Science.
- [159] Gusenbauer, M. & Haddaway, N. R. Which academic search systems are suitable for systematic reviews or meta-analyses? evaluating retrieval qualities of google scholar, pubmed, and 26 other resources. *Research synthesis methods* **11**, 181–217 (2020).
- [160] Park, I. & Lee, D. Understanding news-sharing across different platforms: the effects of newsworthiness and gratifications from news-sharing. *Information Research-an International Electronic Journal* **25**, 882 (2020). WOS:000611778100045.
- [161] Machackova, H. & Tkaczyk, M. THE EFFECT OF MEDIA AND POLITICAL BELIEFS AND ATTITUDES ON TRUST IN POLITICAL INSTITUTIONS: A MULTILEVEL ANALYSIS ON DATA FROM 21 EUROPEAN COUNTRIES. *Communication Today* **11**, 64–82 (2020). WOS:000593125500006.
- [162] Magalhaes, P. C., Aldrich, J. H. & Gibson, R. K. New forms of mobilization, new people mobilized? Evidence from the Comparative Study of Electoral Systems. *Party Politics* **26**, 605–618 (2020). WOS:000556264500007.
- [163] Kwak, N., Lane, D. S., Zhu, Q., Lee, S. S. & Weeks, B. E. Political Rumor Communication on Instant Messaging Platforms: Relationships With Political Participation and Knowledge. *International Journal of Communication* **14**, 5663–5685 (2020). WOS:000616658300200.
- [164] Powers, E., Koliska, M. & Guha, P. "Shouting Matches and Echo Chambers": Perceived Identity Threats and Political Self-Censorship on Social Media. *International Journal of Communication* **13**, 3630–3649 (2019). WOS:000483935600005.
- [165] Adegbola, O. & Gearhart, S. Examining the Relationship Between Media Use and Political Engagement: A Comparative Study Among the United States, Kenya, and Nigeria. *International Journal of Communication* **13**, 1231–1251 (2019). WOS:000475772800037.
- [166] Liu, Y.-I. Online and Offline Communication and Political Knowledge and Participation in Presidential Campaigns: Effects of Geographical Context. *International Journal of Communication* **13**, 1438–1461 (2019). WOS:000475772800051.
- [167] Taneja, H. & Yaeger, K. *Do People Consume the News they Trust? Incidental News Usage and the High-Choice Media Environment* (2019). Publication Title: Chi 2019: Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems WOS:000474467906076.
- [168] Goebel, S. & Munzert, S. Political Advertising on the Wikipedia Marketplace of Information. *Social Science Computer Review* **36**, 157–175 (2018). WOS:000429943900002.

- [169] Wen, N. & Wei, R. Examining Effects of Informational Use of Social Media Platforms and Social Capital on Civic Engagement Regarding Genetically Modified Foods in China. *International Journal of Communication* **12**, 3729–3750 (2018). WOS:000457602100111.
- [170] Quenette, A. M. & Velasquez, A. Shifting Demographics: Understanding How Ethnically Diverse Networks Influence Latinos' Political Uses of Social Media and Offline Political Engagement. *International Journal of Communication* **12**, 4839–4859 (2018). WOS:000457602100175.
- [171] Zhang, N. & Skoric, M. M. Media Use and Environmental Engagement: Examining Differential Gains from News Media and Social Media. *International Journal of Communication* **12**, 380–403 (2018). WOS:000431821400025.
- [172] Mustapha, L. K., Gbonegun, V. O. & Mustapha, M. L. Social Media Use, Social Capital, and Political Participation among Nigerian University Students. *Tripodos* 127–143 (2016). WOS:000395301500008.
- [173] Barredo Ibanez, D., Arcila Calderon, C., Arroyave, J. & Silva, R. Influence of Social Networks in the Decision to Vote: An Exploratory Survey on the Ecuadorian Electorate. *International Journal of E-Politics* **6**, 15–34 (2015). WOS:000446494000003.
- [174] Kim, Y. & Chen, H.-T. Discussion Network Heterogeneity Matters: Examining a Moderated Mediation Model of Social Media Use and Civic Engagement. *International Journal of Communication* **9**, 2344–2365 (2015). WOS:000361977300104.
- [175] Krolo, K. & Puzek, I. Usage of Internet Social Networks and Participatory Dimensions of Social Capital of Youth - The Example of Facebook. *Drustvena Istrazivanja* **23**, 383–405 (2014). WOS:000345812600001.
- [176] de Zuniga, H. G. & Valenzuela, S. The Mediating Path to a Stronger Citizenship: Online and Offline Networks, Weak Ties, and Civic Engagement. *Communication Research* **38**, 397–421 (2011). WOS:000290301400005.
- [177] Popova, O. & Negrov, E. Political communication of youth in the internet space: Effects on influence on political consciousness and behavior. vol. 2813, 181–195 (2021). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85101631373&partnerID=40&md5=dfe78af08271e9b58aece5019ee35f28>.
- [178] Panizo-LLedot, A., Torregrosa, J., Bello-Orgaz, G., Thorburn, J. & Camacho, D. *Describing Alt-Right Communities and Their Discourse on Twitter During the 2018 US Mid-term Elections*, vol. 882 *SCI of Studies in Computational Intelligence* (2020). URL [https://www.scopus.com/inward/record.uri?eid=2-s2.0-85086570383&doi=10.1007%2f978-3-030-36683-4\\_35&partnerID=40&md5=d669494d971cbd00c64cec8aaebd021c](https://www.scopus.com/inward/record.uri?eid=2-s2.0-85086570383&doi=10.1007%2f978-3-030-36683-4_35&partnerID=40&md5=d669494d971cbd00c64cec8aaebd021c). Pages: 439.
- [179] Serhan, F. & Elareshi, M. New media and hate speech: A study of university students in Jordan. *Opcion* **36**, 166–184 (2020). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85084997399&partnerID=40&md5=f09e3d9075f29f1ac39eee7a3adea979>.
- [180] Riikonen, R., Huhtinen, A.-M. & Norri-Sederholm, T. Not a problem for me: Young mens conceptions of their social media use and false information. 240–245 (2020). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85084997399&partnerID=40&md5=f09e3d9075f29f1ac39eee7a3adea979>.

- [//www.scopus.com/inward/record.uri?eid=2-s2.0-85097756848&doi=10.34190%2fESM.20.031&partnerID=40&md5=debc7038da814e1795cf153561edf5a3](https://www.scopus.com/inward/record.uri?eid=2-s2.0-85097756848&doi=10.34190%2fESM.20.031&partnerID=40&md5=debc7038da814e1795cf153561edf5a3).
- [181] Ida, R., Saud, M. & Mashud, M. Persistence of social media on political activism and engagement among Indonesian and Pakistani youths. *International Journal of Web Based Communities* **16**, 321–342 (2020). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85083581686&doi=10.1504%2fijwbc.2020.111377&partnerID=40&md5=d9c679273d1558fc67107d1dd0406c81>.
- [182] KIM, H., KIM, Y. & LEE, D. Understanding the Role of Social Media in Political Participation: Integrating Political Knowledge and Bridging Social Capital From the Social Cognitive Approach. *International Journal of Communication* **14**, 4803–4824 (2020). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85099605102&partnerID=40&md5=ff61266c2ac800c3c940006d98359a9d>.
- [183] Steffan, D. Visual Self-Presentation Strategies of Political Candidates on Social Media Platforms: A Comparative Study. *International Journal of Communication* **14**, 3096–3118 (2020). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85099439857&partnerID=40&md5=08a0c64300c9b63ab205da0d41adc44d>.
- [184] Supovitz, J., Kolouch, C. & Daly, A. The social psychology of homophily: The collective sentiments of education advocacy groups. *Teachers College Record* **122**, 49–66 (2020). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85091826465&partnerID=40&md5=8911eeb637fcd2b7e198fba0a9faa7eb>.
- [185] Zannettou, S., Finkelstein, J., Bradlyn, B. & Blackburn, J. A quantitative approach to understanding on-line antisemitism. 786–797 (2020). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85088394577&partnerID=40&md5=745a787180358f6f711088e19aa12181>.
- [186] Halpern, D., Valenzuela, S., Katz, J. & Miranda, J. *From Belief in Conspiracy Theories to Trust in Others: Which Factors Influence Exposure, Believing and Sharing Fake News*, vol. 11578 LNCS of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2019). URL [https://www.scopus.com/inward/record.uri?eid=2-s2.0-85069833064&doi=10.1007%2f978-3-030-21902-4\\_16&partnerID=40&md5=3bf3b8bc0fcff702ce77fb07fea63db6](https://www.scopus.com/inward/record.uri?eid=2-s2.0-85069833064&doi=10.1007%2f978-3-030-21902-4_16&partnerID=40&md5=3bf3b8bc0fcff702ce77fb07fea63db6). Pages: 232.
- [187] Jones, M. Propaganda, fake news, and fake trends: The weaponization of Twitter bots in the Gulf crisis. *International Journal of Communication* **13**, 1389–1415 (2019). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85068128178&partnerID=40&md5=b6736e5c213f839d390d50131fa03cf4>.
- [188] Jiang, S., Robertson, R. & Wilson, C. Bias misperceived: The role of partisanship and misinformation in YouTube comment moderation. 278–289 (2019). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85070399362&partnerID=40&md5=2d70fd6bbdf3f70d6f91743032db1880>.
- [189] Quintanilha, T., Da Silva, M. & Lapa, T. Fake news and its impact on trust in the news. Using the portuguese case to establish lines of differentiation. *Communication and Society* **32**, 17–



- 33 (2019). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85069803365&doi=10.15581%2f003.32.3.17-33&partnerID=40&md5=1e3f6959b3cb64e2353a5d570bd0b0f9>.
- [190] Emamjomehzadeh, S., Masoudnia, H. & Rahbarqazi, M. The role of online social media in peoples political orientations and tendency to participate in presidential elections. *Teorija in Praksa* **55**, 666–683 (2018). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85055971280&partnerID=40&md5=86333c6dd745e8560bbcd26ce70355e4>.
- [191] Bodrunova, S., Blekanov, I. & Kukarkin, M. *Multi-dimensional echo chambers: Language and sentiment structure of Twitter discussions on the Charlie Hebdo case*, vol. 850 of *Communications in Computer and Information Science* (2018). URL [https://www.scopus.com/inward/record.uri?eid=2-s2.0-85055822021&doi=10.1007%2f978-3-319-92270-6\\_56&partnerID=40&md5=a3951716f2d40f1ef2270ca6713385ee](https://www.scopus.com/inward/record.uri?eid=2-s2.0-85055822021&doi=10.1007%2f978-3-319-92270-6_56&partnerID=40&md5=a3951716f2d40f1ef2270ca6713385ee). Pages: 400.
- [192] Mohamad, B., Dauda, S. & Halim, H. Youth offline political participation: Trends and role of social media. *Jurnal Komunikasi: Malaysian Journal of Communication* **34**, 172–192 (2018). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85056571382&doi=10.17576%2fJKMJC-2018-3403-11&partnerID=40&md5=2fc3b41313a00c22cf48b0ee980329b4>.
- [193] Lee, S. The Role of Social Media in Protest Participation: The Case of Candlelight Vigils in South Korea. *International Journal of Communication* **12**, 1523–1540 (2018). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85061390066&partnerID=40&md5=0909eee742426af3d5e14ade58330961>.
- [194] Herrero-Jiménez, B., Carratalá, A. & Berganza, R. Violent conflicts and the new mediatization: The impact of social media on the European parliamentary agenda regarding the Syrian war. *Communication and Society* **31**, 141–157 (2018). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85050497356&doi=10.15581%2f003.31.3.141-157&partnerID=40&md5=cbc8363444ea0ccbd3bdfff5e5df6791>.
- [195] Xenos, M. *et al.* News media use and the informed public in the digital age. *International Journal of Communication* **12**, 706–724 (2018). URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85041596617&partnerID=40&md5=cb45b2bb28d5937be8998211847e6b09>.
- [196] Garimella, K., Smith, T., Weiss, R. & West, R. Political Polarization in Online News Consumption. *arXiv:2104.06481 [cs]* (2021). URL <http://arxiv.org/abs/2104.06481>. ArXiv: 2104.06481.
- [197] Pagoto, L. & Longhi, R. Platformization, techno populism and disintermediation of sources in attacks on journalism on Instagram. *CHASQUI-REVISTA LATINOAMERICANA DE COMUNICACION* 181–200 (2021).
- [198] de Zuniga, H., Borah, P. & Goyanes, M. How do people learn about politics when inadvertently exposed to news? Incidental news paradoxical Direct and indirect effects on political knowledge. *COMPUTERS IN HUMAN BEHAVIOR* **121** (2021).
- [199] Olise, F. Level of Acceptance of News Stories on Social Media Platforms Among Youth in Nigeria. *JURNAL KOMUNIKASI-MALAYSIAN JOURNAL OF COMMUNICATION* **37**, 210–225 (2021).

- [200] Santini, R., Salles, D. & Tucci, G. When Machine Behavior Targets Future Voters: The Use of Social Bots to Test Narratives for Political Campaigns in Brazil. *INTERNATIONAL JOURNAL OF COMMUNICATION* **15**, 1220–1243 (2021).
- [201] Choi, J. Cross-Cutting Scanning, Integrating, and Interacting: Dimensions of Cross-Cutting Exposure on Social Media and Political Participation. *INTERNATIONAL JOURNAL OF COMMUNICATION* **15**, 1595–1616 (2021).
- [202] Gerosa, T., Gui, M., Hargittai, E. & Nguyen, M. (Mis)informed During COVID-19: How Education Level and Information Sources Contribute to Knowledge Gaps. *INTERNATIONAL JOURNAL OF COMMUNICATION* **15**, 2196–2217 (2021).
- [203] Villanueva, C. & Toscano, G. Legitimation of hate and political violence through memetic images: the Bolsonaro campaign. *COMMUNICATION & SOCIETY-SPAIN* **34**, 449–468 (2021).
- [204] Ahmed, M., Riaz, M., Qamar, M. & Asghar, R. Fake News Shared on WhatsApp During Covid-19: An Analysis of Groups and Statuses in Pakistan. *MEDIA EDUCATION-MEDIAOBRAZOVANIE* 4–17 (2021).
- [205] Gorski, L. C. & Thomas, F. Staying Tuned or Tuning Out? A Longitudinal Analysis of News-Avoiders on the Micro and Macro-Level. *Communication Research* 009365022110259 (2021). URL <http://journals.sagepub.com/doi/10.1177/00936502211025907>.
- [206] Enders, A. M. *et al.* The Relationship Between Social Media Use and Beliefs in Conspiracy Theories and Misinformation. *Political Behavior* (2021). URL <https://link.springer.com/10.1007/s11109-021-09734-6>.
- [207] Müller, P. & Bach, R. L. Populist alternative news use and its role for elections: Web-tracking and survey evidence from two campaign periods. *New Media & Society* 146144482110329 (2021). URL <http://journals.sagepub.com/doi/10.1177/14614448211032969>.
- [208] Ozeren, S., Cubukcu, S. & Cash, G. Exposure to Extremist Content and Public Sympathy for ISIS. *Studies in Conflict & Terrorism* 1–21 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/1057610X.2021.1965728>.
- [209] Hashemi, M. Discovering social media topics and patterns in the coronavirus and election era. *Journal of Information, Communication and Ethics in Society* **20**, 1–17 (2022). URL <https://www.emerald.com/insight/content/doi/10.1108/JICES-04-2021-0039/full/html>.
- [210] Wolfowicz, M., Weisburd, D. & Hasisi, B. Examining the interactive effects of the filter bubble and the echo chamber on radicalization. *Journal of Experimental Criminology* (2021). URL <https://link.springer.com/10.1007/s11292-021-09471-0>.
- [211] Yamamoto, M. & Yang, F. Does news help us become knowledgeable or think we are knowledgeable? Examining a linkage of traditional and social media use with political knowledge. *Journal of Information Technology & Politics* **19**, 269–283 (2022). URL <https://www.tandfonline.com/doi/full/10.1080/19331681.2021.1969611>.

- [212] Goyanes, M., Borah, P. & Gil de Zúñiga, H. Social media filtering and democracy: Effects of social media news use and uncivil political discussions on social media unfriending. *Computers in Human Behavior* **120**, 106759 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0747563221000819>.
- [213] Choli, M. & Kuss, D. J. Perceptions of blame on social media during the coronavirus pandemic. *Computers in Human Behavior* **124**, 106895 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0747563221002181>.
- [214] Cano-Orón, L., Calvo, D., Llorca-Abad, G. & Mestre-PÃ©rez, R. Media crisis and disinformation: the participation of digital newspapers in the dissemination of a denialist hoax. *El Profesional de la informaci3n* e300412 (2021). URL <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/86394/version/4386>.
- [215] Kim, B., Cooks, E. & Kim, S.-K. Exploring incivility and moral foundations toward Asians in English-speaking tweets in hate crime-reporting cities during the COVID-19 pandemic. *Internet Research* **32**, 362–378 (2022). URL <https://www.emerald.com/insight/content/doi/10.1108/INTR-11-2020-0678/full/html>.
- [216] Melki, J. *et al.* Mitigating infodemics: The relationship between news exposure and trust and belief in COVID-19 fake news and social media spreading. *PLOS ONE* **16**, e0252830 (2021). URL <https://dx.plos.org/10.1371/journal.pone.0252830>.
- [217] Martin, J. D. & Hassan, F. Testing Classical Predictors of Public Willingness to Censor on the Desire to Block Fake News Online. *Convergence: The International Journal of Research into New Media Technologies* **28**, 867–887 (2022). URL <http://journals.sagepub.com/doi/10.1177/13548565211012552>.
- [218] Bermes, A. Information overload and fake news sharing: A transactional stress perspective exploring the mitigating role of consumers resilience during COVID-19. *Journal of Retailing and Consumer Services* **61**, 102555 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0969698921001211>.
- [219] Shin, S. H., Ji, H. & Lim, H. Heterogeneity in preventive behaviors during COVID-19: Health risk, economic insecurity, and slanted information. *Social Science & Medicine* **278**, 113944 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0277953621002768>.
- [220] Nah, S., Lee, S. & Liu, W. Community Storytelling Network, Expressive Digital Media Use, and Civic Engagement. *Communication Research* **49**, 327–352 (2022). URL <http://journals.sagepub.com/doi/10.1177/00936502211019677>.
- [221] Dias, T., von Bülöw, M. & Gobbi, D. Populist Framing Mechanisms and the Rise of Right-wing Activism in Brazil. *Latin American Politics and Society* **63**, 69–92 (2021). URL [https://www.cambridge.org/core/product/identifier/S1531426X21000224/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S1531426X21000224/type/journal_article).
- [222] SIHOMBING, S. O. & PRAMONO, R. The Integration of Social Media to the Theory of Planned Behavior: A Case Study in Indonesia. *The Journal of Asian Finance, Economics and Business* **8**, 445–454 (2021). URL <https://doi.org/10.13106/JAFEB.2021.VOL8.NO5.0445>.

- [223] Rao, A. *et al.* Political Partisanship and Antiscience Attitudes in Online Discussions About COVID-19: Twitter Content Analysis. *Journal of Medical Internet Research* **23**, e26692 (2021). URL <https://www.jmir.org/2021/6/e26692>.
- [224] Visentin, M., Tuan, A. & Di Domenico, G. Words matter: How privacy concerns and conspiracy theories spread on twitter. *Psychology & Marketing* **38**, 1828–1846 (2021). URL <https://onlinelibrary.wiley.com/doi/10.1002/mar.21542>.
- [225] Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A. & Petersen, M. B. Partisan Polarization Is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter. *American Political Science Review* **115**, 999–1015 (2021). URL [https://www.cambridge.org/core/product/identifier/S0003055421000290/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S0003055421000290/type/journal_article).
- [226] Herrera-Peco, I. *et al.* Antivaccine Movement and COVID-19 Negationism: A Content Analysis of Spanish-Written Messages on Twitter. *Vaccines* **9**, 656 (2021). URL <https://www.mdpi.com/2076-393X/9/6/656>.
- [227] Nazar, S. & Pieters, T. Plandemic Revisited: A Product of Planned Disinformation Amplifying the COVID-19 infodemic. *Frontiers in Public Health* **9**, 649930 (2021). URL <https://www.frontiersin.org/articles/10.3389/fpubh.2021.649930/full>.
- [228] Hollewell, G. F. & Longpré, N. Radicalization in the Social Media Era: Understanding the Relationship between Self-Radicalization and the Internet. *International Journal of Offender Therapy and Comparative Criminology* **66**, 896–913 (2022). URL <http://journals.sagepub.com/doi/10.1177/0306624X211028771>.
- [229] Tal-Or, N., Cohen, J., Tsfati, Y. & Gunther, A. C. Testing Causal Direction in the Influence of Presumed Media Influence. *Communication Research* **37**, 801–824 (2010). URL <http://journals.sagepub.com/doi/10.1177/0093650210362684>.
- [230] Weeks, B. E., Menchen-Trevino, E., Calabrese, C., Casas, A. & Wojcieszak, M. Partisan media, untrustworthy news sites, and political misperceptions. *New Media & Society* 146144482110333 (2021). URL <http://journals.sagepub.com/doi/10.1177/14614448211033300>.
- [231] Gil de Zúñiga, H., Barnidge, M. & Diehl, T. Political persuasion on social media: A moderated moderation model of political discussion disagreement and civil reasoning. *The Information Society* **34**, 302–315 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/01972243.2018.1497743>.
- [232] Chayinska, M., Miranda, D. & González, R. A longitudinal study of the bidirectional causal relationships between online political participation and offline collective action. *Computers in Human Behavior* **121**, 106810 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0747563221001333>.
- [233] Criss, S. *et al.* Advocacy, Hesitancy, and Equity: Exploring U.S. Race-Related Discussions of the COVID-19 Vaccine on Twitter. *International Journal of Environmental Research and Public Health* **18**, 5693 (2021). URL <https://www.mdpi.com/1660-4601/18/11/5693>.

- [234] Onat, I., Guler, A., Kula, S. & Bastug, M. F. Fear of Terrorism and Fear of Violent Crimes in the United States: A Comparative Analysis. *Crime & Delinquency* 001112872110361 (2021). URL <http://journals.sagepub.com/doi/10.1177/00111287211036130>.
- [235] Chen, H.-T. Second Screening and the Engaged Public: The Role of Second Screening for News and Political Expression in an O-S-R-O-R Model. *Journalism & Mass Communication Quarterly* **98**, 526–546 (2021). URL <http://journals.sagepub.com/doi/10.1177/1077699019866432>.
- [236] Lin, W.-Y., Hope Cheong, P., Kim, Y.-C. & Jung, J.-Y. Becoming Citizens: Youths Civic Uses of New Media in Five Digital Cities in East Asia. *Journal of Adolescent Research* **25**, 839–857 (2010). URL <http://journals.sagepub.com/doi/10.1177/0743558410371125>.
- [237] Young, L. E. Mobilization Under Threat: Emotional Appeals and Pro-Opposition Political Participation Online. *Political Behavior* (2021). URL <https://link.springer.com/10.1007/s11109-021-09711-z>.
- [238] Wang, D. & Qian, Y. Echo Chamber Effect in Rumor Rebuttal Discussions About COVID-19 in China: Social Media Content and Network Analysis Study. *Journal of Medical Internet Research* **23**, e27009 (2021). URL <https://www.jmir.org/2021/3/e27009>.
- [239] Diehl, T., Huber, B., Gil de Zúñiga, H. & Liu, J. Social Media and Beliefs about Climate Change: A Cross-National Analysis of News Use, Political Ideology, and Trust in Science. *International Journal of Public Opinion Research* **33**, 197–213 (2021). URL <https://academic.oup.com/ijpor/article/33/2/197/5628296>.
- [240] Biancovilli, P., Makszin, L. & Jurberg, C. Misinformation on social networks during the novel coronavirus pandemic: a quali-quantitative case study of Brazil. *BMC Public Health* **21**, 1200 (2021). URL <https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-021-11165-1>.
- [241] Apuke, O. D. & Omar, B. Social media affordances and information abundance: Enabling fake news sharing during the COVID-19 health crisis. *Health Informatics Journal* **27**, 146045822110214 (2021). URL <http://journals.sagepub.com/doi/10.1177/14604582211021470>.
- [242] PÃ©rez-Curiel, C., Rivas-de Roca, R. & GarcÃ­a-Gordillo, M. Impact of Trumps Digital Rhetoric on the US Elections: A View from Worldwide Far-Right Populism. *Social Sciences* **10**, 152 (2021). URL <https://www.mdpi.com/2076-0760/10/5/152>.
- [243] Elliott, T. & Earl, J. Online protest participation and the digital divide: Modeling the effect of the digital divide on online petition-signing. *New Media & Society* **20**, 698–719 (2018). URL <http://journals.sagepub.com/doi/10.1177/1461444816669159>.
- [244] Hawkins, I. & Saleem, M. Rise UP!: A content analytic study of how collective action is discussed within White nationalist videos on YouTube. *New Media & Society* 146144482110405 (2021). URL <http://journals.sagepub.com/doi/10.1177/14614448211040520>.

- [245] Gorodnichenko, Y., Pham, T. & Talavera, O. Social media, sentiment and public opinions: Evidence from #Brexit and #USElection. *European Economic Review* **136**, 103772 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0014292121001252>.
- [246] Chan, M., Chen, H.-T. & Lee, F. L. F. Cross-Cutting Discussion on Social Media and Online Political Participation: A Cross-National Examination of Information Seeking and Social Accountability Explanations. *Social Media + Society* **7**, 205630512110356 (2021). URL <http://journals.sagepub.com/doi/10.1177/20563051211035697>.
- [247] Mari, S. *et al.* Conspiracy Theories and Institutional Trust: Examining the Role of Uncertainty Avoidance and Active Social Media Use. *Political Psychology* **43**, 277–296 (2022). URL <https://onlinelibrary.wiley.com/doi/10.1111/pops.12754>.
- [248] Kopacheva, E. How the Internet has changed participation: Exploring distinctive preconditions of online activism. *Communication & Society* 67–85 (2021). URL <https://revistas.unav.edu/index.php/communication-and-society/article/view/40019>.
- [249] Gavazza, A., Nardotto, M. & Valletti, T. Internet and Politics: Evidence from U.K. Local Elections and Local Government Policies. *The Review of Economic Studies* **86**, 2092–2135 (2019). URL <https://academic.oup.com/restud/article/86/5/2092/5035337>.
- [250] De Coninck, D. *et al.* Beliefs in Conspiracy Theories and Misinformation About COVID-19: Comparative Perspectives on the Role of Anxiety, Depression and Exposure to and Trust in Information Sources. *Frontiers in Psychology* **12**, 646394 (2021). URL <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.646394/full>.
- [251] Wang, L. Race, social media news use, and political participation. *Journal of Information Technology & Politics* **19**, 83–97 (2022). URL <https://www.tandfonline.com/doi/full/10.1080/19331681.2021.1923497>.
- [252] Basch, C. E., Basch, C. H., Hillyer, G. C., Meleo-Erwin, Z. C. & Zagnit, E. A. YouTube Videos and Informed Decision-Making About COVID-19 Vaccination: Successive Sampling Study. *JMIR Public Health and Surveillance* **7**, e28352 (2021). URL <https://publichealth.jmir.org/2021/5/e28352>.
- [253] Weeks, B. E., Lane, D. S. & Hahn, L. B. Online Incidental Exposure to News Can Minimize Interest-Based Political Knowledge Gaps: Evidence from Two U.S. Elections. *The International Journal of Press/Politics* **27**, 243–262 (2022). URL <http://journals.sagepub.com/doi/10.1177/1940161221991550>.
- [254] Valenzuela, S., Halpern, D., Katz, J. E. & Miranda, J. P. The Paradox of Participation Versus Misinformation: Social Media, Political Engagement, and the Spread of Misinformation. *Digital Journalism* **7**, 802–823 (2019). URL <https://www.tandfonline.com/doi/full/10.1080/21670811.2019.1623701>.
- [255] Onuch, O., Mateo, E. & Waller, J. G. Mobilization, Mass Perceptions, and (Dis)information: New and Old Media Consumption Patterns and Protest. *Social Media + Society* **7**, 205630512199965 (2021). URL <http://journals.sagepub.com/doi/10.1177/2056305121999656>.

- [256] Hernandez, R. G., Hagen, L., Walker, K., OLeary, H. & Lengacher, C. The COVID-19 vaccine social media *infodemic* : healthcare providers missed dose in addressing misinformation and vaccine hesitancy. *Human Vaccines & Immunotherapeutics* **17**, 2962–2964 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/21645515.2021.1912551>.
- [257] Ruijgrok, K. Illusion of control: how internet use generates anti-regime sentiment in authoritarian regimes. *Contemporary Politics* **27**, 247–270 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/13569775.2020.1851931>.
- [258] Brandtzaeg, P. B. Facebook is no Great equalizer: A big data approach to gender differences in civic engagement across countries. *Social Science Computer Review* **35**, 103–125 (2017). URL <http://journals.sagepub.com/doi/10.1177/0894439315605806>.
- [259] Rodr guez-Virgili, J., Serrano-Puche, J. & Fern ndez, C. B. Digital Disinformation and Preventive Actions: Perceptions of Users from Argentina, Chile, and Spain. *Media and Communication* **9**, 323–337 (2021). URL <https://www.cogitatiopress.com/mediaandcommunication/article/view/3521>.
- [260] Suhay, E., Blackwell, A., Roche, C. & Bruggeman, L. Forging Bonds and Burning Bridges: Polarization and Incivility in Blog Discussions About Occupy Wall Street. *American Politics Research* **43**, 643–679 (2015). URL <http://journals.sagepub.com/doi/10.1177/1532673X14553834>.
- [261] Lilleker, D., Koc-Michalska, K. & Bimber, B. Women learn while men talk?: revisiting gender differences in political engagement in online environments. *Information, Communication & Society* **24**, 2037–2053 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/1369118X.2021.1961005>.
- [262] Schulz, W. Spiraleffekte in der neuen Medienwelt: W hlermobilisierung und die Nutzung politischer Online- und Offline-Information im Bundestagswahlkampf 2013. *Studies in Communication | Media* **8**, 77–114 (2019). URL <https://www.nomos-elibrary.de/index.php?doi=10.5771/2192-4007-2019-1-77>.
- [263] Tolbert, C. J. & McNeal, R. S. Unraveling the Effects of the Internet on Political Participation? *Political Research Quarterly* **56**, 175 (2003). URL <https://www.jstor.org/stable/3219896?origin=crossref>.
- [264] Valenzuela, S., Halpern, D. & Araneda, F. A Downward Spiral? A Panel Study of Misinformation and Media Trust in Chile. *The International Journal of Press/Politics* **27**, 353–373 (2022). URL <http://journals.sagepub.com/doi/10.1177/19401612211025238>.
- [265] Choi, J. & Lee, J. Enthusiasm toward the other side matters: Emotion and willingness to express disagreement in social media political conversation. *The Social Science Journal* 1–17 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/03623319.2021.1949548>.
- [266] Jennings, W. *et al.* Lack of Trust, Conspiracy Beliefs, and Social Media Use Predict COVID-19 Vaccine Hesitancy. *Vaccines* **9**, 593 (2021). URL <https://www.mdpi.com/2076-393X/9/6/593>.
- [267] Nagayoshi, K. The Political Orientation of Japanese Online Right-wingers. *Pacific Affairs* **94**, 5–32 (2021). URL <https://www.ingentaconnect.com/content/10.5509/20219415>.

- [268] Kahne, J. & Bowyer, B. The Political Significance of Social Media Activity and Social Networks. *Political Communication* **35**, 470–493 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/10584609.2018.1426662>.
- [269] Liu, W., Chen, N.-T. N., Ognyanova, K., Nah, S. & Ball-Rokeach, S. Connecting With Hyperlocal News Website: Cause or Effect of Civic Participation? *American Behavioral Scientist* **62**, 1022–1041 (2018). URL <http://journals.sagepub.com/doi/10.1177/0002764218764243>.
- [270] Housholder, E., Watson, B. R. & LoRusso, S. Does Political Advertising Lead to Online Information Seeking? A Real-World Test Using Google Search Data. *Journal of Broadcasting & Electronic Media* **62**, 337–353 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/08838151.2018.1451854>.
- [271] Shahin, S., Saldaña, M. & Gil de Zúñiga, H. Peripheral elaboration model: The impact of incidental news exposure on political participation. *Journal of Information Technology & Politics* **18**, 148–163 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/19331681.2020.1832012>.
- [272] Haenschen, K. & Jennings, J. Mobilizing Millennial Voters with Targeted Internet Advertisements: A Field Experiment. *Political Communication* **36**, 357–375 (2019). URL <https://www.tandfonline.com/doi/full/10.1080/10584609.2018.1548530>.
- [273] Sakya, S. M. *et al.* The impact of COVID-19-related changes in media consumption on public knowledge: results of a cross-sectional survey of Pennsylvania adults. *Current Medical Research and Opinion* **37**, 911–915 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/03007995.2021.1901679>.
- [274] Sridhar, D. & Getoor, L. Estimating Causal Effects of Tone in Online Debates. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 1872–1878 (International Joint Conferences on Artificial Intelligence Organization, Macao, China, 2019). URL <https://www.ijcai.org/proceedings/2019/259>.
- [275] Patra, R. K. & Pandey, N. Disinformation on Novel Coronavirus (COVID 19) A Content Analysis of News Published on Fact Checking Sites in India. *DESIDOC Journal of Library & Information Technology* **41**, 275–283 (2021). URL <https://publications.drdo.gov.in/ojs/index.php/djlit/article/view/16556>.
- [276] Shin, J. How Do Partisans Consume News on Social Media? A Comparison of Self-Reports With Digital Trace Measures Among Twitter Users. *Social Media + Society* **6**, 205630512098103 (2020). URL <http://journals.sagepub.com/doi/10.1177/2056305120981039>.
- [277] Praprotnik, K., Perlot, F., Ingruber, D. & Filzmaier, P. Soziale Medien als politischer Informationskanal (2019). URL <https://webapp.uibk.ac.at/ojs/index.php/OEZP/article/viewFile/2726/2291>. Publisher: Universitaet Innsbruck - Innsbruck University Press.
- [278] Bail, C. A. *et al.* Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017. *Proceedings of the National Academy of Sciences* **117**, 243–250 (2020). URL <https://pnas.org/doi/full/10.1073/pnas.1906420116>.



- [279] Batool, S. H., Ahmed, W., Mahmood, K. & Saeed, H. Twitter dialogue: an analysis of Pakistani politicians information sharing. *Information Discovery and Delivery* **50**, 64–74 (2022). URL <https://www.emerald.com/insight/content/doi/10.1108/IDD-03-2020-0031/full/html>.
- [280] Siongers, J., Keppens, G., Spruyt, B. & Van Droogenbroeck, F. On the Digital Lane to Citizenship? Patterns of Internet Use and Civic Engagement Amongst Flemish Adolescents and Young Adults. *JSSE - Journal of Social Science Education* Vol 18 No 2 (2019) (2019). URL <https://www.jsse.org/index.php/jsse/article/view/901>. Publisher: JSSE - Journal of Social Science Education.
- [281] Linvill, D. L., Boatwright, B. C., Grant, W. J. & Warren, P. L. THE RUSSIANS ARE HACKING MY BRAIN! investigating Russia's internet research agency twitter tactics during the 2016 United States presidential campaign. *Computers in Human Behavior* **99**, 292–300 (2019). URL <https://linkinghub.elsevier.com/retrieve/pii/S074756321930202X>.
- [282] Chan, M.-p. S. *et al.* Legacy and social media respectively influence risk perceptions and protective behaviors during emerging health threats: A multi-wave analysis of communications on Zika virus cases. *Social Science & Medicine* **212**, 50–59 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S0277953618303630>.
- [283] Koivula, A., Kaakinen, M., Oksanen, A. & Räsänen, P. The Role of Political Activity in the Formation of Online Identity Bubbles. *Policy & Internet* **11**, 396–417 (2019). URL <https://onlinelibrary.wiley.com/doi/10.1002/poi3.211>.
- [284] Kleinnijenhuis, J., van Hoof, A. M. J. & van Atteveldt, W. The Combined Effects of Mass Media and Social Media on Political Perceptions and Preferences. *Journal of Communication* **69**, 650–673 (2019). URL <https://academic.oup.com/joc/article/69/6/650/5681988>.
- [285] Bowman, W. M. & Bowman, J. D. Censorship or self-control? Hate speech, the state and the voter in the Kenyan election of 2013. *The Journal of Modern African Studies* **54**, 495–531 (2016). URL [https://www.cambridge.org/core/product/identifier/S0022278X16000380/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S0022278X16000380/type/journal_article).
- [286] Kim, H. H.-s. & Lim, C. From virtual space to public space: The role of online political activism in protest participation during the Arab Spring. *International Journal of Comparative Sociology* **60**, 409–434 (2019). URL <http://journals.sagepub.com/doi/10.1177/0020715219894724>.
- [287] Matuszewski, P. & Szabó, G. Are Echo Chambers Based on Partisanship? Twitter and Political Polarity in Poland and Hungary. *Social Media + Society* **5**, 205630511983767 (2019). URL <http://journals.sagepub.com/doi/10.1177/2056305119837671>.
- [288] Lee, F. L. F., Lee, P. S. N., So, C. Y., Leung, L. & Chan, M. C. Conditional impact of facebook as an information source on political opinions: the case of political reform in Hong Kong. *Asian Journal of Political Science* **25**, 365–382 (2017). URL <https://www.tandfonline.com/doi/full/10.1080/02185377.2017.1352523>.

- [289] Casas, A. & Williams, N. W. Images that Matter: Online Protests and the Mobilizing Role of Pictures. *Political Research Quarterly* **72**, 360–375 (2019). URL <http://journals.sagepub.com/doi/10.1177/1065912918786805>.
- [290] Xiong, J., Feng, X. & Tang, Z. Understanding user-to-User interaction on government microblogs: An exponential random graph model with the homophily and emotional effect. *Information Processing & Management* **57**, 102229 (2020). URL <https://linkinghub.elsevier.com/retrieve/pii/S0306457319310672>.
- [291] Nguyen, T. T. *et al.* Evaluating associations between area-level Twitter-expressed negative racial sentiment, hate crimes, and residents' racial prejudice in the United States. *SSM - Population Health* **13**, 100750 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S2352827321000252>.
- [292] Gondal, M. T., Munir, A., Shabir, G. & Naz, A. Facebook and Propaganda: Following Politics on Facebook and its Impact on Political Behaviors of Youth. *Clinical Social Work and Health Intervention* **10**, 27–33 (2019). URL <https://clinicalsocialwork.eu/wp-content/uploads/2019/11/04-Gondal-1.pdf>.
- [293] Zang, L., Xiong, F. & Gao, Y. Reversing the U: New Evidence on the Internet and Democracy Relationship. *Social Science Computer Review* **37**, 295–314 (2019). URL <http://journals.sagepub.com/doi/10.1177/0894439318767957>.
- [294] Ohme, J. Updating citizenship? The effects of digital media use on citizenship understanding and political participation. *Information, Communication & Society* **22**, 1903–1928 (2019). URL <https://www.tandfonline.com/doi/full/10.1080/1369118X.2018.1469657>.
- [295] Corbu, N., Oprea, D.-A., Negrea-Busuioac, E. & Radu, L. They cant fool me, but they can fool the others! Third person effect and fake news detection. *European Journal of Communication* **35**, 165–180 (2020). URL <http://journals.sagepub.com/doi/10.1177/0267323120903686>.
- [296] Germann, M. & Gemenis, K. Getting Out the Vote With Voting Advice Applications. *Political Communication* **36**, 149–170 (2019). URL <https://www.tandfonline.com/doi/full/10.1080/10584609.2018.1526237>.
- [297] Neely, S., Eldredge, C. & Sanders, R. Health Information Seeking Behaviors on Social Media During the COVID-19 Pandemic Among American Social Networking Site Users: Survey Study. *Journal of Medical Internet Research* **23**, e29802 (2021). URL <https://www.jmir.org/2021/6/e29802>.
- [298] Inguanzo, I., Zhang, B. & Gil de Zúñiga, H. Online cultural backlash? sexism and political user-generated content. *Information, Communication & Society* **24**, 2133–2152 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/1369118X.2021.1962940>.
- [299] Chekol, M. A., Moges, M. A. & Nigatu, B. A. Social media hate speech in the walk of Ethiopian political reform: analysis of hate speech prevalence, severity, and natures. *Information, Communication & Society* 1–20 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/1369118X.2021.1942955>.

- [300] Valenzuela, S., Bachmann, I. & Bargsted, M. The Personal Is the Political? What Do WhatsApp Users Share and How It Matters for News Knowledge, Polarization and Participation in Chile. *Digital Journalism* **9**, 155–175 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/21670811.2019.1693904>.
- [301] Van Duyn, E., Peacock, C. & Stroud, N. J. The Gender Gap in Online News Comment Sections. *Social Science Computer Review* **39**, 181–196 (2021). URL <http://journals.sagepub.com/doi/10.1177/0894439319864876>.
- [302] Wiedlitzka, S., Prati, G., Brown, R., Smith, J. & Walters, M. A. Hate in Word and Deed: The Temporal Association Between Online and Offline Islamophobia. *Journal of Quantitative Criminology* (2021). URL <https://link.springer.com/10.1007/s10940-021-09530-9>.
- [303] Oh, H. J., Lor, Z. & Choi, J. News Repertoires and Political Information Efficacy: Focusing on the Mediating Role of Perceived News Overload. *SAGE Open* **11**, 215824402098868 (2021). URL <http://journals.sagepub.com/doi/10.1177/2158244020988685>.
- [304] Siegel, A. A. *et al.* Trumping Hate on Twitter? Online Hate Speech in the 2016 U.S. Election Campaign and its Aftermath. *Quarterly Journal of Political Science* **16**, 71–104 (2021). URL <http://www.nowpublishers.com/article/Details/QJPS-19045>.
- [305] Kruikeemeier, S. How political candidates use Twitter and the impact on votes. *Computers in Human Behavior* **34**, 131–139 (2014). URL <https://linkinghub.elsevier.com/retrieve/pii/S0747563214000302>.
- [306] van Erkel, P. F. A. & Van Aelst, P. Why Dont We Learn from Social Media? Studying Effects of and Mechanisms behind Social Media News Use on General Surveillance Political Knowledge. *Political Communication* **38**, 407–425 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/10584609.2020.1784328>.
- [307] Chan, N. K. Political Inequality in the Digital World: The Puzzle of Asian American Political Participation Online. *Political Research Quarterly* **74**, 882–898 (2021). URL <http://journals.sagepub.com/doi/10.1177/1065912920945391>.
- [308] Wang, X. & Kobayashi, T. Nationalism and political system justification in China: differential effects of traditional and new media. *Chinese Journal of Communication* **14**, 139–156 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/17544750.2020.1807372>.
- [309] Freudenthaler, R. & Wessler, H. Mapping Emerging and Legacy Outlets Online by Their Democratic Functions—Agonistic, Deliberative, or Corrosive? *The International Journal of Press/Politics* **27**, 417–438 (2022). URL <http://journals.sagepub.com/doi/10.1177/19401612211015077>.
- [310] Cardenal, A. S., Galais, C. & Majó-Vázquez, S. Is Facebook Eroding the Public Agenda? Evidence From Survey and Web-Tracking Data. *International Journal of Public Opinion Research* (2018). URL <https://academic.oup.com/ijpor/advance-article/doi/10.1093/ijpor/edy025/5250416>.

- [311] Zhou, C., Xiu, H., Wang, Y. & Yu, X. Characterizing the dissemination of misinformation on social media in health emergencies: An empirical study based on COVID-19. *Information Processing & Management* **58**, 102554 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0306457321000583>.
- [312] Hajj, N., McEwan, P. J. & Turkington, R. Women, information ecology, and political protest in the Middle East. *Mediterranean Politics* **24**, 62–83 (2019). URL <https://www.tandfonline.com/doi/full/10.1080/13629395.2017.1380116>.
- [313] Zhu, A. Y. F., Chan, A. L. S. & Chou, K. L. Creative social media use and political participation in young people: The moderation and mediation role of online political expression. *Journal of Adolescence* **77**, 108–117 (2019). URL <https://onlinelibrary.wiley.com/doi/10.1016/j.adolescence.2019.10.010>.
- [314] Towner, T. L. & Muñoz, C. L. Baby Boom or Bust? the New Media Effect on Political Participation. *Journal of Political Marketing* **17**, 32–61 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/15377857.2016.1153561>.
- [315] Licari, P. R. Sharp as a Fox: Are foxnews.com Visitors Less Politically Knowledgeable? *American Politics Research* **48**, 792–806 (2020). URL <http://journals.sagepub.com/doi/10.1177/1532673X20915222>.
- [316] Krueger, B. S. A Comparison of Conventional and Internet Political Mobilization. *American Politics Research* **34**, 759–776 (2006). URL <http://journals.sagepub.com/doi/10.1177/1532673X06290911>.
- [317] Makhortykh, M., de Vreese, C., Helberger, N., Harambam, J. & Bountouridis, D. We are what we click: Understanding time and content-based habits of online news readers. *New Media & Society* **23**, 2773–2800 (2021). URL <http://journals.sagepub.com/doi/10.1177/1461444820933221>.
- [318] Sydnor, E. Platforms for Incivility: Examining Perceptions Across Different Media Formats. *Political Communication* **35**, 97–116 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/10584609.2017.1355857>.
- [319] Piazza, J. A. & Guler, A. The Online Caliphate: Internet Usage and ISIS Support in the Arab World. *Terrorism and Political Violence* **33**, 1256–1275 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/09546553.2019.1606801>.
- [320] Dahlgren, P. M., Shehata, A. & Strömback, J. Reinforcing spirals at work? Mutual influences between selective news exposure and ideological leaning. *European Journal of Communication* **34**, 159–174 (2019). URL <http://journals.sagepub.com/doi/10.1177/0267323119830056>.
- [321] Garrett, R. K. Politically Motivated Reinforcement Seeking: Reframing the Selective Exposure Debate. *Journal of Communication* **59**, 676–699 (2009). URL <https://academic.oup.com/joc/article/59/4/676-699/4098408>.
- [322] Milani, E., Weitkamp, E. & Webb, P. The Visual Vaccine Debate on Twitter: A Social Network Analysis. *Media and Communication* **8**, 364–375 (2020). URL <https://www.cogitatiopress.com/mediaandcommunication/article/view/2847>.

- [323] Nguyen, A. & Western, M. Socio-structural correlates of online news and information adoption/use: Implications for the digital divide. *Journal of Sociology* **43**, 167–185 (2007). URL <http://journals.sagepub.com/doi/10.1177/1440783307076894>.
- [324] Kane, B. & Luo, J. Do the Communities We Choose Shape our Political Beliefs? A Study of the Politicization of Topics in Online Social Groups. In *2018 IEEE International Conference on Big Data (Big Data)*, 3665–3671 (IEEE, Seattle, WA, USA, 2018). URL <https://ieeexplore.ieee.org/document/8622535/>.
- [325] Wollebæk, D., Karlsen, R., Steen-Johnsen, K. & Enjolras, B. Anger, Fear, and Echo Chambers: The Emotional Basis for Online Behavior. *Social Media + Society* **5**, 205630511982985 (2019). URL <http://journals.sagepub.com/doi/10.1177/2056305119829859>.
- [326] Scheffauer, R., Goyanes, M. & Gil de Zúñiga, H. Beyond social media news use algorithms: how political discussion and network heterogeneity clarify incidental news exposure. *Online Information Review* **45**, 633–650 (2021). URL <https://www.emerald.com/insight/content/doi/10.1108/OIR-04-2020-0133/full/html>.
- [327] Chan, M., Chen, H.-T. & Lee, F. L. F. Examining the Roles of Political Social Network and Internal Efficacy on Social Media News Engagement: A Comparative Study of Six Asian Countries. *The International Journal of Press/Politics* **24**, 127–145 (2019). URL <http://journals.sagepub.com/doi/10.1177/1940161218814480>.
- [328] Miao, H. Media Use and Political Participation in China: Taking Three National Large-N Surveys as Examples. *Asian Journal for Public Opinion Research* **7**, 1–22 (2019). URL <https://doi.org/10.15206/AJPOR.2019.7.1.1>.
- [329] Peterson, E., Goel, S. & Iyengar, S. Partisan selective exposure in online news consumption: evidence from the 2016 presidential campaign. *Political Science Research and Methods* **9**, 242–258 (2021). URL [https://www.cambridge.org/core/product/identifier/S2049847019000554/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S2049847019000554/type/journal_article).
- [330] Hjorth, F. & Adler-Nissen, R. Ideological Asymmetry in the Reach of Pro-Russian Digital Disinformation to United States Audiences. *Journal of Communication* **69**, 168–192 (2019). URL <https://academic.oup.com/joc/article/69/2/168/5425470>.
- [331] Dozier, D. M., Shen, H., Sweetser, K. D. & Barker, V. Demographics and Internet behaviors as predictors of active publics. *Public Relations Review* **42**, 82–90 (2016). URL <https://linkinghub.elsevier.com/retrieve/pii/S0363811115001435>.
- [332] Asker, D. & Dinas, E. Thinking Fast and Furious: Emotional Intensity and Opinion Polarization in Online Media. *Public Opinion Quarterly* **83**, 487–509 (2019). URL <https://academic.oup.com/poq/article/83/3/487/5566253>.
- [333] Sugihartati, R., Suyanto, B. & Sirry, M. The Shift from Consumers to Prosumers: Susceptibility of Young Adults to Radicalization. *Social Sciences* **9**, 40 (2020). URL <https://www.mdpi.com/2076-0760/9/4/40>.

- [334] Johnson, T. J., Kaye, B. K. & Lee, A. M. Blinded by the Spite? Path Model of Political Attitudes, Selectivity, and Social Media. *Atlantic Journal of Communication* **25**, 181–196 (2017). URL <https://www.tandfonline.com/doi/full/10.1080/15456870.2017.1324454>.
- [335] Yang, J. & Grabe, M. E. Knowledge acquisition gaps: A comparison of print versus online news sources. *New Media & Society* **13**, 1211–1227 (2011). URL <http://journals.sagepub.com/doi/10.1177/1461444811401708>.
- [336] Bode, L., Vraga, E. K., Borah, P. & Shah, D. V. A New Space for Political Behavior: Political Social Networking and its Democratic Consequences. *Journal of Computer-Mediated Communication* **19**, 414–429 (2014). URL <https://academic.oup.com/jcmc/article/19/3/414-429/4067544>.
- [337] Shim, K. & Oh, S.-K. K. Who creates the bandwagon? The dynamics of fear of isolation, opinion congruency and anonymity-preference on social media in the 2017 South Korean presidential election. *Computers in Human Behavior* **86**, 181–189 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S0747563218301766>.
- [338] Rosenbusch, H., Evans, A. M. & Zeelenberg, M. Multilevel Emotion Transfer on YouTube: Disentangling the Effects of Emotional Contagion and Homophily on Video Audiences. *Social Psychological and Personality Science* **10**, 1028–1035 (2019). URL <http://journals.sagepub.com/doi/10.1177/1948550618820309>.
- [339] Munger, K., Luca, M., Nagler, J. & Tucker, J. The (Null) Effects of Clickbait Headlines on Polarization, Trust, and Learning. *Public Opinion Quarterly* **84**, 49–73 (2020). URL <https://academic.oup.com/poq/article/84/1/49/5827235>.
- [340] Eady, G., Nagler, J., Guess, A., Zilinsky, J. & Tucker, J. A. How Many People Live in Political Bubbles on Social Media? Evidence From Linked Survey and Twitter Data. *SAGE Open* **9**, 215824401983270 (2019). URL <http://journals.sagepub.com/doi/10.1177/2158244019832705>.
- [341] Guerrero-Solá, F. Interactive Behavior in Political Discussions on Twitter: Politicians, Media, and Citizens Patterns of Interaction in the 2015 and 2016 Electoral Campaigns in Spain. *Social Media + Society* **4**, 205630511880877 (2018). URL <http://journals.sagepub.com/doi/10.1177/2056305118808776>.
- [342] Theocharis, Y., Moor, J. & Deth, J. W. Digitally Networked Participation and Lifestyle Politics as New Modes of Political Participation. *Policy & Internet* **13**, 30–53 (2021). URL <https://onlinelibrary.wiley.com/doi/10.1002/poi3.231>.
- [343] Robles, J. M., Velez, D., De Marco, S., Rodríguez, J. T. & Gomez, D. Affective homogeneity in the Spanish general election debate. A comparative analysis of social networks political agents. *Information, Communication & Society* **23**, 216–233 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/1369118X.2018.1499792>.

- [344] Costello, M. & Hawdon, J. Who Are the Online Extremists Among Us? Sociodemographic Characteristics, Social Networking, and Online Experiences of Those Who Produce Online Hate Materials. *Violence and Gender* **5**, 55–60 (2018). URL <https://www.liebertpub.com/doi/10.1089/vio.2017.0048>.
- [345] Vaccari, C. & Valeriani, A. Digital Political Talk and Political Participation: Comparing Established and Third Wave Democracies. *SAGE Open* **8**, 215824401878498 (2018). URL <http://journals.sagepub.com/doi/10.1177/2158244018784986>.
- [346] Park, C. S. & Karan, K. Unraveling the relationships between smartphone use, exposure to heterogeneity, political efficacy, and political participation: a mediation model approach. *Asian Journal of Communication* **24**, 370–389 (2014). URL <http://www.tandfonline.com/doi/abs/10.1080/01292986.2014.892146>.
- [347] David, C. C., San Pascual, M. R. S. & Torres, M. E. S. Reliance on Facebook for news and its influence on political engagement. *PLOS ONE* **14**, e0212263 (2019). URL <https://dx.plos.org/10.1371/journal.pone.0212263>.
- [348] Arshad, S. & Khurram, S. Can governments presence on social media stimulate citizens online political participation? Investigating the influence of transparency, trust, and responsiveness. *Government Information Quarterly* **37**, 101486 (2020). URL <https://linkinghub.elsevier.com/retrieve/pii/S0740624X1930574X>.
- [349] Machackova, H. & Å erek, J. Does clicking matter? The role of online participation in adolescents civic development. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* **11** (2017). URL <https://cyberpsychology.eu/article/view/8741>.
- [350] Jeroense, T., Luimers, J., Jacobs, K. & Spierings, N. Political social media use and its linkage to populist and postmaterialist attitudes and vote intention in the Netherlands. *European Political Science* **21**, 193–215 (2022). URL <https://link.springer.com/10.1057/s41304-020-00306-6>.
- [351] Bosi, L., Lavizzari, A. & Portos, M. The impact of intolerance on young peoples online political participation. *Politics* **42**, 95–127 (2022). URL <http://journals.sagepub.com/doi/10.1177/02633957211014453>.
- [352] Sommariva, S., Vamos, C., Mantzaris, A., Uyên-Loan Đào, L. U.-L. & Martinez Tyson, D. Spreading the (Fake) News: Exploring Health Messages on Social Media and the Implications for Health Professionals Using a Case Study. *American Journal of Health Education* **49**, 246–255 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/19325037.2018.1473178>.
- [353] Kim, D. H., Jones-Jang, S. M. & Kenski, K. Why Do People Share Political Information on Social Media? *Digital Journalism* **9**, 1123–1140 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/21670811.2020.1827966>.
- [354] Kulshrestha, J. *et al.* Search bias quantification: investigating political bias in social media and web search. *Information Retrieval Journal* **22**, 188–227 (2019). URL <http://link.springer.com/10.1007/s10791-018-9341-2>.

- [355] Eddington, S. M. The Communicative Constitution of Hate Organizations Online: A Semantic Network Analysis of Make America Great Again. *Social Media + Society* **4**, 205630511879076 (2018). URL <http://journals.sagepub.com/doi/10.1177/2056305118790763>.
- [356] Feezell, J. T. & Ortiz, B. I saw it on Facebook: an experimental analysis of political learning through social media. *Information, Communication & Society* **24**, 1283–1302 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/1369118X.2019.1697340>.
- [357] Ejaz, W., Ittefaq, M., Seo, H. & Naz, F. Factors associated with the belief in COVID-19 related conspiracy theories in Pakistan. *Health, Risk & Society* **23**, 162–178 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/13698575.2021.1929865>.
- [358] Omotayo, F. & Folorunso, M. B. Use of Social Media for Political Participation by Youths. *JeDEM - eJournal of eDemocracy and Open Government* **12**, 132–157 (2020). URL <https://jedem.org/index.php/jedem/article/view/585>.
- [359] Hasangani, S. Religious Identification on Facebook Visuals and (Online) Out-group Intolerance: Experimenting the Sri Lankan Case. *Journal of Asian and African Studies* **57**, 247–268 (2022). URL <http://journals.sagepub.com/doi/10.1177/00219096211008466>.
- [360] Mueller-Herbst, J. M., Xenos, M. A., Scheufele, D. A. & Brossard, D. Saw It on Facebook: The Role of Social Media in Facilitating Science Issue Awareness. *Social Media + Society* **6**, 205630512093041 (2020). URL <http://journals.sagepub.com/doi/10.1177/2056305120930412>.
- [361] Lu, Y., Lee, J. K. & Kim, E. Network characteristics matter in politics on Facebook: evidence from a US national survey. *Online Information Review* **42**, 372–386 (2018). URL <https://www.emerald.com/insight/content/doi/10.1108/OIR-09-2016-0262/full/html>.
- [362] Forati, A. M. & Ghose, R. Geospatial analysis of misinformation in COVID-19 related tweets. *Applied Geography* **133**, 102473 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0143622821000898>.
- [363] Hong, S. & Kim, S. H. Political polarization on twitter: Implications for the use of social media in digital governments. *Government Information Quarterly* **33**, 777–782 (2016). URL <https://linkinghub.elsevier.com/retrieve/pii/S0740624X16300375>.
- [364] Westerwick, A., Sude, D., Robinson, M. & Knobloch-Westerwick, S. Peers Versus Pros: Confirmation Bias in Selective Exposure to User-Generated Versus Professional Media Messages and Its Consequences. *Mass Communication and Society* **23**, 510–536 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/15205436.2020.1721542>.
- [365] Stoica, A.-A., Riederer, C. & Chaintreau, A. Algorithmic Glass Ceiling in Social Networks: The effects of social recommendations on network diversity. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, 923–932 (ACM Press, Lyon, France, 2018). URL <http://dl.acm.org/citation.cfm?doid=3178876.3186140>.



- [366] Vissenberg, J., Coninck, D. D. & dHaenens, L. Relating adolescents exposure to legacy and digital news media and intergroup contact to their attitudes towards immigrants. *Communications* **46**, 373–393 (2021). URL <https://www.degruyter.com/document/doi/10.1515/commun-2021-0047/html>.
- [367] Yu, R. P. & Oh, Y. W. Social media and expressive citizenship: Understanding the relationships between social and entertainment expression on Facebook and political participation. *Telematics and Informatics* **35**, 2299–2311 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S073658531830460X>.
- [368] Lake, J. S., Alston, A. T. & Kahn, K. B. How Social Networking Use and Beliefs About Inequality Affect Engagement With Racial Justice Movements. *Race and Justice* **11**, 500–519 (2021). URL <http://journals.sagepub.com/doi/10.1177/2153368718809833>.
- [369] Barnidge, M., Huber, B., de Zúñiga, H. G. & Liu, J. H. Social Media as a Sphere for Risky Political Expression: A Twenty-Country Multilevel Comparative Analysis. *The International Journal of Press/Politics* **23**, 161–182 (2018). URL <http://journals.sagepub.com/doi/10.1177/1940161218773838>.
- [370] Mustapha, L. K. & Omar, B. Do social media matter? Examining social media use and youths political participation during the 2019 Nigerian general elections. *The Round Table* **109**, 441–457 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/00358533.2020.1788766>.
- [371] Qin, A. Y. Judging Them by My Media Use: Exploring the Cause and Consequences of Perceived Selective Exposure. *Mass Communication and Society* **25**, 237–259 (2022). URL <https://www.tandfonline.com/doi/full/10.1080/15205436.2021.1977327>.
- [372] Blank, G. & Lutz, C. Benefits and harms from Internet use: A differentiated analysis of Great Britain. *New Media & Society* **20**, 618–640 (2018). URL <http://journals.sagepub.com/doi/10.1177/1461444816667135>.
- [373] Bail, C. A., Merhout, F. & Ding, P. Using Internet search data to examine the relationship between anti-Muslim and pro-ISIS sentiment in U.S. counties. *Science Advances* **4**, eaao5948 (2018). URL <https://www.science.org/doi/10.1126/sciadv.aao5948>.
- [374] Hendriks Vettehen, P., Troost, J., Boerboom, L., Steijaert, M. & Scheepers, P. The Relationship Between Media Content Preferences and Political Participation in 25 European Countries: The Moderating Role of Broadband Penetration and Broadband Access. *Communication Research* **47**, 967–987 (2020). URL <http://journals.sagepub.com/doi/10.1177/0093650217701033>.
- [375] Arlt, D. & Wolling, J. Bias wanted! Examining peoples information exposure, quality expectations and bias perceptions in the context of the refugees debate among different segments of the German population. *Communications* **43**, 75–99 (2018). URL <https://www.degruyter.com/document/doi/10.1515/commun-2017-0045/html>.
- [376] Cardenal, A. S., Aguilar-Paredes, C., Cristancho, C. & Majó-Vázquez, S. Echo-chambers in online news consumption: Evidence from survey and navigation data in Spain. *European Journal of Communication* **34**, 360–376 (2019). URL <http://journals.sagepub.com/doi/10.1177/0267323119844409>.

- [377] Gallego, J., Mart nez, J. D., Munger, K. & V squez-Cort s, M. Tweeting for peace: Experimental evidence from the 2016 Colombian Plebiscite. *Electoral Studies* **62**, 102072 (2019). URL <https://linkinghub.elsevier.com/retrieve/pii/S026137941930006X>.
- [378] Allington, D., McAndrew, S., Moxham-Hall, V. L. & Duffy, B. Media usage predicts intention to be vaccinated against SARS-CoV-2 in the US and the UK. *Vaccine* **39**, 2595–2603 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0264410X21002346>.
- [379] Wagner, K. M., Gainous, J. & Abbott, J. P. Gender Differences in Critical Digital Political Engagement in China: The Consequences for Protest Attitudes. *Social Science Computer Review* **39**, 211–225 (2021). URL <http://journals.sagepub.com/doi/10.1177/0894439319865907>.
- [380] Foos, F., Kostadinov, L., Marinov, N. & Schimmelfennig, F. Does Social Media Promote Civic Activism? A Field Experiment with a Civic Campaign. *Political Science Research and Methods* **9**, 500–518 (2021). URL [https://www.cambridge.org/core/product/identifier/S2049847020000138/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S2049847020000138/type/journal_article).
- [381] Erdem, R. & Ozejder, I. Use of Social Media for Political Purposes: The Case of Diyarbakir. *Revista de Cercetare si Interventie Sociala* **72**, 187–209 (2021). URL [https://www.rcis.ro/images/documente/rcis72\\_13.pdf](https://www.rcis.ro/images/documente/rcis72_13.pdf).
- [382] Bimber, B. Information and Political Engagement in America: The Search for Effects of Information Technology at the Individual Level. *Political Research Quarterly* **54**, 53 (2001). URL <https://www.jstor.org/stable/449207?origin=crossref>.
- [383] Sindermann, C., Elhai, J. D., Moshagen, M. & Montag, C. Age, gender, personality, ideological attitudes and individual differences in a person’s news spectrum: how many and who might be prone to filter bubbles and echo chambers online? *Heliyon* **6**, e03214 (2020). URL <https://linkinghub.elsevier.com/retrieve/pii/S2405844020300591>.
- [384] Guntuku, S. C., Buttenheim, A. M., Sherman, G. & Merchant, R. M. Twitter discourse reveals geographical and temporal variation in concerns about COVID-19 vaccines in the United States. *Vaccine* **39**, 4034–4038 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0264410X21007386>.
- [385] Kim, H. & Joshanloo, M. Internet Access and Voicing Opinions: The Moderating Roles of Age and the National Economy. *Social Indicators Research* **150**, 121–141 (2020). URL <http://link.springer.com/10.1007/s11205-020-02298-8>.
- [386] Heinsohn, T., Fatke, M., Israel, J., Marschall, S. & Schultze, M. Effects of voting advice applications during election campaigns. Evidence from a panel study at the 2014 European elections. *Journal of Information Technology & Politics* **16**, 250–264 (2019). URL <https://www.tandfonline.com/doi/full/10.1080/19331681.2019.1644265>.

- [387] Shmargad, Y. & Klar, S. Sorting the News: How Ranking by Popularity Polarizes Our Politics. *Political Communication* **37**, 423–446 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/10584609.2020.1713267>.
- [388] Chen, H.-T. Spiral of silence on social media and the moderating role of disagreement and publicness in the network: Analyzing expressive and withdrawal behaviors. *New Media & Society* **20**, 3917–3936 (2018). URL <http://journals.sagepub.com/doi/10.1177/1461444818763384>.
- [389] Romer, D. & Jamieson, K. H. Patterns of Media Use, Strength of Belief in COVID-19 Conspiracy Theories, and the Prevention of COVID-19 From March to July 2020 in the United States: Survey Study. *Journal of Medical Internet Research* **23**, e25215 (2021). URL <https://www.jmir.org/2021/4/e25215>.
- [390] Klein, E. & Robison, J. Like, Post, and Distrust? How Social Media Use Affects Trust in Government. *Political Communication* **37**, 46–64 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/10584609.2019.1661891>.
- [391] Kim, C. & Lee, S. Does social media type matter to politics? Investigating the difference in political participation depending on preferred social media sites. *Social Science Quarterly* **102**, 2942–2954 (2021). URL <https://onlinelibrary.wiley.com/doi/10.1111/ssqu.13055>.
- [392] Garrett, R. K. & Bond, R. M. Conservatives susceptibility to political misperceptions. *Science Advances* **7**, eabf1234 (2021). URL <https://www.science.org/doi/10.1126/sciadv.abf1234>.
- [393] van Tubergen, F., Cinjee, T., Menshikova, A. & Veldkamp, J. Online activity of mosques and Muslims in the Netherlands: A study of Facebook, Instagram, YouTube and Twitter. *PLOS ONE* **16**, e0254881 (2021). URL <https://dx.plos.org/10.1371/journal.pone.0254881>.
- [394] Gherghina, S. & Rusu, E. Begin Again: Election Campaign and Own Opinions Among First-Time Voters in Romania. *Social Science Quarterly* **102**, 1311–1329 (2021). URL <https://onlinelibrary.wiley.com/doi/10.1111/ssqu.12979>.
- [395] Karakaya, S. & Glazier, R. A. Media, information, and political participation: The importance of online news sources in the absence of a free press. *Journal of Information Technology & Politics* **16**, 290–306 (2019). URL <https://www.tandfonline.com/doi/full/10.1080/19331681.2019.1645784>.
- [396] Heatherly, K. A., Lu, Y. & Lee, J. K. Filtering out the other side? Cross-cutting and like-minded discussions on social networking sites. *New Media & Society* **19**, 1271–1289 (2017). URL <http://journals.sagepub.com/doi/10.1177/1461444816634677>.
- [397] Kaakinen, M., Oksanen, A. & Räsänen, P. Did the risk of exposure to online hate increase after the November 2015 Paris attacks? A group relations approach. *Computers in Human Behavior* **78**, 90–97 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S0747563217305484>.

- [398] Kurfi, M. Y., Msughter, M. E. & Mohamed, I. Digital Images on Social Media and Proliferation of Fake News on Covid-19 in Kano, Nigeria. *Galactica Media: Journal of Media Studies* **3**, 103–124 (2021). URL <https://galacticamedia.com/index.php/gmd/article/view/111>.
- [399] Zhang, X. & Lin, W.-Y. Stoking the Fires of Participation: Extending the Gamson Hypothesis on Social Media Use and Elite-challenging Political Engagement. *Computers in Human Behavior* **79**, 217–226 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S074756321730609X>.
- [400] Costello, M., Barrett-Fox, R., Bernatzky, C., Hawdon, J. & Mendes, K. Predictors of Viewing Online Extremism Among Americas Youth. *Youth & Society* **52**, 710–727 (2020). URL <http://journals.sagepub.com/doi/10.1177/0044118X18768115>.
- [401] Dvir-Gvirsman, S., Tsfati, Y. & Menchen-Trevino, E. The extent and nature of ideological selective exposure online: Combining survey responses with actual web log data from the 2013 Israeli Elections. *New Media & Society* **18**, 857–877 (2016). URL <http://journals.sagepub.com/doi/10.1177/1461444814549041>.
- [402] Chang, K. & Park, J. Social Media Use and Participation in Dueling Protests: The Case of the 2016-2017 Presidential Corruption Scandal in South Korea. *The International Journal of Press/Politics* **26**, 547–567 (2021). URL <http://journals.sagepub.com/doi/10.1177/1940161220940962>.
- [403] Lee, H. & Hahn, K. S. Partisan selective following on twitter over time: polarization or depolarization? *Asian Journal of Communication* **28**, 227–246 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/01292986.2017.1384845>.
- [404] Boxell, L., Gentzkow, M. & Shapiro, J. M. Greater Internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences* **114**, 10612–10617 (2017). URL <https://pnas.org/doi/full/10.1073/pnas.1706588114>.
- [405] Stier, S., Kirkizh, N., Froio, C. & Schroeder, R. Populist Attitudes and Selective Exposure to Online News: A Cross-Country Analysis Combining Web Tracking and Surveys. *The International Journal of Press/Politics* **25**, 426–446 (2020). URL <http://journals.sagepub.com/doi/10.1177/1940161220907018>.
- [406] Garrett, R. K. *et al.* Implications of Pro- and Counterattitudinal Information Exposure for Affective Polarization: Partisan Media Exposure and Affective Polarization. *Human Communication Research* **40**, 309–332 (2014). URL <https://academic.oup.com/hcr/article/40/3/309-332/4093810>.
- [407] Sharma, I., Jain, K. & Singh, G. Effect of online political incivility on partisan attitude: role of issue involvement, moral identity and incivility accountability. *Online Information Review* **44**, 1421–1441 (2020). URL <https://www.emerald.com/insight/content/doi/10.1108/OIR-03-2020-0084/full/html>.
- [408] Stella, M., Ferrara, E. & De Domenico, M. Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences* **115**, 12435–12440 (2018). URL <https://pnas.org/doi/full/10.1073/pnas.1803470115>.

- [409] Ackland, R., O'Neil, M. & Park, S. Engagement with news on Twitter: insights from Australia and Korea. *Asian Journal of Communication* **29**, 235–251 (2019). URL <https://www.tandfonline.com/doi/full/10.1080/01292986.2018.1462393>.
- [410] Mothes, C. & Ohme, J. Partisan Selective Exposure in Times of Political and Technological Upheaval: A Social Media Field Experiment. *Media and Communication* **7**, 42–53 (2019). URL <https://www.cogitatiopress.com/mediaandcommunication/article/view/2183>.
- [411] Popa, S. A., Theocharis, Y. & Schnaudt, C. From seeing the writing on the wall, to getting together for a bowl: Direct and compensating effects of Facebook use on offline associational membership. *Journal of Information Technology & Politics* **13**, 222–238 (2016). URL <http://www.tandfonline.com/doi/full/10.1080/19331681.2016.1194241>.
- [412] Karlsen, R., Beyer, A. & Steen-Johnsen, K. Do High-Choice Media Environments Facilitate News Avoidance? A Longitudinal Study 1997–2016. *Journal of Broadcasting & Electronic Media* **64**, 794–814 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/08838151.2020.1835428>.
- [413] Goyanes, M. Antecedents of Incidental News Exposure: The Role of Media Preference, Use and Trust. *Journalism Practice* **14**, 714–729 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/17512786.2019.1631710>.
- [414] Siegel, A. A., Nagler, J., Bonneau, R. & Tucker, J. A. Tweeting Beyond Tahrir: Ideological Diversity and Political Intolerance in Egyptian Twitter Networks. *World Politics* **73**, 243–274 (2021). URL [https://www.cambridge.org/core/product/identifier/S0043887120000295/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S0043887120000295/type/journal_article).
- [415] Mashuri, A. *et al.* The socio-psychological predictors of support for post-truth collective action. *The Journal of Social Psychology* **162**, 504–522 (2022). URL <https://www.tandfonline.com/doi/full/10.1080/00224545.2021.1935678>.
- [416] Alshareef, M. & Alotiby, A. Prevalence and Perception Among Saudi Arabian Population About Resharing of Information on Social Media Regarding Natural Remedies as Protective Measures Against COVID-19. *International Journal of General Medicine* **Volume 14**, 5127–5137 (2021). URL <https://www.dovepress.com/prevalence-and-perception-among-saudi-arabian-population-about-reshari-peer-reviewed-fulltext-art>
- [417] North, S., Piwek, L. & Joinson, A. Battle for Britain: Analyzing Events as Drivers of Political Tribalism in Twitter Discussions of Brexit. *Policy & Internet* **13**, 185–208 (2021). URL <https://onlinelibrary.wiley.com/doi/10.1002/poi3.247>.
- [418] Florio, K., Basile, V., Lai, M. & Patti, V. Leveraging Hate Speech Detection to Investigate Immigration-related Phenomena in Italy. In *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 1–7 (IEEE, Cambridge, United Kingdom, 2019). URL <https://ieeexplore.ieee.org/document/8925079/>.

- [419] Dilliplane, S. All the News You Want to Hear: The Impact of Partisan News Exposure on Political Participation. *Public Opinion Quarterly* **75**, 287–316 (2011). URL <https://academic.oup.com/poq/article-lookup/doi/10.1093/poq/nfr006>.
- [420] Wasisto, A. Electoral volatility of the 2019 presidential election: A study in Jakarta and Depok, Indonesia. *Masyarakat, Kebudayaan dan Politik* **34**, 281 (2021). URL <https://e-journal.unair.ac.id/MKP/article/view/18842>.
- [421] Sturm Wilkerson, H., Riedl, M. J. & Whipple, K. N. Affective Affordances: Exploring Facebook Reactions as Emotional Responses to Hyperpartisan Political News. *Digital Journalism* **9**, 1040–1061 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/21670811.2021.1899011>.
- [422] Zunino, E. Medios digitales y COVID-19: sobreinformación, polarización y desinformación. *Universitas* 133–154 (2021). URL <https://universitas.ups.edu.ec/index.php/universitas/article/view/34.2021.06>.
- [423] Strauß, N., Huber, B. & Gil de Zúñiga, H. Structural Influences on the News Finds Me Perception: Why People Believe They Dont Have to Actively Seek News Anymore. *Social Media + Society* **7**, 205630512110249 (2021). URL <http://journals.sagepub.com/doi/10.1177/20563051211024966>.
- [424] Tai, K.-T., Porumbescu, G. & Shon, J. Can e-participation stimulate offline citizen participation: an empirical test with practical implications. *Public Management Review* **22**, 278–296 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/14719037.2019.1584233>.
- [425] Kim, H. The mere exposure effect of tweets on vote choice. *Journal of Information Technology & Politics* **18**, 455–465 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/19331681.2021.1900979>.
- [426] Marozzo, F. & Bessi, A. Analyzing polarization of social media users and news sites during political campaigns. *Social Network Analysis and Mining* **8**, 1 (2018). URL <http://link.springer.com/10.1007/s13278-017-0479-5>.
- [427] Granberg-Rademacker, J. S. & Parsneau, K. Lets Get Ready to Tweet! An Analysis of Twitter Use by 2018 Senate Candidates. *Congress & the Presidency* **48**, 78–100 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/07343469.2020.1728425>.
- [428] Chan, M. Media Use and the Social Identity Model of Collective Action: Examining the Roles of Online Alternative News and Social Media News. *Journalism & Mass Communication Quarterly* **94**, 663–681 (2017). URL <http://journals.sagepub.com/doi/10.1177/1077699016638837>.
- [429] Kitchens, B., Johnson, S. L. & Gray, P. Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption. *MIS Quarterly* **44**, 1619–1649 (2020). URL <https://misq.org/understanding-echo-chambers-and-filter-bubbles-the-impact-of-social-media-on-diversification-and.html>.

- [430] Amit, S., Mannan, S. & Islam, A. BANGLADESH: TIME SPENT ONLINE, CONFLICT AND RADICALIZATION. *Conflict Studies Quarterly* 3–21 (2020). URL <http://www.csq.ro/wp-content/uploads/Sajid-AMIT-et-al.pdf>.
- [431] Casero-RipollÃ©s, A. Influencia de los medios de comunicaci3n en la conversaci3n polÃ©tica en Twitter. *Revista ICONO14 Revista cientÃ©fica de Comunicaci3n y TecnologÃ©as emergentes* 18, 33–57 (2020). URL <https://icono14.net/ojs/index.php/icono14/article/view/1527>.
- [432] Saud, M., El Hariri, D. B. & Ashfaq, A. The role of social media in promoting political participation: The Lebanon experience. *Masyarakat, Kebudayaan dan Politik* 33, 248 (2020). URL <https://e-journal.unair.ac.id/MKP/article/view/19054>.
- [433] Sismeiro, C. & Mahmood, A. Competitive vs. Complementary Effects in Online Social Networks and News Consumption: A Natural Experiment. *Management Science* 64, 5014–5037 (2018). URL <http://pubsonline.informs.org/doi/10.1287/mnsc.2017.2896>.
- [434] Johannesson, M. P. & Knudsen, E. Disentangling the Influence of Recommender Attributes and News-Story Attributes: A Conjoint Experiment on Exposure and Sharing Decisions on Social Networking Sites. *Digital Journalism* 9, 1141–1161 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/21670811.2020.1805780>.
- [435] Yamamoto, M. & Nah, S. Mobile information seeking and political participation: A differential gains approach with offline and online discussion attributes. *New Media & Society* 20, 2070–2090 (2018). URL <http://journals.sagepub.com/doi/10.1177/1461444817712722>.
- [436] Lee, J. M., Park, Y. & Kim, G. D. Social Media and Regionalism in South Korean Voting Behavior: The Case of the 19th South Korean Presidential Election. *Issues & Studies* 54, 1840006 (2018). URL <https://www.worldscientific.com/doi/abs/10.1142/S1013251118400064>.
- [437] Allcott, H. & Gentzkow, M. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31, 211–236 (2017). URL <https://pubs.aeaweb.org/doi/10.1257/jep.31.2.211>.
- [438] Theocharis, Y., Barberá, P., Fazekas, Z. & Poppo, S. A. The Dynamics of Political Incivility on Twitter. *SAGE Open* 10, 215824402091944 (2020). URL <http://journals.sagepub.com/doi/10.1177/2158244020919447>.
- [439] Nah, S. & Yamamoto, M. The Integrated Media Effect: Rethinking the Effect of Media Use on Civic Participation in the Networked Digital Media Environment. *American Behavioral Scientist* 62, 1061–1078 (2018). URL <http://journals.sagepub.com/doi/10.1177/0002764218764240>.
- [440] Pang, H. Can microblogs motivate involvement in civic and political life? Examining uses, gratifications and social outcomes among Chinese youth. *Online Information Review* 42, 663–680 (2018). URL <https://www.emerald.com/insight/content/doi/10.1108/OIR-04-2017-0136/full/html>.

- [441] Moeller, J., Kühne, R. & De Vreese, C. Mobilizing Youth in the 21st Century: How Digital Media Use Fosters Civic Duty, Information Efficacy, and Political Participation. *Journal of Broadcasting & Electronic Media* **62**, 445–460 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/08838151.2018.1451866>.
- [442] Cinelli, M. *et al.* Selective exposure shapes the Facebook news diet. *PLOS ONE* **15**, e0229129 (2020). URL <https://dx.plos.org/10.1371/journal.pone.0229129>.
- [443] Yamamoto, M. & Morey, A. C. Incidental News Exposure on Social Media: A Campaign Communication Mediation Approach. *Social Media + Society* **5**, 205630511984361 (2019). URL <http://journals.sagepub.com/doi/10.1177/2056305119843619>.
- [444] Levy, R. Social Media, News Consumption, and Polarization: Evidence from a Field Experiment. *American Economic Review* **111**, 831–870 (2021). URL <https://pubs.aeaweb.org/doi/10.1257/aer.20191777>.
- [445] Vozab, D. Generational Patterns of Digital News Consumption: From Traditionalists to Millennial Minimalists. *Medijske studije* **10**, 107–126 (2020). URL [https://hrcak.srce.hr/index.php?show=clanak&id\\_clanak\\_jezik=343064](https://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=343064).
- [446] Jamal, A., Kizgin, H., Rana, N. P., Laroche, M. & Dwivedi, Y. K. Impact of acculturation, online participation and involvement on voting intentions. *Government Information Quarterly* **36**, 510–519 (2019). URL <https://linkinghub.elsevier.com/retrieve/pii/S0740624X1830488X>.
- [447] Shen, F., Xia, C. & Skoric, M. Examining the roles of social media and alternative media in social movement participation: A study of Hong Kongs Umbrella Movement. *Telematics and Informatics* **47**, 101303 (2020). URL <https://linkinghub.elsevier.com/retrieve/pii/S0736585319307956>.
- [448] Dohle, M., Bernhard, U. & Kelm, O. Presumed Media Influences and Demands for Restrictions: Using Panel Data to Examine the Causal Direction. *Mass Communication and Society* **20**, 595–613 (2017). URL <https://www.tandfonline.com/doi/full/10.1080/15205436.2017.1303072>.
- [449] Marquart, F., Goldberg, A. C. & de Vreese, C. H. This time Im (not) voting: A comprehensive overview of campaign factors influencing turnout at European Parliament elections. *European Union Politics* **21**, 680–705 (2020). URL <http://journals.sagepub.com/doi/10.1177/1465116520943670>.
- [450] Hoffmann, C. P. & Lutz, C. Digital Divides in Political Participation: The Mediating Role of Social Media Self-Efficacy and Privacy Concerns. *Policy & Internet* **13**, 6–29 (2021). URL <https://onlinelibrary.wiley.com/doi/10.1002/poi3.225>.
- [451] Mitts, T. From Isolation to Radicalization: Anti-Muslim Hostility and Support for ISIS in the West. *American Political Science Review* **113**, 173–194 (2019). URL [https://www.cambridge.org/core/product/identifier/S0003055418000618/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S0003055418000618/type/journal_article).
- [452] Lukito, J. Coordinating a Multi-Platform Disinformation Campaign: Internet Research Agency Activity on Three U.S. Social Media Platforms, 2015 to 2017. *Political Communication* **37**, 238–255 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/10584609.2019.1661889>.



- [453] Hong, S., Choi, H. & Kim, T. K. Why Do Politicians Tweet? Extremists, Underdogs, and Opposing Parties as Political Tweeters. *Policy & Internet* **11**, 305–323 (2019). URL <https://onlinelibrary.wiley.com/doi/10.1002/poi3.201>.
- [454] Chadwick, A. *et al.* Online Social Endorsement and Covid-19 Vaccine Hesitancy in the United Kingdom. *Social Media + Society* **7**, 205630512110088 (2021). URL <http://journals.sagepub.com/doi/10.1177/20563051211008817>.
- [455] Zumárraga-Espinosa, M. Redes sociales y protesta política: Un análisis del rol moderador del estatus socioeconómico y la pertenencia a grupos políticos. *Doxa Comunicación. Revista interdisciplinaria de estudios de comunicación y ciencias sociales* **55–77** (2020). URL <https://revistascientificas.uspceu.com/doxacomunicacion/article/view/500>.
- [456] Fletcher, R. & Nielsen, R. K. Automated Serendipity: The effect of using search engines on news repertoire balance and diversity. *Digital Journalism* **6**, 976–989 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/21670811.2018.1502045>.
- [457] Sell, T. K., Hosangadi, D. & Trotochaud, M. Misinformation and the US Ebola communication crisis: analyzing the veracity and content of social media messages related to a fear-inducing infectious disease outbreak. *BMC Public Health* **20**, 550 (2020). URL <https://bmcpublihealth.biomedcentral.com/articles/10.1186/s12889-020-08697-3>.
- [458] Brugnoli, E., Cinelli, M., Quattrociocchi, W. & Scala, A. Recursive patterns in online echo chambers. *Scientific Reports* **9**, 20118 (2019). URL <http://www.nature.com/articles/s41598-019-56191-7>.
- [459] Barnidge, M., Kim, B., Sherrill, L. A., Luknar, Å. & Zhang, J. Perceived exposure to and avoidance of hate speech in various communication settings. *Telematics and Informatics* **44**, 101263 (2019). URL <https://linkinghub.elsevier.com/retrieve/pii/S0736585319307555>.
- [460] Kim, B. & Hoewe, J. Developing contemporary factors of political participation. *The Social Science Journal* **1–15** (2020). URL <https://www.tandfonline.com/doi/full/10.1080/03623319.2020.1782641>.
- [461] Barnidge, M., Sayre, B. & Rojas, H. Perceptions of the Media and the Public and their Effects on Political Participation in Colombia. *Mass Communication and Society* **18**, 259–280 (2015). URL <http://www.tandfonline.com/doi/full/10.1080/15205436.2014.923463>.
- [462] Petrova, M., Sen, A. & Yildirim, P. Social Media and Political Contributions: The Impact of New Technology on Political Competition. *Management Science* **67**, 2997–3021 (2021). URL <http://pubsonline.informs.org/doi/10.1287/mnsc.2020.3740>.
- [463] Bryson, B. P. Polarizing the middle: internet exposure and public opinion. *International Journal of Sociology and Social Policy* **40**, 99–113 (2020). URL <https://www.emerald.com/insight/content/doi/10.1108/IJSSP-09-2019-0181/full/html>.

- [464] Bovet, A. & Makse, H. A. Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications* **10**, 7 (2019). URL <http://www.nature.com/articles/s41467-018-07761-2>.
- [465] Germani, F. & Biller-Andorno, N. The anti-vaccination infodemic on social media: A behavioral analysis. *PLOS ONE* **16**, e0247642 (2021). URL <https://dx.plos.org/10.1371/journal.pone.0247642>.
- [466] Wei, R. & Lo, V.-h. News Media Use and Knowledge about the 2006 U.S. Midterm Elections: Why Exposure Matters in Voter Learning. *International Journal of Public Opinion Research* **20**, 347–362 (2008). URL <https://academic.oup.com/ijpor/article-lookup/doi/10.1093/ijpor/edn032>.
- [467] Zhu, A. Y. F., Chan, A. L. S. & Chou, K. L. The Pathway Toward Radical Political Participation Among Young People in Hong Kong: A Communication Mediation Approach. *East Asia* **37**, 45–62 (2020). URL <http://link.springer.com/10.1007/s12140-019-09326-6>.
- [468] Balcells, J. & Padró-Solanet, A. Crossing Lines in the Twitter Debate on Catalonias Independence. *The International Journal of Press/Politics* **25**, 28–52 (2020). URL <http://journals.sagepub.com/doi/10.1177/1940161219858687>.
- [469] Grover, P., Kar, A. K., Dwivedi, Y. K. & Janssen, M. Polarization and acculturation in US Election 2016 outcomes - Can twitter analytics predict changes in voting preferences. *Technological Forecasting and Social Change* **145**, 438–460 (2019). URL <https://linkinghub.elsevier.com/retrieve/pii/S004016251731421X>.
- [470] Choi, D.-H. & Shin, D.-H. A dialectic perspective on the interactive relationship between social media and civic participation: the moderating role of social capital. *Information, Communication & Society* **20**, 151–166 (2017). URL <https://www.tandfonline.com/doi/full/10.1080/1369118X.2016.1154586>.
- [471] Thorson, K., Cotter, K., Medeiros, M. & Pak, C. Algorithmic inference, political interest, and exposure to news and politics on Facebook. *Information, Communication & Society* **24**, 183–200 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/1369118X.2019.1642934>.
- [472] Kim, Y., Hsu, S.-H. & de Zúñiga, H. G. Influence of Social Media Use on Discussion Network Heterogeneity and Civic Engagement: The Moderating Role of Personality Traits: Social Media & Personality Traits. *Journal of Communication* **63**, 498–516 (2013). URL <https://academic.oup.com/joc/article/63/3/498-516/4085989>.
- [473] Guess, A. M., Nyhan, B. & Reifler, J. Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behaviour* **4**, 472–480 (2020). URL <http://www.nature.com/articles/s41562-020-0833-x>.
- [474] Hokka, J. & Nelimarkka, M. Affective economy of national-populist images: Investigating national and transnational online networks through visual big data. *New Media & Society* **22**, 770–792 (2020). URL <http://journals.sagepub.com/doi/10.1177/1461444819868686>.
- [475] Schumann, S., Boer, D., Hanke, K. & Liu, J. Social media use and support for populist radical right parties: assessing exposure and selection effects in a two-wave panel study. *Information, Communication & Society* **24**, 921–940 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/1369118X.2019.1668455>.

- [476] Kim, M. How does Facebook news use lead to actions in South Korea? The role of Facebook discussion network heterogeneity, political interest, and conflict avoidance in predicting political participation. *Telematics and Informatics* **35**, 1373–1381 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S0736585317308407>.
- [477] Mosca, L. & Quaranta, M. Are digital platforms potential drivers of the populist vote? A comparative analysis of France, Germany and Italy. *Information, Communication & Society* **24**, 1441–1459 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/1369118X.2021.1894211>.
- [478] Justwan, F., Baumgaertner, B., Carlisle, J. E., Clark, A. K. & Clark, M. Social media echo chambers and satisfaction with democracy among Democrats and Republicans in the aftermath of the 2016 US elections. *Journal of Elections, Public Opinion and Parties* **28**, 424–442 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/17457289.2018.1434784>.
- [479] Smith, S. T., Kao, E. K., Shah, D. C., Simek, O. & Rubin, D. B. Influence Estimation on Social Media Networks Using Causal Inference. In *2018 IEEE Statistical Signal Processing Workshop (SSP)*, 328–332 (IEEE, Freiburg, 2018). URL <https://ieeexplore.ieee.org/document/8450823/>.
- [480] Bode, L. *et al.* Participation in Contentious Politics: Rethinking the Roles of News, Social Media, and Conversation Amid Divisiveness. *Journal of Information Technology & Politics* **15**, 215–229 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/19331681.2018.1485607>.
- [481] Cinelli, M., Cresci, S., Galeazzi, A., Quattrociocchi, W. & Tesconi, M. The limited reach of fake news on Twitter during 2019 European elections. *PLOS ONE* **15**, e0234689 (2020). URL <https://dx.plos.org/10.1371/journal.pone.0234689>.
- [482] Guenther, L., Ruhrmann, G., Bischoff, J., Penzel, T. & Weber, A. Strategic Framing and Social Media Engagement: Analyzing Memes Posted by the German Identitarian Movement on Facebook. *Social Media + Society* **6**, 205630511989877 (2020). URL <http://journals.sagepub.com/doi/10.1177/2056305119898777>.
- [483] Song, H., Cho, J. & Benefield, G. A. The Dynamics of Message Selection in Online Political Discussion Forums: Self-Segregation or Diverse Exposure? *Communication Research* **47**, 125–152 (2020). URL <http://journals.sagepub.com/doi/10.1177/0093650218790144>.
- [484] Rice, L. L. & Moffett, K. W. Snapchat and civic engagement among college students. *Journal of Information Technology & Politics* **16**, 87–104 (2019). URL <https://www.tandfonline.com/doi/full/10.1080/19331681.2019.1574249>.
- [485] Beam, M. A., Hmielowski, J. D. & Hutchens, M. J. Democratic Digital Inequalities: Threat and Opportunity in Online Citizenship From Motivation and Ability. *American Behavioral Scientist* **62**, 1079–1096 (2018). URL <http://journals.sagepub.com/doi/10.1177/0002764218764253>.

- [486] Hermann, E., Eisend, M. & Bayón, T. Facebook and the cultivation of ethnic diversity perceptions and attitudes. *Internet Research* **30**, 1123–1141 (2020). URL <https://www.emerald.com/insight/content/doi/10.1108/INTR-10-2019-0423/full/html>.
- [487] Powell, A., Scott, A. J. & Henry, N. Digital harassment and abuse: Experiences of sexuality and gender minority adults. *European Journal of Criminology* **17**, 199–223 (2020). URL <http://journals.sagepub.com/doi/10.1177/1477370818788006>.
- [488] Li, L., Chen, J. & Raghunathan, S. Informative Role of Recommender Systems in Electronic Marketplaces: A Boon or a Bane for Competing Sellers. *MIS Quarterly* **44**, 1957–1985 (2020). URL <https://misq.org/informative-role-of-recommender-systems-in-electronic-marketplaces-a-boon-or-a-bane-for-competing.html>.
- [489] Choi, D.-H., Yoo, W., Noh, G.-Y. & Park, K. The impact of social media on risk perceptions during the MERS outbreak in South Korea. *Computers in Human Behavior* **72**, 422–431 (2017). URL <https://linkinghub.elsevier.com/retrieve/pii/S074756321730153X>.
- [490] Samuel-Azran, T. & Hayat, T. Online news recommendations credibility: The tie is mightier than the source. *Comunicar* **27**, 71–80 (2019). URL <https://www.revistacomunicar.com/index.php?contenido=detalles&numero=60&articulo=60-2019-07>.
- [491] Kushin, M. J., Yamamoto, M. & Dalisay, F. Societal Majority, Facebook, and the Spiral of Silence in the 2016 US Presidential Election. *Social Media + Society* **5**, 205630511985513 (2019). URL <http://journals.sagepub.com/doi/10.1177/2056305119855139>.
- [492] Ardi, R. Partisan Selective Exposure to Fake News Content. *Makara Human Behavior Studies in Asia* **23**, 6 (2019). URL <https://scholarhub.ui.ac.id/hubsasia/vol23/iss1/3/>.
- [493] Feld, S. L. & McGail, A. Egonets as systematically biased windows on society. *Network Science* **8**, 399–417 (2020). URL [https://www.cambridge.org/core/product/identifier/S2050124220000053/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S2050124220000053/type/journal_article).
- [494] Hedayatifar, L., Rigg, R. A., Bar-Yam, Y. & Morales, A. J. US social fragmentation at multiple scales. *Journal of The Royal Society Interface* **16**, 20190509 (2019). URL <https://royalsocietypublishing.org/doi/10.1098/rsif.2019.0509>.
- [495] Bale, T., Webb, P. & Poletti, M. Participating Locally and Nationally: Explaining the Offline and Online Activism of British Party Members. *Political Studies* **67**, 658–675 (2019). URL <http://journals.sagepub.com/doi/10.1177/0032321718794740>.
- [496] Celik, S. Experiences of internet users regarding cyberhate. *Information Technology & People* **32**, 1446–1471 (2019). URL <https://www.emerald.com/insight/content/doi/10.1108/ITP-01-2018-0009/full/html>.

- [497] Sainudiin, R., Yogeewaran, K., Nash, K. & Sahioun, R. Characterizing the Twitter network of prominent politicians and SPLC-defined hate groups in the 2016 US presidential election. *Social Network Analysis and Mining* **9**, 34 (2019). URL <http://link.springer.com/10.1007/s13278-019-0567-9>.
- [498] Lee, S. & Xenos, M. Incidental news exposure via social media and political participation: Evidence of reciprocal effects. *New Media & Society* **24**, 178–201 (2022). URL <http://journals.sagepub.com/doi/10.1177/1461444820962121>.
- [499] Casero-Ripollés, A., Micó-Sanz, J.-L. & D'Áez-Bosch, M. Digital Public Sphere and Geography: The Influence of Physical Location on Twitters Political Conversation. *Media and Communication* **8**, 96–106 (2020). URL <https://www.cogitatiopress.com/mediaandcommunication/article/view/3145>.
- [500] Kim, Y., Chen, H.-T. & Wang, Y. Living in the Smartphone Age: Examining the Conditional Indirect Effects of Mobile Phone Use on Political Participation. *Journal of Broadcasting & Electronic Media* **60**, 694–713 (2016). URL <https://www.tandfonline.com/doi/full/10.1080/08838151.2016.1203318>.
- [501] Elvestad, E., Phillips, A. & Feuerstein, M. Can Trust in Traditional News Media Explain Cross-National Differences in News Exposure of Young People Online?: A comparative study of Israel, Norway and the United Kingdom. *Digital Journalism* **6**, 216–235 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/21670811.2017.1332484>.
- [502] Steffan, D. & Venema, N. New medium, old strategies? Comparing online and traditional campaign posters for German Bundestag elections, 2013-2017. *European Journal of Communication* **35**, 370–388 (2020). URL <http://journals.sagepub.com/doi/10.1177/0267323120903681>.
- [503] Vraga, E. K. & Tully, M. News literacy, social media behaviors, and skepticism toward information on social media. *Information, Communication & Society* **24**, 150–166 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/1369118X.2019.1637445>.
- [504] Kofi Frimpong, A. N., Li, P., Nyame, G. & Hossin, M. A. The Impact of Social Media Political Activists on Voting Patterns. *Political Behavior* **44**, 599–652 (2022). URL <https://link.springer.com/10.1007/s11109-020-09632-3>.
- [505] Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F. & Meira, W. Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 131–141 (ACM, Barcelona Spain, 2020). URL <https://dl.acm.org/doi/10.1145/3351095.3372879>.
- [506] López-Rabadán, P. & Domínech-Fabregat, H. Nuevas funciones de Instagram en el avance de la política espectáculo. Claves profesionales y estrategia visual de Vox en su despegue electoral. *El profesional de la información* e300220 (2021). URL <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/85530>.

- [507] Gainous, J., Abbott, J. P. & Wagner, K. M. Active vs. Passive Social Media Engagement with Critical Information: Protest Behavior in Two Asian Countries. *The International Journal of Press/Politics* **26**, 464–483 (2021). URL <http://journals.sagepub.com/doi/10.1177/1940161220963606>.
- [508] Skoric, M. M., Zhu, Q. & Lin, J.-H. T. What Predicts Selective Avoidance on Social Media? A Study of Political Unfriending in Hong Kong and Taiwan. *American Behavioral Scientist* **62**, 1097–1115 (2018). URL <http://journals.sagepub.com/doi/10.1177/0002764218764251>.
- [509] Zannettou, S. *et al.* Disinformation Warfare: Understanding State-Sponsored Trolls on Twitter and Their Influence on the Web. In *Companion Proceedings of The 2019 World Wide Web Conference*, 218–226 (ACM, San Francisco USA, 2019). URL <https://dl.acm.org/doi/10.1145/3308560.3316495>.
- [510] Alsaad, A., Taamneh, A. & Al-Jedaiah, M. N. Does social media increase racist behavior? An examination of confirmation bias theory. *Technology in Society* **55**, 41–46 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S0160791X18300630>.
- [511] Nanz, A., Heiss, R. & Matthes, J. Antecedents of intentional and incidental exposure modes on social media and consequences for political participation: a panel study. *Acta Politica* **57**, 235–253 (2022). URL <https://link.springer.com/10.1057/s41269-020-00182-4>.
- [512] Ardèvol-Abreu, A., Hooker, C. M. & Gil de Zúñiga, H. Online news creation, trust in the media, and political participation: Direct and moderating effects over time. *Journalism* **19**, 611–631 (2018). URL <http://journals.sagepub.com/doi/10.1177/1464884917700447>.
- [513] Lu, Y. & Pan, J. Capturing Clicks: How the Chinese Government Uses Clickbait to Compete for Visibility. *Political Communication* **38**, 23–54 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/10584609.2020.1765914>.
- [514] Davidson, B. I., Jones, S. L., Joinson, A. N. & Hinds, J. The evolution of online ideological communities. *PLOS ONE* **14**, e0216932 (2019). URL <https://dx.plos.org/10.1371/journal.pone.0216932>.
- [515] Mahmood, Q. K., Bhutta, M. H. & Haq, M. A. u. Effects of sociodemographic variables and Facebook group membership on students political participation. *Education and Information Technologies* **23**, 2235–2247 (2018). URL <http://link.springer.com/10.1007/s10639-018-9715-5>.
- [516] Kyung Hee University, Min, H., Yun, S. & Kyung Hee University. Selective Exposure and Political Polarization of Public Opinion on the Presidential Impeachment in South Korea : Facebook vs. KakaoTalk. *Korea Observer - Institute of Korean Studies* **49**, 137–159 (2018). URL [http://dcontent.dkyobobook.co.kr/genomad\\_gift/001/article/1/03/81/10381508.pdf](http://dcontent.dkyobobook.co.kr/genomad_gift/001/article/1/03/81/10381508.pdf).
- [517] Yamamoto, M., Kushin, M. J. & Dalisay, F. How informed are messaging app users about politics? A linkage of messaging app use and political knowledge and participation. *Telematics and Informatics* **35**, 2376–2386 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S073658531830412X>.

- [518] Martiñez-Torres, H. & Gámez, C. Is internet access bad news for media-capturing incumbents? *Journal of Applied Economics* **22**, 527–553 (2019). URL <https://www.tandfonline.com/doi/full/10.1080/15140326.2019.1673082>.
- [519] Imran, M. S., Fatima, M. & Kosar, G. Connectivism: E-learning of democratic values on social media public spheres. In *2017 International Conference on Information and Communication Technologies (ICICT)*, 82–89 (IEEE, Karachi, 2017). URL <https://ieeexplore.ieee.org/document/8320169/>.
- [520] Scharkow, M., Mangold, F., Stier, S. & Breuer, J. How social network sites and other online intermediaries increase exposure to news. *Proceedings of the National Academy of Sciences* **117**, 2761–2763 (2020). URL <https://pnas.org/doi/full/10.1073/pnas.1918279117>.
- [521] Kwak, N. *et al.* Perceptions of Social Media for Politics: Testing the Slacktivism Hypothesis. *Human Communication Research* **44**, 197–221 (2018). URL <https://academic.oup.com/hcr/article/44/2/197/4957051>.
- [522] Schmidt, A. L., Zollo, F., Scala, A., Betsch, C. & Quattrocioni, W. Polarization of the vaccination debate on Facebook. *Vaccine* **36**, 3606–3612 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S0264410X18306601>.
- [523] Vaccari, C. & Valeriani, A. Dual Screening, Public Service Broadcasting, and Political Participation in Eight Western Democracies. *The International Journal of Press/Politics* **23**, 367–388 (2018). URL <http://journals.sagepub.com/doi/10.1177/1940161218779170>.
- [524] Sharar, B. & Abd-El-Barr, M. Citizens' Perspective on the Impact of Social Media on Politics in Kuwait. In *2018 International Conference on Computing Sciences and Engineering (ICCSE)*, 1–6 (IEEE, Kuwait City, 2018). URL <https://ieeexplore.ieee.org/document/8374207/>.
- [525] Lee, S. H. & Fu, K.-w. INTERNET USE AND PROTEST POLITICS IN SOUTH KOREA AND TAIWAN. *Journal of East Asian Studies* **19**, 89–109 (2019). URL [https://www.cambridge.org/core/product/identifier/S1598240818000255/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S1598240818000255/type/journal_article).
- [526] Piatak, J. & Mikkelsen, I. Does Social Media Engagement Translate to Civic Engagement Offline? *Nonprofit and Voluntary Sector Quarterly* **50**, 1079–1101 (2021). URL <http://journals.sagepub.com/doi/10.1177/0899764021999444>.
- [527] Baek, Y. M. Political mobilization through social network sites: The mobilizing power of political messages received from SNS friends. *Computers in Human Behavior* **44**, 12–19 (2015). URL <https://linkinghub.elsevier.com/retrieve/pii/S0747563214006190>.
- [528] Ferrucci, P., Hopp, T. & Vargo, C. J. Civic engagement, social capital, and ideological extremity: Exploring online political engagement and political expression on Facebook. *New Media & Society* **22**, 1095–1115 (2020). URL <http://journals.sagepub.com/doi/10.1177/1461444819873110>.

- [529] Corrigan-Brown, C. & Wilkes, R. Media exposure and the engaged citizen: How the media shape political participation. *The Social Science Journal* **51**, 408–421 (2014). URL <https://www.tandfonline.com/doi/full/10.1016/j.sosci.2014.03.009>.
- [530] Valenzuela, S., Correa, T. & Gil de Zúñiga, H. Ties, Likes, and Tweets: Using Strong and Weak Ties to Explain Differences in Protest Participation Across Facebook and Twitter Use. *Political Communication* **35**, 117–134 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/10584609.2017.1334726>.
- [531] Popan, J. R., Coursey, L., Acosta, J. & Kenworthy, J. Testing the effects of incivility during internet political discussion on perceptions of rational argument and evaluations of a political outgroup. *Computers in Human Behavior* **96**, 123–132 (2019). URL <https://linkinghub.elsevier.com/retrieve/pii/S074756321930069X>.
- [532] Ā tÄ>tkA, V., Mazák, J. & Vochocová, L. Nobody Tells us what to Write about: The Disinformation Media Ecosystem and its Consumers in the Czech Republic. *Javnost - The Public* **28**, 90–109 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/13183222.2020.1841381>.
- [533] Kobayashi, T. Depolarization through social media use: Evidence from dual identifiers in Hong Kong. *New Media & Society* **22**, 1339–1358 (2020). URL <http://journals.sagepub.com/doi/10.1177/1461444820910124>.
- [534] Talwar, S., Dhir, A., Kaur, P., Zafar, N. & Alrasheedy, M. Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior. *Journal of Retailing and Consumer Services* **51**, 72–82 (2019). URL <https://linkinghub.elsevier.com/retrieve/pii/S0969698919301407>.
- [535] Yarchi, M., Baden, C. & Kligler-Vilenchik, N. Political Polarization on the Digital Sphere: A Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media. *Political Communication* **38**, 98–139 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/10584609.2020.1785067>.
- [536] Bobba, G. Social media populism: features and likeability of Lega Nord communication on Facebook. *European Political Science* **18**, 11–23 (2019). URL <http://link.springer.com/10.1057/s41304-017-0141-8>.
- [537] Tewksbury, D. & Riles, J. M. Polarization as a Function of Citizen Predispositions and Exposure to News on the Internet. *Journal of Broadcasting & Electronic Media* **59**, 381–398 (2015). URL <http://www.tandfonline.com/doi/full/10.1080/08838151.2015.1054996>.
- [538] Giglietto, F., Righetti, N., Rossi, L. & Marino, G. It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 Italian elections. *Information, Communication & Society* **23**, 867–891 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/1369118X.2020.1739732>.
- [539] Marcinkowski, F. & DoÄ;enoviÄ;P. From incidental exposure to intentional avoidance: Psychological reactance to political communication during the 2017 German national election campaign. *New Media & Society* **23**, 457–478 (2021). URL <http://journals.sagepub.com/doi/10.1177/1461444820902104>.



- [540] Bastien, F., Koop, R., Small, T. A., Giasson, T. & Jansen, H. The role of online technologies and digital skills in the political participation of citizens with disabilities. *Journal of Information Technology & Politics* **17**, 218–231 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/19331681.2020.1742264>.
- [541] Nelson, J. L. & Taneja, H. The small, disloyal fake news audience: The role of audience availability in fake news consumption. *New Media & Society* **20**, 3720–3737 (2018). URL <http://journals.sagepub.com/doi/10.1177/1461444818758715>.
- [542] Lelkes, Y., Sood, G. & Iyengar, S. The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect: EFFECT OF BROADBAND INTERNET ACCESS ON PARTISAN AFFECT. *American Journal of Political Science* **61**, 5–20 (2017). URL <https://onlinelibrary.wiley.com/doi/10.1111/ajps.12237>.
- [543] Hunter, L. Y., Griffith, C. E. & Warren, T. Internet connectivity and domestic terrorism in democracies. *International Journal of Sociology* **50**, 201–219 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/00207659.2020.1757297>.
- [544] Heiss, R. & Matthes, J. Does incidental exposure on social media equalize or reinforce participatory gaps? Evidence from a panel study. *New Media & Society* **21**, 2463–2482 (2019). URL <http://journals.sagepub.com/doi/10.1177/1461444819850755>.
- [545] Bobba, G., Cremonesi, C., Mancosu, M. & Seddone, A. Populism and the Gender Gap: Comparing Digital Engagement with Populist and Non-populist Facebook Pages in France, Italy, and Spain. *The International Journal of Press/Politics* **23**, 458–475 (2018). URL <http://journals.sagepub.com/doi/10.1177/1940161218787046>.
- [546] Troian, J., Arciszewski, T. & Apostolidis, T. The dynamics of public opinion following terror attacks: Evidence for a decrease in equalitarian values from Internet Search Volume Indices. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* **13** (2019). URL <https://cyberpsychology.eu/article/view/12015>.
- [547] Frissen, T. Internet, the great radicalizer? Exploring relationships between seeking for online extremist materials and cognitive radicalization in young adults. *Computers in Human Behavior* **114**, 106549 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0747563220302995>.
- [548] Park, B., Kang, M. Y. & Lee, J. Sustainable Political Social Media Marketing: Effects of Structural Features in Plain Text Messages. *Sustainability* **12**, 5997 (2020). URL <https://www.mdpi.com/2071-1050/12/15/5997>.
- [549] Yun, G. W., Park, S.-Y., Holody, K., Yoon, K. S. & Xie, S. Selective Moderation, Selective Responding, and Balkanization of the Blogosphere: A Field Experiment. *Media Psychology* **16**, 295–317 (2013). URL <http://www.tandfonline.com/doi/abs/10.1080/15213269.2012.759462>.

- [550] Allen, J., Howland, B., Mobius, M., Rothschild, D. & Watts, D. J. Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances* **6**, eaay3539 (2020). URL <https://www.science.org/doi/10.1126/sciadv.aay3539>.
- [551] Gainous, J., Abbott, J. P. & Wagner, K. M. Traditional Versus Internet Media in a Restricted Information Environment: How Trust in the Medium Matters. *Political Behavior* **41**, 401–422 (2019). URL <http://link.springer.com/10.1007/s11109-018-9456-6>.
- [552] Boulianne, S., Koc-Michalska, K. & Bimber, B. Mobilizing media: comparing TV and social media effects on protest mobilization. *Information, Communication & Society* **23**, 642–664 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/1369118X.2020.1713847>.
- [553] Waechter, N. The participative role of social media for the disadvantaged young generation in the Arab Spring. *Österreichische Zeitschrift für Soziologie* **44**, 217–236 (2019). URL <http://link.springer.com/10.1007/s11614-019-00356-1>.
- [554] Wang, T. & Shen, F. Perceived party polarization, news attentiveness, and political participation: a mediated moderation model. *Asian Journal of Communication* **28**, 620–637 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/01292986.2018.1494204>.
- [555] Guess, A., Nagler, J. & Tucker, J. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances* **5**, eaau4586 (2019). URL <https://www.science.org/doi/10.1126/sciadv.aau4586>.
- [556] Kaur, M. & Verma, R. Demographics, Social Media Usage, and Political Engagement in Punjab. *Indian Journal of Marketing* **48**, 43 (2018). URL <http://www.indianjournalofmarketing.com/index.php/ijom/article/view/137984>.
- [557] Castillo-DÁaz, A. & Castillo-Esparcia, A. Relación entre la participación en foros y blogs de debate político en internet y el seguimiento de información política en medios profesionales: análisis evolutivo 2011-2016. *El Profesional de la Información* **27**, 1248 (2018). URL <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2018.nov.08>.
- [558] Heiss, R., Knoll, J. & Matthes, J. Pathways to political (dis-)engagement: motivations behind social media use and the role of incidental and intentional exposure modes in adolescents political engagement. *Communications* **45**, 671–693 (2020). URL <https://www.degruyter.com/document/doi/10.1515/commun-2019-2054/html>.
- [559] Enjolras, B., Steen-Johnsen, K. & Wollebæk, D. Social media and mobilization to offline demonstrations: Transcending participatory divides? *New Media & Society* **15**, 890–908 (2013). URL <http://journals.sagepub.com/doi/10.1177/1461444812462844>.

- [560] Bhat, S. I., Arif, T., Malik, M. B. & Sheikh, A. A. Browser simulation-based crawler for online social network profile extraction. *International Journal of Web Based Communities* **16**, 321 (2020). URL <http://www.inderscience.com/link.php?id=111377>.
- [561] Aruguete, N., Calvo, E. & Ventura, T. News Sharing, Gatekeeping, and Polarization: A Study of the #Bolsonaro Election. *Digital Journalism* **9**, 1–23 (2021). URL <https://www.tandfonline.com/doi/full/10.1080/21670811.2020.1852094>.
- [562] Bae, S. Y. The social mediation of political rumors: Examining the dynamics in social media and belief in political rumors. *Journalism* **21**, 1522–1538 (2020). URL <http://journals.sagepub.com/doi/10.1177/1464884917722657>.
- [563] Nikolov, D., Lalmas, M., Flammini, A. & Menczer, F. Quantifying Biases in Online Information Exposure. *Journal of the Association for Information Science and Technology* **70**, 218–229 (2019). URL <https://onlinelibrary.wiley.com/doi/10.1002/asi.24121>.
- [564] Akpan, I. J. *et al.* Association Between What People Learned About COVID-19 Using Web Searches and Their Behavior Toward Public Health Guidelines: Empirical Infodemiology Study. *Journal of Medical Internet Research* **23**, e28975 (2021). URL <https://www.jmir.org/2021/9/e28975>.
- [565] Yamamoto, M., Nah, S. & Bae, S. Y. Social media presumption and online political participation: An examination of online communication processes. *New Media & Society* **22**, 1885–1902 (2020). URL <http://journals.sagepub.com/doi/10.1177/1461444819886295>.
- [566] Song, T., Tang, Q. & Huang, J. Triadic Closure, Homophily, and Reciprocation: An Empirical Investigation of Social Ties Between Content Providers. *Information Systems Research* **30**, 912–926 (2019). URL <http://pubsonline.informs.org/doi/10.1287/isre.2019.0838>.
- [567] Strauß, N., Huber, B. & Gil de Zúñiga, H. Yes, I Saw It - But Didnt Read It... A Cross-Country Study, Exploring Relationships between Incidental News Exposure and News Use across Platforms. *Digital Journalism* **8**, 1181–1205 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/21670811.2020.1832130>.
- [568] Ahmad, S. Political behavior in virtual environment: Role of social media intensity, internet connectivity, and political affiliation in online political persuasion among university students. *Journal of Human Behavior in the Social Environment* **30**, 457–473 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/10911359.2019.1698485>.
- [569] Lee, H. Voters involvement, attitude, and confidence in the era of new media. *Palgrave Communications* **6**, 1 (2020). URL <http://www.nature.com/articles/s41599-019-0368-9>.
- [570] Čábelková, I., Smutka, L. & Strielkowski, W. Public support for sustainable development and environmental policy: A case of the Czech Republic. *Sustainable Development* **30**, 110–126 (2022). URL <https://onlinelibrary.wiley.com/doi/10.1002/sd.2232>.

- [571] Lu, J. & Yu, X. Does The internet make us more intolerant? A contextual analysis in 33 countries. *Information, Communication & Society* **23**, 252–266 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/1369118X.2018.1499794>.
- [572] David, Y. Public opinion, media and activism: the differentiating role of media use and perceptions of public opinion on political behaviour. *Social Movement Studies* **21**, 334–354 (2022). URL <https://www.tandfonline.com/doi/full/10.1080/14742837.2021.1875321>.
- [573] Back, E. A., Back, H., FredÃ©n, A. & Gustafsson, N. A social safety net? Rejection sensitivity and political opinion sharing among young people in social media. *New Media & Society* **21**, 298–316 (2019). URL <http://journals.sagepub.com/doi/10.1177/1461444818795487>.
- [574] Ye, Y., Xu, P. & Zhang, M. Social media, public discourse and civic engagement in modern China. *Telematics and Informatics* **34**, 705–714 (2017). URL <https://linkinghub.elsevier.com/retrieve/pii/S0736585316302908>.
- [575] Salman, A., Yusoff, M. A., Mohamad Salleh, M. A. & Hj Abdullah, M. Y. PENGUNAAN MEDIA SOSIAL UNTUK SOKONGAN POLITIK DI MALAYSIA [THE USE OF SOCIAL MEDIA FOR POLITICAL SUPPORT IN MALAYSIA]. *Journal of Nusantara Studies (JONUS)* **3**, 51 (2018). URL <https://journal.unisza.edu.my/jonus/index.php/jonus/article/view/222>.
- [576] Mueller, K. & Schwarz, C. Fanning the Flames of Hate: Social Media and Hate Crime. *Journal of the European Economic Association* **19**, 2131–2167 (2021). URL <https://academic.oup.com/jeea/article/19/4/2131/5917396>.
- [577] Bursztyn, L., Egorov, G., Enikolopov, R. & Petrova, M. Social Media and Xenophobia: Evidence from Russia. Tech. Rep. w26567, National Bureau of Economic Research, Cambridge, MA (2019). URL <http://www.nber.org/papers/w26567.pdf>.
- [578] Lee, C., Shin, J. & Hong, A. Does social media use really make people politically polarized? Direct and indirect effects of social media use on political polarization in South Korea. *Telematics and Informatics* **35**, 245–254 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S0736585317305208>.

## **P2. More than News! Mapping the Deliberative Potential of Political Online Ecosystems with Digital Trace Data.**

### **Preprint Publication**

Oswald, L. (2022). More than News! Mapping the Deliberative Potential of Political Online Ecosystems with Digital Trace Data.

<https://osf.io/preprints/socarxiv/68udg/>

### **Supplementary Materials**

Supplementary information including documented code, available data and supplementary tables and figures available under <https://osf.io/atj5u/>.

# More than News! Mapping the Deliberative Potential of a Political Online Ecosystem with Digital Trace Data


Lisa Oswald  
Hertie School

## Abstract

Does the internet facilitate everyday public deliberation? Most of the previous research on this question has focused on specific aspects, such as online news media diets or political discussions on social media platforms. However, increasingly complex media environments are composed of different arenas with different respective potential for democracy. Using digital trace data from Germany, I map out which websites are relevant for the online public discourse, I develop a measure of deliberative potential along six dimensions (information, communication, participation, connectivity, inclusivity and heterogeneity), and explore different types of websites alongside high-level usage patterns. Besides a class of mainstream informational hubs, I find a class of quality information providers, such as public broadcasting outlets, that are not characterized by a diverse user base. A third class of niche online forums hosts political discussions among more tightly-knit online communities, supporting previous findings of incidental exposure to political content online. While the mainstream information hubs in the sample attract a much larger volume of clicks, users spend relatively more time consuming political information on quality information sites as well as on niche online forums to engage with politics online. This project takes a more holistic perspective of the diverse ecosystem of online deliberation, while presenting a first quantitative exploration of a deliberative system.

*Keywords:* Online environments, political information, political communication, deliberative potential, digital trace data, latent class analysis, deliberative systems

---

Lisa Oswald  <https://orcid.org/0000-0002-8418-282X>

Correspondence concerning this article should be addressed to Lisa Oswald, Hertie School, Friedrichstraße 180, 10117 Berlin, Germany. E-mail: [l.oswald@phd.hertie-school.org](mailto:l.oswald@phd.hertie-school.org)

## Introduction

The question as to what extent the internet enables (or hinders) public deliberation is a much disputed issue that has, so far, only partially been addressed and from rather specific angles. Researchers with a focus on social media platforms have considered active user communication by analyzing online discussion threads (e.g. Esau et al., 2020; Halpern & Gibbs, 2013), whereas researchers with an interest in online news media diets, for example, have examined web browsing histories with a distinct news media frame (Guess, 2021).

The academic discourse concerning the extent of homophily and selective exposure in so-called online ‘echo chambers’ (Pariser, 2011; Sunstein, 2002) considerably diverges between disciplines and methodological approaches. While within platform studies are finding robust evidence for homophily in social networks (Cinelli et al., 2021; Cota et al., 2019; Guerrero-Solé & Lopez-Gonzalez, 2019; Koiranen et al., 2019; Rivero, 2019), between platform studies are increasingly finding evidence of a diversity of exposure (Fletcher & Nielsen, 2018; Guess, 2021; Lelkes, 2020; Strauss et al., 2020; Yang et al., 2020). Facing this dispute, a more holistic consideration of online arenas for public discourse can help avoid the underestimation of exposure while, at the same time, avoiding the overestimation of siloed information. In line with theorists of deliberative democracy (e.g. Bächtiger & Parkinson, 2019), I argue that the political usage of the internet or the engagement with political topics online goes far beyond the categories of news media consumption and social media discussion but should be measured more holistically, by mapping the diverse ecosystem of online deliberation.

In this project, I link and expand upon existing streams of research on online communication and information, and focus on the *deliberative potential* of websites as the structural basis for a constructive online public discourse. I regard bridging those streams of research as a necessary condition for a systematic and systemic assessment of the online public sphere (Bächtiger & Parkinson, 2019). I address the following three research questions:

1. Which websites hold potential for online public discourse, including political information consumption and discussion online?
2. How is the political online ecosystem structured along infrastructural and usage char-

acteristics?

3. How does the interplay between user demographics and different classes of websites look like?

Using digital trace data from Germany in combination with survey data and manual content coding to characterize a wide range of politically relevant websites, I empirically tackle various questions of the online public sphere for the first time. With passive web tracking, the data collection is not geared towards one specific platform or website type (e.g. news media), but provides a more complete picture of online behavior, which is crucial for gaining a more holistic and realistic perspective of the online public sphere. I consider the deliberative potential of websites as a latent construct which is in line with an understanding of deliberation as the summative quality of a deliberative system in which different sites fulfill different democratic functions (Bächtiger & Parkinson, 2019; Elstub et al., 2019; Esau et al., 2020). Furthermore, a latent classification of websites goes beyond a xylographic distinction between news sites and social media platforms, for example. Using a broad initial scope together with a latent approach, one does not risk overlooking important, potentially overlapping arenas in which political communication is taking place and where people get their information from online.

Overall, I find that only a small proportion of online activity (1%) is concerned with politics. To the disappointment of social scientists, the majority of people spend most of their time accessing various forms of entertainment, shopping and work-related URLs. However, the vast majority of users in the sample (1,190 out of 1,282 individuals) engaged with some political content during the six-month observation period that included the 2017 national elections in Germany. Originally starting with a web tracking dataset of more than 56 million website visits, without setting a predefined exclusion frame for the type of website and using automated approaches together with manual cross validation, the analysis is narrowed down to a set of 69 central domains featuring content on a wide range of political topics relevant to the German public discourse.

Besides a cluster of highly-popular ‘mainstream’ sites that are visited by a broad range of users to read and discuss political information, I find a cluster of public broadcasting



and journalistic news outlets, the ‘quality information providers’ that cover the highest density of political information. However, they are not characterized by a diverse user base. I find a third cluster of niche online forums, often dedicated to specific topics or communities, that are mostly neglected in current analyses of the online public sphere. Manual cross checking confirmed that they host in-depth political discussions among sometimes tightly-knit online communities such as fan forums. While the mainstream sites in the sample attract a much larger volume of clicks, users spend relatively more time consuming political information via public broadcasting and online news outlets as well as on niche online forums to discuss politics online.

In other words, while the German deliberative system seems to be a rather small fraction of the wider online environment, the consumption of political content is not as exclusive as the visible discussion patterns of few very active users on social media may imply. While mainstream platforms are most central to the topical flow of political information consumption within the system, the latent structure of deliberative potential highlights the persisting relevance of high quality public broadcasting as the backbone for democratic deliberation in Germany. Niche online forums provide particular potential for mutual trust building through prior exchanges about shared a-political interests as potential basis for the deliberation of conflicting political views.

### **Deliberation in Online Environments**

Online communication has often been connected to an increase in affective political polarization (Bail, 2021; Rathje et al., 2021), the spread of misinformation (Vosoughi et al., 2018) and the rise of radical counter publics (Bright, 2018; Douglas et al., 2017; Lorenz-Spreen et al., 2021).

However, in theory, constructive discussions among informed citizens should help to identify the best arguments for complex societal questions and therefore mitigate opinion polarization (Grönlund et al., 2015; Habermas, 1984; Ugarriza & Caluwaerts, 2014). While more and more deliberation researchers are concentrating their research efforts in the area of online discussions (Strandberg & Grönlund, 2018), contradicting evidence is emerging on the quality of online discussions - negative in many regards (Anderson et al., 2014; Coe

et al., 2014; Sunstein, 2002; Ziegele et al., 2020) with some positive exceptions, for example when it comes to discussions in the comments section of online newspapers (Manosevitch & Walker, 2009; Rowe, 2015).

There is a nuanced empirical discourse around the measurement of deliberative quality, including some widely-established coding schemes (Steenbergen et al., 2003) and the development of novel, more inclusive criteria of deliberative quality (Graham, 2008, 2012). Additional concepts discussed in the field are for example story-telling (Esau et al., 2020; Gerber et al., 2018), humor (Basu, 1999; Coleman & Moss, 2012), emotions (Krause, 2008), power (Follesdal, 2010) and the role of non-verbal communication for deliberative democracy (Mendonça et al., 2020). However, there is little empirical investigation of the infrastructural foundation of online deliberation—the potential for deliberation supplied online by websites providing political information and discussion spaces.

A promising but today mostly theoretical development in the field are systemic perspectives on deliberation. Deliberative systems theory, that appears particularly applicable in the digital domain, argues that different arenas fulfill different functions for democracy (Bächtiger & Parkinson, 2019; Ercan et al., 2017; Mansbridge et al., 2012). However, the empirical conceptualization of the deliberative *potential* of websites as the basis for a constructive public discourse remains largely unresolved. Even though online political deliberation might be a niche phenomenon rather than mainstream behavior, it is crucial to understand its structural foundation. Beauchamp (2020) describes the deliberativeness of discussions in online environments as a function of membership and structure. In this project, I empirically map these structures, an ecosystem of politically relevant websites, as the foundation of an online deliberative system.

### **Deliberative Potential of Websites**

While the theoretical term ‘deliberative potential’ is not a novel concept in the field (Conover et al., 2002; Mendonça & Ercan, 2015), the deliberative potential of websites has, so far, only been explored theoretically or with regard to specific domains. For example, Wiklund (2005) analyzed different municipal websites in Sweden with a focus on two dimensions, information services and communication services provided by the websites.

In contrast, Richardson and Stanyer (2011) examined British online news outlets. They consider manifest characteristics of websites while still keeping a focus on communicative features such as online forums and the deliberative quality of communication.

I root the assessment of deliberative potential in the theory of deliberative democracy and describe the six dimensions of the concept in detail below. The notion that ‘different types of public online deliberation can be expected to display different characteristics and fulfill different functions in democratic opinion and will formation, as well as in decision making.’ (Esau et al., 2020, p. 2) has especially influenced the definition of deliberative potential used throughout this project. The dimensions of deliberative potential were structured along three core functional dimensions that are determined by the infrastructure of the website as the basis for deliberation (see Table 1). Three additional higher-level dimensions are defined by the respective usage patterns or demand-side characteristics. For example, I regard a forum that enables reciprocity in communication as necessary basis for online deliberation. However, in line with theories of deliberative democracy (Bächtiger & Parkinson, 2019; Habermas et al., 1974) only the consideration of heterogeneous arguments within an inclusive debate that is accessible for a diverse citizenship can make a discussion truly deliberative. This conception is not particular to the online sphere. Already in face-to-face citizen assemblies, the deliberative quality as well as the outcomes of deliberation depend on who is deliberating, regardless of the venue’s infrastructure (Warren, 2021).

Therefore, in this project, I map the structural pre-conditions for deliberative discourse, including both infrastructural aspects and patterns of how this infrastructure is used. In doing so, I stay on the level of describing the deliberative potential of online environments, rather than assessing the substance of the discourse.

**Information.** The first dimension in the assessment of the deliberative potential of websites is the provision of relevant information. In 1789, Thomas Jefferson had already concluded that well informed citizens are the key to a healthy democracy (Jefferson, 1789), a notion that still constitutes a core feature of deliberative democracy (Bächtiger & Parkinson, 2019; Fishkin, 2018). Information on parties, policies, institutions and procedures are the building blocks of political knowledge and are extensively researched concepts in the social sciences for good reason (Carpini & Keeter, 1993; Prior, 2005). Civic knowledge

about institutions and processes can help citizens to better understand their interests as individuals and members of a group, it increases the consistency of views across issues and over time, and it increases trust, political participation and support for democratic values, such as tolerance for the needs of minorities (e.g. Galston, 2001). Moving the perspective to the digital era, in the context of online deliberation, websites on which users find political information can serve as a resource for opinion and will formation (Esau et al., 2020). While information on political parties and issues potentially influence political opinions and inform vote choices, administrative information help citizens to understand democratic structures and procedures (Döring, 2021).

Drawing upon both empirical findings on political and civic knowledge (Carpini & Keeter, 1993; Munzert & Selb, 2017) as well as previous research on the role of political information for deliberation (Wiklund, 2005), three distinct criteria are included for the assessment to what extent a site provides relevant political information. I assess whether the site provides (1) information on political actors, institutions and political issues, (2) information on administrative procedures and local information, and (3) whether information provided by the site is journalistically curated or has, at least, undergone some other form of fact checking (such as e.g. on Wikipedia). Local information, for example on local initiatives and regulations is integrated into the category with administrative information, because they have the similarly enabling potential for civic engagement and political participation. The third information criterion serves as a basic manifest proxy for information quality.

Measuring the exposure to diverse news media is one important component to assess the informative potential of the online sphere for a functioning democracy. Previous projects focused on news access through social media sites, for example (Bakshy et al., 2015; Barberá et al., 2015; Eady et al., 2019), which however risks neglecting less mainstream media outlets. Other studies, collecting data from the user perspective risk suffering from biased self-reports in surveys (Boxell et al., 2017; Lelkes, 2016). Facing these methodological challenges, web tracking data provide unique insights into real-life media diets. While Guess (2021) illuminates important aspects of online news media diets, for example, I aim to capture the overall deliberative potential of the complex online public sphere using an

even more inclusive scope.

**Communication.** It is important to note that this project does not consider the deliberative quality or the content of discussions taking place on a certain website, when looking at the communication dimension of deliberative potential. Instead, I assess whether the website provides users with the possibility to express and/or exchange political opinions with other users (Wiklund, 2005). Such sites can serve as communicative spaces for interactional opinion and will formation (Esau et al., 2020). I split the dimension into two levels of communication. First, I measure whether the website enables the expression of political opinions through the provision of comment sections, for example. In a second step, I assess whether the site fulfills the deliberative core criterion of potential reciprocity (Steenbergen et al., 2003). Communicative situations can only be characterized by reciprocity, if users have the option of replying to each other. Therefore, websites only fulfill the second criterion if a reference to previous comments is enabled, for example in online forums and on social media platforms, but also in comment sections of online news papers where ‘reply’ functions are enabled. The sole possibility to up-vote or down-vote comments, or to react to comments through ‘likes’ is not regarded as reciprocity. In doing so, I do not assess actual communication or specific elements such as listening to the arguments of others (Kriplean et al., 2012; Scudder, 2020), but I assess the structural foundation for deliberative communication.

**Participation.** Websites that provide users with the possibility of online political participation can have a more or less direct impact on political decision-making or, at least, serve as a platform for the aggregation of interests (Esau et al., 2020).

It is a disputed issue whether a link to decision-making is necessary to consider political communication as deliberation. While Thompson (2008) insists on the link to decision-making, the idea of deliberative polls (Fishkin et al., 2018), the Habermasian idea of diffuse communication in the public sphere as well as the deliberative systems approach adopt a broader definition of deliberation. By mapping the deliberative potential of the online ecosystem, I include opportunities for online political participation as desirable features of online political ecosystems without making a judgment about the definition of deliberation itself.

To assess the structural foundation of the link to decision-making, I examine three distinct criteria: (1) whether the website hosts petitions and/or opinion polls (Richardson & Stanyer, 2011) to collect, aggregate or organize public opinion, (2) whether the website enables citizens to get in contact with political actors (Wiklund, 2005), and (3) whether the website enables the political organization of citizens, for example by the formation of political interest groups or events such as discussion forums, demonstrations and other forms of political protest. Wiklund (2005) considered some of these criteria under the framework of the communicative services of a website. However, it might be worth distinguishing between forums for discussions among citizens and communicative acts that can have a more direct influence on political decision-making.

While I consider the first three dimensions: information, communication and participation, as core dimensions of the deliberative potential of a platform, three additional criteria are assessed: connectivity, inclusiveness and heterogeneity that are defined through usage patterns and user characteristics.

**Connectivity.** The dimension of connectivity considers whether the website is connected to other politically relevant websites. These connections can, for example, enable further research by citizens on political issues or facilitate the implementation of intention to participate in the political process. Following the conceptualization of deliberative systems, an understanding of the links and flows between different sites is crucial for a systemic understanding of the online public discourse (Dryzek, 2012; Esau et al., 2020; Fleuß et al., 2018). For example, Fleuß et al. (2018) emphasized the transmissions between different loci as being an important aspect to measure deliberation in a systemic way. They proposed tracking the transmission of topics as they evolve within the system as well as tracking individuals who transmit ideas from one locus to another. While the analysis mainly operates within the arena of informal deliberation, the approach to operationalize connectivity, by tracking users' subsequent visits to different websites featuring the same topics, gets very close to their theoretical idea of formalizing transmissions.

A body of literature outside the field of deliberation research that examines those links and flows between different online sites can be found in the field of inter-media agenda-setting research. For example, media researchers have analyzed to what extent certain online

**Table 1**

*Deliberative Potential Criteria*

<b>Dimension</b>	<b>Operational Definition</b>	<b>Criteria</b>	<b>Measurement</b>
Information	Users can find political information on this website. Such sites serve as resource for opinion and will-formation.	<ol style="list-style-type: none"> <li>1. Information on political issues, actors and institutions</li> <li>2. Administrative or local information</li> <li>3. Primary or (journalistically) curated source of information</li> </ol>	Human rating (binary 0/1) for each criterion
Communication	The website provides users with the possibility to express and/or exchange political opinions with other users. Such sites serve as communicative spaces for (interactional) opinion and will-formation.	<ol style="list-style-type: none"> <li>1. Enables commenting / political expression (and potentially ratings of comments)</li> <li>2. Enables reciprocity / replies to comments of other users (open replies, not only ratings)</li> </ol>	Human rating (binary 0/1) for each criterion
Participation	The website provides users with the possibility of online political participation or organization, implying a potential (direct) impact on political decision making.	<ol style="list-style-type: none"> <li>1. Enables contact to political actors</li> <li>2. Enables political participation (petitions, polls, etc.)</li> <li>3. Enables political organization (events, groups, etc.)</li> </ol>	Human rating (binary 0/1) for each criterion
Connectivity	The website is connected to other relevant websites. This facilitates further research on political issues or the implementation of intentions of political participation.	<ol style="list-style-type: none"> <li>1. From this website users move to other relevant websites</li> <li>2. Users land on this website after visiting other relevant websites</li> </ol>	Network centrality, in/outgoing relevant traffic
Inclusivity	The website is used by a comparably diverse set of individuals. This dimension serves as indicator for low barriers of access.	<ol style="list-style-type: none"> <li>1. Educational backgrounds</li> <li>2. Gender</li> <li>3. Age</li> </ol>	Quantitative assessment using web-tracking data in combination with survey information
Heterogeneity	The website connects people holding a comparably wide range of political opinions. This stands in contrast to the notion of ideological online ‘echo chambers’.	<ol style="list-style-type: none"> <li>1. Political orientation</li> <li>2. Party preferences</li> </ol>	Quantitative assessment using web-tracking data in combination with survey information

*Note.* Curation represents a basic form of fact checking taking place (no obvious misinformation or strongly opinionated statements presented as factual statements). Connectivity, inclusivity and heterogeneity are optional categories as only possible to assess with digital trace data + survey information. Criteria developed with reference to Esau et al. (2020), Mansbridge et al. (2012), Richardson and Stanyer (2011), and Wiklund (2005).

publics are able to stimulate mass media publics (Messner & Distaso, 2008; Wallsten, 2007), while others have examined news diffusion processes from a temporal perspective (Buhl et al., 2018), or the Twitter networks of journalists as important nodes between digital and mass media (Neuberger et al., 2019).

In this course, Digital trace data provides a unique opportunity to explore linkages between websites through the observation of real-life online behavior with network-analytical quantities. At the time of the study and to the best of my knowledge, this project is the first to empirically formalize the connectivity of the different arenas of an online deliberative system.

**Inclusivity.** The dimension of inclusivity appears to be an increasingly prominent aspect in the scientific discourse on deliberation. Mansbridge et al. (2012) describe three overall functions of a deliberative system: (1) an epistemic function to produce appropriately informed preferences and opinions, which I consider under the dimension of information, (2) an ethical function that creates respect between experts and citizens (these aspects could, for example, be a consequence of communication among citizens and contact between citizens and political actors, both captured in the dimensions of communication and participation), and (3) a democratic function. Mansbridge et al. (2012) conceptualize the democratic function as promoting ‘an inclusive political process in terms of equality’ (p. 12), which implies the inclusion of multiple and plural voices.

I explicitly consider the demographic variables of gender, age and educational background in the assessment of inclusivity, to answer the question whether a website is used by a diverse set of individuals. This dimension can further serve as indicator for low barriers of access. The unique combination of digital trace data with high-quality survey data allows a precise measurement of this dimension of deliberative potential.

**Heterogeneity.** An interesting aspect, also implied in the conceptualization of the democratic functions of a deliberative system by Parkinson and Mansbridge (2012) is the inclusion of a variety of interests, concerns and claims. I consider this aspect under the dimension of heterogeneity and assess whether the website connects people holding a wide range of political opinions. This concept stands in contrast to the notion of ideological online ‘echo chambers’ (Pariser, 2011; Sunstein, 2002) in which users are mainly surrounded by



similar others, holding opinion-reinforcing views. In contrast to other researchers, who focused on the heterogeneity of information diets in online environments (Bright et al., 2020; Dubois & Blank, 2018; Guess, 2021), I consider the composition of users that visit a website. More specifically, I capture the heterogeneity of political orientations of users within a platform through their explicit indication of political orientation on a Left-to-Right scale and their expressed party preferences in the context of the German federal election in 2017.

Both dimensions, inclusivity and heterogeneity are rooted in representation literature (Warren, 2021). Random sampling would, under ideal experimental conditions with high compliance, ensure inclusivity and heterogeneity. Under natural conditions in online environments, the issues of inclusivity and heterogeneity as basis for discursive representation are more complex. I consider the concepts as theoretically distinct, as inclusivity builds on manifest demographic variables of the user whereas heterogeneity is a more latent construct of political attitudes and party preferences that possibly links more closely to political content featured online.

### **Methodological Challenges**

Around the beginning of the millennium, Steyaert (2000) had already emphasized the need for analytical tools that enable a systematic empirical analysis of digital democracy. However, most research in the field is still being conducted as explorative case studies, mostly with a focus on the content or the deliberative quality of communicative acts online (Felicetti et al., 2016; Jensen, 2003; Jonsson, 2015; Pedrini, 2014). Also the rooting theorist of deliberative democracy and the concept of the public sphere, Jürgen Habermas, explicitly emphasized the methodological challenge of empirically examining online deliberation in a recent piece on the restructuring of the public sphere in the digital age (Habermas, 2021). The conceptualization of deliberation as the emergent property of a system, involving the dynamics of contexts and platform design elements with different functions for democracy (Esau et al., 2017), comes with serious questions for empirical research (Boswell & Corbett, 2017; Fleuß et al., 2018; Niemeyer et al., 2015).

Based on the current state of the empirical literature, I identify two key method-

ological challenges in the study of online deliberation. First, given the ever-growing online landscape, it is crucial to know *where* on the web different branches of the public discourse are taking place in order to make assumptions about their content and quality. The question as to which websites are used for political issues is not trivial as deliberation in online environments is getting increasingly pluralistic (Esau et al., 2017; Janssen & Kies, 2005) and incidental news exposure occurs regularly (Feezell, 2018; Kim et al., 2013; Tewksbury et al., 2001; Yadamsuren & Erdelez, 2010).

Second, most empirical research in the field of on online deliberation, especially the assessment of the deliberative quality of communication, is researched on individual (active) behavior. However, most people on the web are passive consumers of content, also called ‘lurkers’, (Nonnecke & Preece, 1999; Sun et al., 2014). This passive majority does not leave obvious digital traces in online forums and comment sections but they certainly do obtain political information from the web that shape their political opinions and actions. In the context of social media platforms, for example, passive users also experience social learning and constantly adapt their understandings of social norms by observing other people’s communication, while a highly active minority, also called ‘power users’ (Bright et al., 2019), creates the majority of content online.

In consequence, it remains largely unclear how this silent majority uses the web for political issues facing a heavy focus on communication data. It would be important to also examine passive exposure instead, to meaningfully define the boundaries of the public arena online. The question as to what extent websites *enable* public deliberation, under the further consideration of a systemic perspective, is what I aim to answer with the assessment of the deliberative potential of websites.

I would like to stress that deliberative potential does not directly imply deliberation. Online environments may provide accurate political information but also misinformation, they may enable deliberative discussion but also host toxic exchanges, they may provide platforms for civic engagement but also hostile participation (Freudenthaler & Wessler, 2022; Quandt, 2018). However, if the fundamental conditions of deliberative potential are not met in the infrastructure and usage of online environments, deliberation itself is impossible.

## Method

The analysis is based on digital trace data that was collected within a six-month period in Germany, including the national elections in September 2017. The data is linked to rich survey data, including users' demographics, political attitudes and other relevant political variables. This unique combination of two data sources allows for the assessment of the deliberative potential of websites enriched by components that only become measurable in the interaction of user characteristics and usage behavior (connectivity and especially, inclusivity and heterogeneity).

## Data

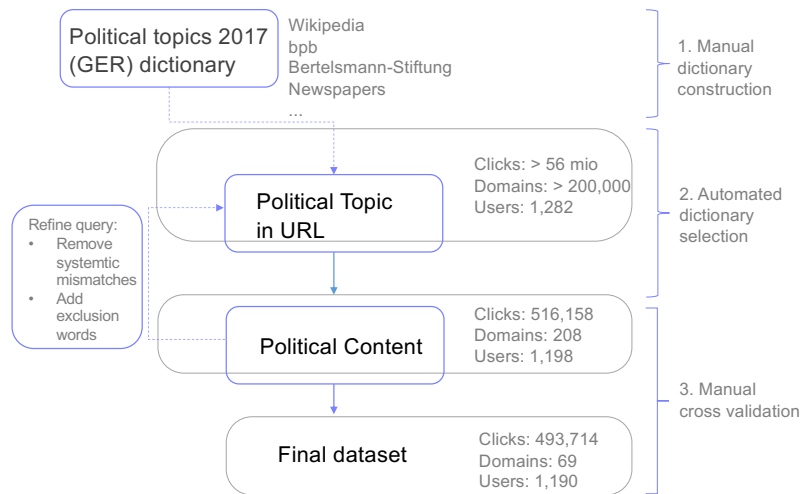
The survey data was collected via the German YouGov Pulse panel with survey questions originally fielded to 1,500 respondents in five waves. Using a quota-sampling procedure on the basis of the marginals from Best for Planning (2017), the sample mirrors the German online population with respect to gender, age and, to some degree, education. Respondents were asked to install a URL tracking software that uses passive metering technology to record detailed browser histories on an opt-in basis. Tracking could be paused for 15 minutes and respondents could end their participation at any time. This digital trace data includes more than 56 million website visits to almost 200,000 different domains by 1,282 different individuals between July 2017 and December 2017. As this paper draws on data collected in a prior research project, details on the sampling procedure, the deployment of the passive metering software as well as privacy and ethical considerations can be found in SI B. The collection of survey and tracking data used in this study was approved by the IRBs of Princeton University (protocols 8327, 10014, and 10041) and the University of Southern California (UP-17-00513) and authorized by the University of Illinois via a designated IRB agreement.

In a first step, the top 1,000 domains were manually classified into categories (see Table C12). Those 1,000 domains account for about 83% of website visits. This is characteristic for web browsing data, in which central websites accumulate most activity while the majority of sites are only visited by very few users. The pre-labelled web tracking data was then merged with the survey data to allow for the description of the demographic profile of

the sample<sup>1</sup>.

**Measures**

To develop a measurement for the deliberative potential of websites, a reference sample of websites had to be identified first. Considering the context of deliberative democracy, a focus on websites that, in the wider sense, play a role in the online public discourse or feature political content appeared the most appropriate. The approach to be particularly inclusive in the first step sets this project apart from previous research, for example on online news media diets (Guess, 2021), that also use URL browsing histories but focuses only on news websites. I argue that, in order to gain a realistic picture of the online public arena, it is important to consider all possible channels through which political information, communication and participation is enabled, especially because the exposure to political information makes up only a small proportion of users’ total online engagement.



**Figure 1**

*Topic driven reference sample selection process.*

Exploiting the fact that the data almost mirror the German online public demo-

<sup>1</sup>The distribution of the self-reported political orientation of the sample approaches a normal distribution and also geographically, online activity patterns in the sample distribute about evenly across Germany.

graphically and include individuals' browser histories for a period of about six months around the federal elections in Germany, websites accessed in a 'political context' were selected with a combination of automated keyword search and manual cross checking (see Fig. 1). The relevance of these websites for the online public discourse in Germany in 2017 was then approximated using the number of website visits on the respective domain, aggregated across the sample, as a measure of engagement. The process of political website identification is described in detail in the SI A.

In a second step, the deliberative potential of those politically relevant websites was determined. To this end, I assessed the outlined six dimensions of deliberative potential: information, communication, participation, connectivity, inclusivity and heterogeneity (see Table 1). While the first three dimensions were evaluated with manual content coding only, the latter three were determined through a consideration of digital trace data in combination with linked survey data.

**Content coding.** In order to assess the first three core dimensions of the concept of deliberative potential, the full sample of relevant websites was assessed using quantitative content analysis on the websites' infrastructure. The theoretical definitions of the dimensions were translated into operational definitions including concrete criteria that could be assessed using a binary rating system (0 representing 'not present', 1 representing 'present'). The unit of analysis was one website domain and the coding was conducted after manually visiting the website and assessing the overall structure of the page, posts, articles, and comment sections. A standardized code book (see Table 1), including all dimensions and sub-criteria was used to streamline the coding process.

**Digital trace data.** The availability of web tracking data in combination with survey data allows the enrichment of the manual assessment of deliberative potential with granular quantitative measures of online behavior. I used this micro-level behavioral data to assess criteria on a more macro level, the unit being websites rather than single users. I constructed the connectivity measure through network analytical measures of in-going and out-going traffic, and I added the measures of inclusivity, based on demographic variables, and heterogeneity, based on the political preferences of users.

More specifically, in order to exploit the benefits of digital trace data to build the

connectivity measure, I construct a network of website visits with websites represented as nodes, and temporally subsequent website visits for one user, featuring the same topic, represented as edges. For example, if a user reads an online newspaper article featuring the name ‘Merkel’ in the URL and, following this, visits a social media discussion featuring ‘Merkel’, an edge is created between the nodes of the online newspaper and the social media platform. I only counted subsequent visits to different websites, while I excluded self-loops from the network. This way, instead of considering the ‘dead’ hyperlink-infrastructure from the html text of the websites, I created a measure of actual in-going and out-going politically-relevant traffic for each website. These traffic flows can be quantified using the network analytical measures of in-degrees (in-going traffic) and out-degrees (out-going traffic) (see Fig. C4 in the SI).

To assess the degree of inclusivity of a website, I calculated three distinct diversity indicators for each website for the variables age, gender and education. I opted for the widely-used entropy-based Shannon-Wiener diversity index (see SI A), as it is implemented in R (Dixon, 2003; Grafton et al., 2012; Kiernan, 2014; Oksanen, 2013). A high inclusivity means that a website is accessed by individuals from different age groups, education levels or genders. The more different categories (for example age groups) and the more similar the engagement levels of those different groups, the higher the estimated inclusivity value of a particular website.

For the construction of heterogeneity criteria, I followed a similar approach. I applied the diversity assessment to a variable measuring the political orientation of participants on a Left-to-Right scale and to their reported first votes in the 2017 federal election in Germany. According to the Shannon-Wiener diversity index, the heterogeneity of a website is comparably high if it is visited equally by individuals with different political orientation.

**Clustering websites with latent class analysis.** After the assessment of all six dimensions of deliberative potential of websites, I considered patterns of commonalities and differences between websites to explore different ‘profiles’ of deliberative potential. In line with the latent understanding of deliberative potential, I followed a latent class approach to identify groups of websites according to their deliberative potential. Besides this theoretical reason, an examination of the empirical relationships between different criteria, suggests the

use off a latent composite measure as there are both, correlations within, but also between different dimensions of deliberative potential (see Fig. C5). More details on the latent class modeling approach can be found in the SI A. Finally, after the identification of classes, an individual class membership prediction value was assigned to each website, allowing the categorization of websites into latent classes.

## Results

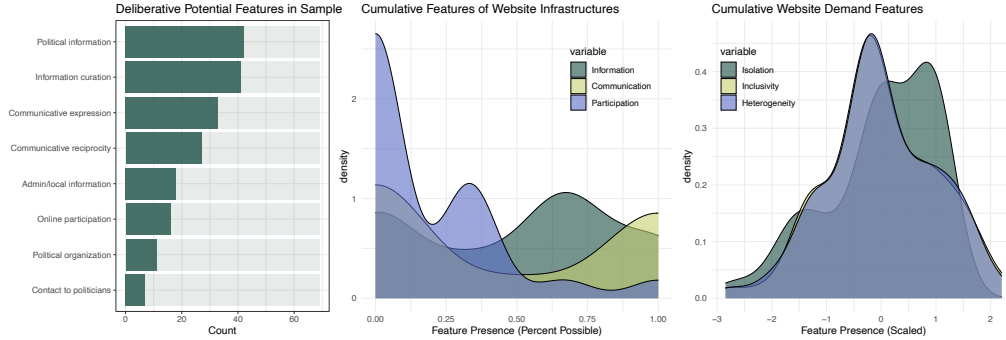
### Politically relevant sites in Germany in 2017

Applying the two-stage process of website selection, consisting of (1) the automated dictionary-based classification of websites as ‘politically relevant’ and (2) the following manual cross validation, 69 central domains were identified in the sample that have played a role in the online public discourse in Germany in the second half of 2017. Those websites were visited by 1,190 unique users, which included a large proportion of the original sample ( $N = 1,282$ ). It is important to note that this does not mean that, for example because highly frequented websites such as ‘Google’ and ‘Facebook’ are part of this set of 69 websites, those 1,190 individuals simply used those platforms at least once in the six-month period. Instead, it means that they ‘googled’ some political keyword or visited political content on Facebook because the political filtering step took place *before* the compression of website visits into domains.

Starting from the original sample of more than 56 million tracked website visits, less than 1% (493,714 clicks) were politically relevant visits to those 69 domains. Table C12 summarizes the big picture of the overall online activity of the sample, illustrating that the engagement with political issues is not the dominant motive for many users to use the web. In contrast, the most frequented websites were social media platforms and search engines (mostly for apolitical content), email providers, online shopping, gaming, streaming, porn and online banking.

Furthermore, only slightly more than half of the politically relevant websites (52%) in the sample are explicitly labeled as news websites, and only 12% of the websites featuring political discussions are social media platforms.

**Deliberative Potential of Sites**

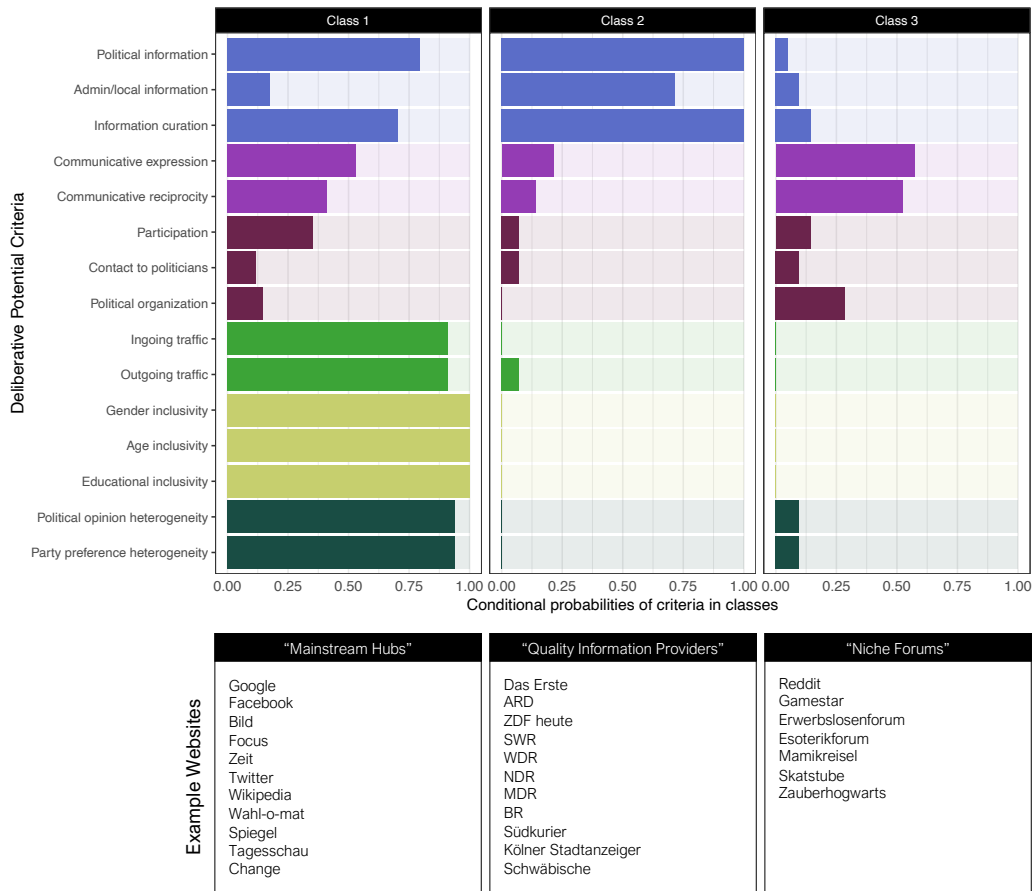


**Figure 2**

*Deliberative Feature Descriptives. Left: How many of the 69 websites fulfill criteria? Center: How does the cumulative feature presence look like for the three infrastructural criteria? E.g. most websites fulfill 0 out of 4 participation criteria, 2 out of 3 information criteria and either 2/2 or 0/2 communication criteria. Right: How does the cumulative feature presence look like for the three usage based criteria? Metrics scaled for better comparability.*

Figure 2 summarizes the descriptives on the prevalence of deliberative potential criteria across the sample of politically relevant websites. While the majority of websites fulfills two out of three information criteria (most provide political information that underlie some form of journalistic curation or fact checking), only very few websites fulfill the criteria of participation. When it comes to the potential to host political discussions, about half of the platforms provide the possibility to express *and* discuss political opinions online while the other half *neither* enables expression nor reciprocity in communication. Only very few platforms enable the expression of political opinions in the form of comment sections without the possibility to reply to other comments. Considering the ‘demand side’ characteristics of demographic inclusivity and political opinion heterogeneity within websites’ user bases, both measured with the entropy-based Shannon-Wiener diversity index (more details see SI A), there is a considerable overlap of density distributions. Websites attracting users of diverse age groups, genders and education levels appear to also attract users of diverse political orientations and party preferences. The distributions of both measures, though highly correlated with the overall engagement on a website, does not mirror the rather left-skewed metric of connectivity that reflects engagement links and flows between politically relevant platforms.





**Figure 3**

*Conditional response probabilities, by deliberative potential criterion, of belonging to each latent class. Based on response probability patterns and class membership, class 1 was named 'mainstream hubs', class 2 was named 'quality information providers' and class 3 was named 'niche forums'. Full list of domains provided in SI Table C1*

To structure the political online environment along the complex set of deliberative potential criteria, a latent class analysis was conducted. Considering various model fit criteria and rounds of validation, a model with three latent classes was selected (see Fig. C6, Table C2 and more description in the SI).

### Latent Class Structure of the Online Ecosystem

Figure 3 summarizes the conditional probabilities of websites belonging to each of the three latent classes dependent on their fulfillment of each of the deliberative potential criteria. It also present exemplary sets of websites that were previously identified as politically relevant and sorted into the three estimated latent classes based on their respective predicted probabilities of class membership (See SI Table C1 for the full lists).

In summary, websites in class 1, from now on referred to as the ‘mainstream hubs’, show especially high class-conditional probabilities of fulfilling the dimensions of connectivity, inclusivity and heterogeneity, while websites assigned to class 2, the ‘quality information providers’ appear strong with regard to information criteria. Websites assigned to class 3, the ‘niche forums’ show rather low class-conditional probabilities for most criteria of deliberative potential, except for the communication dimension and political organization.

More specifically, the class of mainstream hubs (class 1) is composed of a diverse set of websites that fulfill the core criteria of information, communication and participation to some extent but which are especially characterized by a high degree of connectivity, demographic inclusivity and political opinion heterogeneity. Overall, those websites have the highest level of engagement measured by the number of website visits in the sample. Such sites are, for example, prominent high quality national newspapers like ‘Zeit’ and ‘Spiegel’, more tabloid outlets like ‘Bild’, social media platforms like ‘Facebook’ or ‘Twitter’, but also sites with particular functions, such as the online petitioning platform ‘Change’ or the voting advice application ‘Wahl-O-Mat’. What most of the websites in this class have in common is that they are highly-frequented websites that are nationally well known and relevant for political content across diverse German-speaking audiences.

The quality information providers (class 2) include almost exclusively established local, regional and national online news outlets and informative TV channels hosted by public service broadcasting with the exception of ‘RTL’ and ‘Sat1’, two private TV channels with broad online news sections. While ‘ARD’ is the leading national public service broadcasting channel in Germany, ‘MDR’, ‘WDR’, ‘SWR’ and ‘NDR’ are their regional channels. Websites like ‘Südkurier’ and ‘KStA’ (Kölner Stadtanzeiger) are examples of large regional and local news outlets, while ‘Berlin’ is the information platform hosted by the Berlin municipal

government. All of those sites provide high quality, journalistically-curated information, often with specific local focus, but apparently, neither do they offer extensive possibilities for political discussion, nor do they attract attention from diverse audiences.

Finally, the class of niche forums (class 3) contains websites with rather low conditional probabilities of fulfilling explicit criteria of deliberative potential, except for the potential of political expression and reciprocity in communication, and potential for political organization. In this class, I found rather niche online forums for specific communities, or forums that are dedicated to specific topics like gaming, cooking or anime content. While many domains in this class do not appear politically relevant at first glance, it is important to note that I manually validated whether political discussion were indeed taking place on those websites. Examples of websites in class 3 are an esoteric forum that vividly discussed the upcoming federal elections, computer forums in which I found discussions on the military intervention in Afghanistan, a forum for children's second hand clothing ('Mamikreisel') and a forum dealing with issues of unemployment ('Eloforum') that hosted, partly in-depth, political discussions in niches of the forum.

In total, 34 websites were assigned to the mainstream hubs (class 1), 20 belong to the quality information providers (class 2) and 25 to the class of niche forums (class 3)<sup>2</sup>. The estimated mixing proportions  $p_r$  corresponding to the share of observations belonging to each latent class are 49% for the mainstream hubs, 22% for the information providers and 29% for the niche forums.

The input criteria of deliberative potential form two natural groups: information, communication and participation are criteria that were coded manually and belong to the supply side of a website whereas connectivity, inclusivity and heterogeneity are coded computationally based on usage characteristics. This fundamental distinction is also reflected in the correlation-matrix between criteria. I therefore also repeated the clustering process separately for the two groups of criteria (see SI C9 and C11). For the computationally-coded, demand-side criteria, a simple two factor solution was suggested with one class including all websites with high probabilities of fulfilling each criterion and one class with overall very low scores for connectivity, inclusivity and heterogeneity - in other words, high and low

---

<sup>2</sup>The order of classes has no deeper meaning but is determined by configurations in the estimation process.

engagement websites. The model including only the manually-coded infrastructural criteria of information, communication and participation possibilities suggested a more interesting pattern that is in line with the findings from the main model including all criteria. A first class contains websites with an strong information profile, including all public broadcasting pages. A second class contains websites with an especially strong forum component or communication profile with pages that also enable participation to some extent. The last class is rather a residual class including websites with overall low probabilities of fulfilling any criteria. The overall pattern largely mirrors the findings from the main model, the difference being that the two meaningful classes of the infrastructural model also contain the highly popular mainstream hubs that are, in the main model, separated through distinct patterns in the engagement based metrics.

### **Engagement with different classes of sites**

Our measurement framework for the assessment of the deliberative potential of websites could of course be applied to various contexts for analytical and practical purposes. As one application, I considered simple user-level engagement patterns, measured in the number of website visits as well as the duration of engagement.

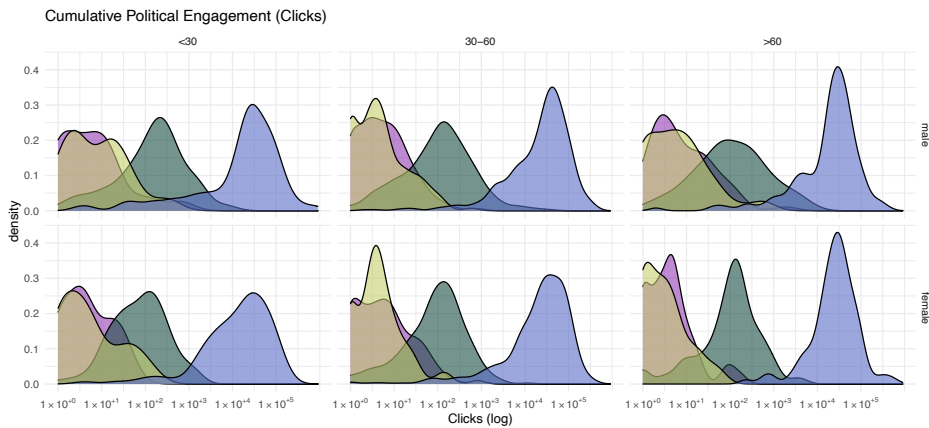
Given the underlying latent structure of deliberative potential dimensions, it does not surprise that the mainstream hubs are more frequently<sup>3</sup> accessed than quality information providers and niche forums (see Fig. C1a). However, if engagement is measured as duration instead of clicks, the engagement distributions become more similar (see Fig. C1b). This implies that people often access prominent websites like Google and Facebook in political contexts but that they tend to spend more time on public broadcasting platforms as well as small online forums to read news more carefully and, potentially, discuss political issues in depth within tighter communities compared to major social media platforms. More specifically, the duration per click ratio is only 35 seconds for mainstream hubs, on average, but 48 seconds for niche forums and almost a minute (59 seconds) for quality information providers. If the data were to be split, for example, just into news websites and social media

---

<sup>3</sup>Cumulative engagement measures are baseline corrected, meaning that they show the share of website visits that users spend on e.g. quality information providers in relation to their total number of website visits in the measurement time frame.

platforms, this pattern would not have been observed (see Fig. C2a and C2b).

Another application is to switch from the perspective of the ‘supply side’ characteristics to the ‘demand side’ characteristics, namely the demographics of users engaging with different classes of sites. Figs. 4 and C3 summarize the engagement with different classes of sites for different genders, age groups and levels of formal education. Despite some minor, though intuitive tendencies (e.g. the engagement with quality information providers is stronger than the engagement with niche online forums in the subgroup with the highest level of formal education (Abitur) in Germany) there is no clear pattern of selection visible within subgroups according to those three rough demographic indicators. The exploration of more sophisticated variables such as political orientation, political efficacy or political knowledge as possible driving factors for the selection into engaging with political content online remains subject to a subsequent project.



**Figure 4**

*Engagement with different classes of sites by age group and gender. Class 1: ‘mainstream hubs’, class 2: ‘quality information providers’ and class 3: ‘niche forums’.*

## Discussion

The deliberative nature of an online environment is, as Beauchamp (2020) puts it, a function of membership and structures. In order to examine this function empirically, as a first step, this project systematically mapped the deliberative potential of those structures for the online public sphere in Germany.

While political online engagement only makes up a small proportion (about 1% of website visits) of the overall online engagement, a large part of the sample (1,190 out of 1,282) did engage with some political topics at least at some point around the federal elections in 2017. It is worth noting that the website selection approach, including a strict manual cross validation of whether a website actually featured political content, focuses on the minimization of false positives rather than false negatives. This implies quite a strict definition of ‘politically relevant’ and tends to rather underestimate the prevalence of political engagement online. However, possibly to the disappointment of many social scientists, engagement with political content online is by no means the dominant form of engagement.

The results of the study clearly align with Guess (2021) who found a considerable overlap of news media diets within a US sample that goes against the common notion of selective exposure in online ‘echo chambers’. According to Guess (2021), this overlap originates from individuals’ common use of large mainstream hubs for political information. Correspondingly, in the sample, the largest cluster of websites are indeed highly-frequented sites that are commonly visited by a large proportion of users. These informational hubs can be understood to be a kind of general-interest intermediary that may indeed facilitate a common arena within the digital public arena that offers shared experiences and the possibility of incidental encounters with diverse perspectives (Sunstein, 2018).

The results of the latent class analysis further suggest that public service broadcasting still plays a major role in the online public discourse in Germany even though these websites did not reach a particularly diverse audience within the sample. Furthermore, the reference to local or regional issues and information is a commonality of many websites assigned to the class of quality information providers. The local nature of political issues is often neglected when studying political online communication or when using digital trace data that do not have a geospatial component. However, on an interesting side note, Ellger et al. (2021) find that the decline of local newspapers can be related to an increase in political polarization, a relationship that could be given more attention in the study of online politics. While digital technology does let information flows transcend physical constraints, people still live in specific local contexts.

Furthermore, the analysis points us to a latent class of websites that is only mentioned in a small proportion of empirical studies on online deliberation. Wright (2012) coined the term ‘third spaces’ for non-political online spaces where political talk emerges based on case studies, similar to Graham (2012). This study demonstrates the importance of their early work on a much larger basis. The class contains mostly niche forums dedicated to specific topics and communities which points to the phenomenon of incidental exposure to political issues online (Valeriani & Vaccari, 2016; Yadamsuren & Erdelez, 2010). Furthermore, these online communities might be comparably more tightly knit because of shared (apolitical) interests and fewer overall user numbers, which allows individuals to recognize each other (despite usually being pseudonymous, Moore et al. (2020)). These forums, which, in comparison to large social media platforms, might be closer to offline social groups in which a basic form of trust can be established between members, can provide interesting possibilities for informal political discussions among citizens and might operate as important ‘weak ties’ between large online information and communication platforms within a deliberative online system (Esau et al., 2017; Granovetter, 1973; S. W. Rosenberg, 2014).

As visible among the mainstream hubs, website popularity is heavily ensconced in the three additional dimensions of deliberative potential (connectivity, inclusivity and heterogeneity). One obvious reason for this finding is that the degree of centrality of a node in a social network increases with the frequency of its interactions. Another measurement related explanation could be that the Shannon-Wiener diversity index puts more weight on richness than on evenness (Zeleny, 2021), implying a rising index with more users. Therefore, caution must be taken against a substantive interpretation of the finding that the most heavily used platforms in the sample are, according to the measures, also the most ‘inclusive’ and ‘heterogeneous’. While they are indeed a common source of information and a common arena of political communication for citizens with different demographic profiles and heterogeneous political attitudes, it is still important to keep in mind that this does not prevent the formation of niche corners and sub-groups that might not speak to each other.

While the study focuses on the concept of deliberative *potential*, in order to allow claims about the deliberative *quality* and the substance of discussions, a much more in-

depth analysis of the communicative acts on the subset communication platforms would be necessary. This, however, was not the objective of this contribution, but the findings could be highly informative for future analyses of platforms' deliberative quality. A discrepancy between the deliberative potential and the actual quality measured on the platform may indicate a form of 'unused potential', for example.

Another limitation to consider is that when classifying the content of the sites as political or not, I considered the full URL-string. While this often features the most important keywords of the page accessed, in other cases, scraping the entire HTML text of the site might have been helpful. For computational efficiency reasons, I used the URL based method for this work. Furthermore, an extremely robust scraper would have to be built in order to process hundreds of thousands of different domain structures (in the original full dataset). Future researchers could try to build such a scraper, web-scrape all the sites and search for political topics in the full HTML text of websites instead of the URL-strings. The reference body (sites explicitly dealing with the 2017 German public discourse that I selected to generate keywords, see SI D) would then be similar enough to the target body (now being the full-text of websites instead of URL-text only) to use the semi-automated keyword extraction method proposed by King et al. (2017).

The manually selected set of keywords naturally comes with certain boundary conditions. It is systematically easier to rigorously identify specific political terms, such as the names of politicians and terms referring to party politics and administrative processes in comparison to political issues like education and social policy because terms like 'family' or 'housing' appear in many different political and apolitical contexts. I implemented various efforts to reduce this imbalance (see SI D) but still, cannot be sure to have eliminated it entirely.

Finally, the web tracking data is based on desktop use and does not include mobile devices. This certainly overlooks parts of users' political online engagement and may even introduce non-random blind spots. Furthermore, due to the temporal asymmetry between the browser histories (collected in 2017) and the content analysis on the respective websites (conducted in 2021), I cannot rule out the possibility that some websites might have changed in terms of structure, content and function for the online public discourse.



### Conclusion

Increasingly complex media environments are composed of different arenas with different potential functions for democracy. While most of the previous research focused on specific aspects, such as online news media diets or the content of discussions on social media platforms, I took a step back and examined the infrastructure and usage patterns as the basis for online deliberation.

This project illustrates that the internet provides a plethora of sources for political information, arenas for political communication and some opportunities for online participation. I clearly found the *potential* for public deliberation in the German speaking web in 2017. Even though political content is only a small proportion of the overall content accessed online—the German deliberative system seems to be a rather small fraction of the wider online environment—almost everyone in the sample engaged with some political content around the federal election in 2017. This implies that the consumption of political content is not as exclusive as the visible discussion patterns of few very active users on social media may imply.

The infrastructure of a deliberative system goes far beyond news websites and social media platforms but includes wide range of different types of popular and niche platforms with different primary functions. On some platforms, users get political information. However, it is not clear if those are accurate or misinformation. On other platforms, they can discuss political issues, deliberatively or not. While only very few websites in the sample offer possibilities for participation, the demand also seemed limited.

Mainstream hubs are most central in the network of topical links, whereas public broadcasting outlets and especially the nice forums are more at the periphery of the network. Considering the definition of links within the connectivity measure, this implies that users move beyond the quick bites of political information on mainstream platforms but read more on the topic elsewhere. Those platforms appear to act as general-interest intermediary that may indeed facilitate a common arena within the digital public sphere that, against the notion of online ‘echo chambers’, offers shared experiences and the possibility of encounters with diverse perspectives.

The class of information providers can be interpreted as evidence for the persisting

centrality of high quality public broadcasting as the backbone for democratic deliberation in Germany. The question as to whether we stand at the beginning or the end of the public broadcasting era online could be determined using detailed information on the user base.

This project demonstrated the presence of a-political spaces in which political discussion emerges on a large empirical basis. While niche online forums are especially characteristic for the earlier years of the internet, it will be interesting to see in which spaces more tightly knit online communities will form in the future as previous exchange around a-political shared interests may build mutual trust as important basis for the discussion of conflicting political views.

While this project empirically mapped the online media structures underlying online deliberation for the first time, the logical next step in the research agenda is the quantitative description of membership, the profiles of internet users engaging with political information and communication online. In particular because online political deliberation itself may not be a mainstream behavior, the mechanisms of selection into the online public discourse need to be determined in the future.

### **Data availability**

Extensive supplementary material, including all R scripts and publicly available data, can be found in the project's repository on OSF under <https://osf.io/atj5u/>.

### **Supplementary Information**

Supplementary materials, including supporting tables and figures, the dictionary used for website selection and a software statement can be found under <https://osf.io/wn798>.

### **Acknowledgements**

I thank Simon Munzert, Pablo Barberá, Andrew Guess and JungHwan Yang for the provision of the data and the inspiring and constructive discussions during the 2022 MEOF workshop in Princeton. I thank Simon Munzert for his continuous guidance and supervision on the project, as well as André Bächtiger and all anonymous reviewers for their helpful

comments to the manuscript. The collection of data used in this study was generously funded by a grant from the Volkswagen Foundation Computational Social Science Initiative, reference 92 143.

### References

- Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The “nasty effect:” Online incivility and risk perceptions of emerging technologies [Publisher: Oxford University Press Oxford, UK]. *Journal of computer-mediated communication*, *19*(3), 373–387.
- Auguie, B., & Antonov, A. (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. Retrieved December 22, 2021, from <https://CRAN.R-project.org/package=gridExtra>
- Bacher, J., & Vermunt, J. K. (2010). Analyse latenter Klassen. In C. Wolf & H. Best (Eds.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (pp. 553–574). VS Verlag für Sozialwissenschaften. Retrieved September 27, 2021, from [https://doi.org/10.1007/978-3-531-92038-2\\_22](https://doi.org/10.1007/978-3-531-92038-2_22)
- Bächtiger, A., & Parkinson, J. (2019). *Mapping and measuring deliberation: Towards a new deliberative quality*. Oxford University Press.
- Bail, C. A. (2021). *Breaking the Social Media Prism*. Princeton University Press.
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132.
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science*, *26*(10), 1531–1542.
- Basu, S. (1999). Dialogic ethics and the virtue of humor. *Journal of Political Philosophy*, *7*(4), 378–403. <https://doi.org/10.1111/1467-9760.00082>
- Beauchamp, N. (2020). Modeling and Measuring Deliberation Online. In *The Oxford Handbook of Networked Communication*.
- Best for Planning. (2017). *Berichtsband b4p 2017*. Gesellschaft für integrierte Kommunikationsforschung.

- Boswell, J., & Corbett, J. (2017). Why and how to compare deliberative systems. *European Journal of Political Research*, 56(4), 801–819.
- 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.
- Bright, J. (2018). Explaining the emergence of political fragmentation on social media: The role of ideology and extremism. *Journal of Computer-Mediated Communication*, 23(1), 17–33.
- Bright, J., Bermudez, S., Pilet, J.-B., & Soubiran, T. (2019). Power users in online democracy: Their origins and impact. *Information, communication & society*, 1–16.
- Bright, J., Marchal, N., Ganesh, B., & Rudinac, S. (2020). Echo Chambers Exist!(But They're Full of Opposing Views). *arXiv preprint arXiv:2001.11461*.
- Buhl, F., Günther, E., & Quandt, T. (2018). Observing the dynamics of the online news ecosystem: News diffusion processes among german news sites. *Journalism studies*, 19(1), 79–104.
- Carpini, M. X. D., & Keeter, S. (1993). Measuring political knowledge: Putting first things first. *American Journal of Political Science*, 1179–1206.
- Cinelli, M., de Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences of the United States of America*, 118(9). <https://doi.org/10.1073/pnas.2023301118>
- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, 64(4), 658–679.
- Coleman, S., & Moss, G. (2012). Under construction: The field of online deliberation research.
- Conover, P. J., Searing, D. D., & Crewe, I. M. (2002). The Deliberative Potential of Political Discussion. *British Journal of Political Science*, 32(1), 21–62. <https://doi.org/10.1017/S0007123402000029>

- Cota, W., Ferreira, S., Pastor-Satorras, R., & Starnini, M. (2019). Quantifying echo chamber effects in information spreading over political communication networks. *EPJ Data Science*, 8(1). <https://doi.org/10.1140/epjds/s13688-019-0213-9>
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695.
- Dahl, D. B. (2016). *Xtable: Export Tables to LaTeX or HTML*.
- Díaz, E., & Koutra, C. (2013). Evaluation of the persuasive features of hotel chains websites: A latent class segmentation analysis. *International Journal of hospitality management*, 34, 338–347.
- Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, 14(6), 927–930.
- Döring, M. (2021). How-to Bureaucracy: A Concept of Citizens' Administrative Literacy. *Administration & Society*, 0095399721995460. <https://doi.org/10.1177/0095399721995460>
- Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017). The Psychology of Conspiracy Theories. *Current Directions in Psychological Science*, 26(6), 538–542. <https://doi.org/10.1177/0963721417718261>
- Dryzek, J. S. (2012). *Foundations and frontiers of deliberative governance*. Oxford University Press.
- Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729–745.
- Eady, G., Nagler, J., Guess, A., Zilinsky, J., & Tucker, J. A. (2019). How many people live in political bubbles on social media? Evidence from linked survey and Twitter data. *Sage Open*, 9(1), 2158244019832705.
- Ellger, F., Hilbig, H., Riaz, S., & Tillmann, P. (2021). Local Newspaper Decline and Political Polarization.
- Elstub, S., Ercan, S. A., & Mendonça, R. F. (2019). *Deliberative systems in theory and practice*. Routledge.

- Ercan, S. A., Hendriks, C. M., & Boswell, J. (2017). Studying public deliberation after the systemic turn: The crucial role for interpretive research. *Policy & Politics*, *45*(2), 195–212.
- Esau, K., Fleuß, D., & Nienhaus, S.-M. (2020). Different Arenas, Different Deliberative Quality? Using a Systemic Framework to Evaluate Online Deliberation on Immigration Policy in Germany. *Policy & Internet*.
- Esau, K., Friess, D., & Eilders, C. (2017). Design Matters! An Empirical Analysis of Online Deliberation on Different News Platforms. *Policy & Internet*, *9*(3), 321–342. <https://doi.org/https://doi.org/10.1002/poi3.154>
- Federal Agency for Civic Education. (2017). *Jahresrückblick 2017: Bpb*. Bundeszentrale für politische Bildung. <https://www.bpb.de/politik/hintergrund-aktuell/261923/rueckblick-2017>
- Feezell, J. T. (2018). Agenda setting through social media: The importance of incidental news exposure and social filtering in the digital era. *Political Research Quarterly*, *71*(2), 482–494.
- Felicetti, A., Niemeyer, S., & Curato, N. (2016). Improving deliberative participation: Connecting mini-publics to deliberative systems. *European Political Science Review*, *8*(3), 427–448.
- Fishkin, J. S. (2018). *Democracy When the People Are Thinking P: Revitalizing Our Politics Through Public Deliberation*. Oxford University Press.
- Fishkin, J. S., Senges, M., Donahoe, E., Diamond, L., & Siu, A. (2018). Deliberative polling for multistakeholder internet governance: Considered judgments on access for the next billion. *Information, Communication & Society*, *21*(11), 1541–1554.
- Fletcher, R., & Nielsen, R. (2018). Are people incidentally exposed to news on social media? A comparative analysis. *New Media and Society*, *20*(7), 2450–2468. <https://doi.org/10.1177/1461444817724170>
- Fleuß, D., Helbig, K., & Schaal, G. S. (2018). Four parameters for measuring democratic deliberation: Theoretical and methodological challenges and how to respond. *Politics and Governance*, *6*(1), 11–21.

- Follesdal, A. (2010). The place of self-interest and the role of power in the deliberative democracy. *Journal of political philosophy*, 18(1), 64–100.
- Freudenthaler, R., & Wessler, H. (2022). Mapping Emerging and Legacy Outlets Online by Their Democratic Functions—Agonistic, Deliberative, or Corrosive? [Publisher: SAGE Publications Inc]. *The International Journal of Press/Politics*, 27(2), 417–438. <https://doi.org/10.1177/19401612211015077>
- Galston, W. A. (2001). Political knowledge, political engagement, and civic education. *Annual review of political science*, 4(1), 217–234.
- Gerber, M., Bächtiger, A., Shikano, S., Reber, S., & Rohr, S. (2018). Deliberative abilities and influence in a transnational deliberative poll (EuroPolis). *British Journal of Political Science*, 48(4), 1093–1118.
- Grafton, R. Q., Nelson, H. W., Lambie, N. R., & Wyrwoll, P. R. (2012). *Shannon–Wiener diversity index*. Edward Elgar Publishing Limited. Retrieved June 28, 2021, from [https://www.elgaronline.com/view/nlm-book/9781849803878/c\\_3011.xml](https://www.elgaronline.com/view/nlm-book/9781849803878/c_3011.xml)
- Graham, T. (2008). Needles in a haystack: A new approach for identifying and assessing political talk in nonpolitical discussion forums. *Javnost-the public*, 15(2), 17–36.
- Graham, T. (2012). Beyond “political” communicative spaces: Talking politics on the Wife Swap discussion forum. *Journal of Information Technology & Politics*, 9(1), 31–45.
- Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 78(6), 1360–1380.
- Grönlund, K., Herne, K., & Setälä, M. (2015). Does enclave deliberation polarize opinions? *Political Behavior*, 37(4), 995–1020.
- Guerrero-Solé, F., & Lopez-Gonzalez, H. (2019). Government Formation and Political Discussions in Twitter: An Extended Model for Quantifying Political Distances in Multiparty Democracies. *Social Science Computer Review*, 37(1), 3–21. <https://doi.org/10.1177/0894439317744163>
- Guess, A. M. (2021). (Almost) Everything in Moderation: New Evidence on Americans’ Online Media Diets. *American Journal of Political Science*.
- Habermas, J. (1984). *The Theory of Communicative Action*. Beacon Press.

- Habermas, J. (2021). Überlegungen und Hypothesen zu einem erneuten Strukturwandel der politischen Öffentlichkeit. In *Ein neuer Strukturwandel der Öffentlichkeit?* (pp. 470–500). Nomos Verlagsgesellschaft mbH & Co. KG.
- Habermas, J., Lennox, S., & Lennox, F. (1974). The public sphere: An encyclopedia article (1964). *New German Critique*, (3), 49–55.
- Halpern, D., & Gibbs, J. (2013). Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior*, 29(3), 1159–1168.
- Huling, J. (2019). Jcolors: Colors Palettes for R and 'ggplot2', Additional Themes for 'ggplot2'. Retrieved December 22, 2021, from <https://CRAN.R-project.org/package=jcolors>
- Janssen, D., & Kies, R. (2005). Online forums and deliberative democracy. *Acta política*, 40(3), 317–335.
- Jefferson, T. (1789). Thomas Jefferson to Richard Price - Thomas Jefferson | Exhibitions - Library of Congress. Retrieved July 6, 2021, from <https://www.loc.gov/exhibits/jefferson/60.html>
- Jensen, J. L. (2003). Public spheres on the Internet: Anarchic or government-sponsored—A comparison. *Scandinavian political studies*, 26(4), 349–374.
- Jonsson, M. E. (2015). Democratic innovations in deliberative systems—the case of the Estonian citizens' assembly process. *Journal of Public Deliberation*, 11(1), 7.
- Keyes, O., Jacobs, J., Schmidt, D., Greenaway, M., Rudis, B., Pinto, A., Khezzadeh, M., Meilstrup, P., Costello, A., Bezanson, J., et al. (2019). *Urltools: Vectorised tools for URL handling and parsing*.
- Kiernan, D. (2014). Quantitative Measures of Diversity, Site Similarity, and Habitat Suitability. In *Natural Resources Biometrics*. Open SUNY Textbooks. Retrieved June 28, 2021, from <https://milnepublishing.geneseo.edu/natural-resources-biometrics/chapter/chapter-10-quantitative-measures-of-diversity-site-similarity-and-habitat-suitability/>



- Kim, Y., Chen, H.-T., & De Zúñiga, H. G. (2013). Stumbling upon news on the Internet: Effects of incidental news exposure and relative entertainment use on political engagement. *Computers in human behavior*, *29*(6), 2607–2614.
- King, G., Lam, P., & Roberts, M. (2017). Computer-Assisted Keyword and Document Set Discovery from Unstructured Text. *American Journal of Political Science*, *61*(4), 971–988.
- Koiranen, I., Koivula, A., Keipi, T., & Saarinen, A. (2019). Shared contexts, shared background, shared values – Homophily in Finnish parliament members’ social networks on Twitter. *Telematics and Informatics*, *36*, 117–131. <https://doi.org/10.1016/j.tele.2018.11.009>
- Krause, S. R. (2008). *Civil passions*. Princeton University Press.
- Kriplean, T., Toomim, M., Morgan, J., Borning, A., & Ko, A. J. (2012). Is this what you meant? Promoting listening on the web with reflect. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1559–1568.
- Lelkes, Y. (2016). Mass polarization: Manifestations and measurements. *Public Opinion Quarterly*, *80*(S1), 392–410.
- Lelkes, Y. (2020). A bigger pie: The effects of high-speed Internet on political behavior. *Journal of Computer-Mediated Communication*, *25*(3), 199–216.
- Lin, T. H., & Dayton, C. M. (1997). Model selection information criteria for non-nested latent class models. *Journal of Educational and Behavioral Statistics*, *22*(3), 249–264.
- Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of statistical software*, *42*(10), 1–29.
- Long, J. A. (2021). Panelr: Regression Models and Utilities for Repeated Measures and Panel Data. Retrieved December 22, 2021, from <https://CRAN.R-project.org/package=panelr>
- Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2021). Digital Media and Democracy: A Systematic Review of Causal and Correlational Evidence Worldwide.

- Manosevitch, E., & Walker, D. (2009). Reader comments to online opinion journalism: A space of public deliberation. *International Symposium on Online Journalism*, 10, 1–30.
- Mansbridge, J., Bohman, J., Chambers, S., Christiano, T., Fung, A., Parkinson, J., Thompson, D. F., & Warren, M. E. (2012). A systemic approach to deliberative democracy. In J. Parkinson & J. Mansbridge (Eds.), *Deliberative Systems* (pp. 1–26). Cambridge University Press.
- Menchen-Trevino, E. (2013). Collecting vertical trace data: Big possibilities and big challenges for multi-method research. *Policy & Internet*, 5(3), 328–339.
- Mendonça, R. F., & Ercan, S. A. (2015). Deliberation and protest: Strange bedfellows? Revealing the deliberative potential of 2013 protests in Turkey and Brazil. *Policy Studies*, 36(3), 267–282.
- Mendonça, R. F., Ercan, S. A., & Asenbaum, H. (2020). More than Words: A Multidimensional Approach to Deliberative Democracy. *Political Studies*, 0032321720950561.
- Messner, M., & Distaso, M. W. (2008). The source cycle: How traditional media and weblogs use each other as sources. *Journalism studies*, 9(3), 447–463.
- Moore, A., Fredheim, R., Wyss, D., & Beste, S. (2020). Deliberation and Identity Rules: The Effect of Anonymity, Pseudonyms and Real-Name Requirements on the Cognitive Complexity of Online News Comments. *Political Studies*, 0032321719891385.
- Munzert, S., Ramirez-Ruiz, S., Barberá, P., Guess, A. M., & Yang, J. (2021). Cheating in Online Assessments of Political Knowledge: Evidence from Survey and Digital Trace Data.
- Munzert, S., & Selb, P. (2017). Measuring political knowledge in web-based surveys: An experimental validation of visual versus verbal instruments. *Social Science Computer Review*, 35(2), 167–183.
- Neuberger, C., Nuernbergk, C., & Langenohl, S. (2019). Journalism as multichannel communication: A newsroom survey on the multiple uses of social media. *Journalism Studies*, 20(9), 1260–1280.

- Niemeyer, S., Curato, N., & Bächtiger, A. (2015). Assessing the deliberative capacity of democratic polities and the factors that contribute to it. *Democracy: A Citizens' Perspective, Abo (Turku), Finland*.
- Nonnecke, B., & Preece, J. (1999). Shedding light on lurkers in online communities. *Ethnographic studies in real and virtual environments: Inhabited information spaces and connected communities, Edinburgh, 123128*.
- Ohlsen, N. (2015). Example for a latent class analysis with the polCA-package in R – ahoi data. Retrieved June 28, 2021, from <https://statistics.ohlsen-web.de/latent-class-analysis-polca/>
- Oksanen, J. (2013). Vegan: Ecological diversity. *R Project, 368*.
- Ooms, J. (2018). *Writexl: Export Data Frames to Excel 'xlsx' Format*. <https://CRAN.R-project.org/package=writexl>
- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- Parkinson, J., & Mansbridge, J. (2012). *Deliberative systems: Deliberative democracy at the large scale*. Cambridge University Press.
- Pebesma, E., Bivand, R., Racine, E., Sumner, M., Cook, I., Keitt, T., et al. (2018). Sf: Simple Features for R. *R package version 0.6-1*.
- Pedrini, S. (2014). Deliberative capacity in the political and civic sphere. *Swiss Political Science Review, 20(2)*, 263–286.
- Prior, M. (2005). News vs. entertainment: How increasing media choice widens gaps in political knowledge and turnout. *American Journal of Political Science, 49(3)*, 577–592.
- Quandt, T. (2018). Dark Participation. *Media and Communication, 6(4)*, 36–48. <https://doi.org/10.17645/mac.v6i4.1519>
- R Core Team. (2018). R: A language and environment for statistical computing.
- Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences, 118(26)*.

- Richardson, J. E., & Stanyer, J. (2011). Reader opinion in the digital age: Tabloid and broadsheet newspaper websites and the exercise of political voice. *Journalism*, *12*(8), 983–1003.
- Rivero, G. (2019). Preaching to the choir: Ideology and following behaviour in social media. *Contemporary Social Science*, *14*(1), 54–70. <https://doi.org/10.1080/21582041.2017.1325924>
- Rivers, D. (2006). Sample matching: Representative sampling from internet panels. *Polimetrix White Paper Series*.
- Rosenberg, J. M., Beymer, P. N., Anderson, D. J., Van Lissa, C., & Schmidt, J. A. (2019). tidyLPA: An R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *Journal of Open Source Software*, *3*(30), 978.
- Rosenberg, S. W. (2014). Citizen competence and the psychology of deliberation. *Deliberative democracy: Issues and cases*, 98–117.
- Rowe, I. (2015). Deliberation 2.0: Comparing the deliberative quality of online news user comments across platforms. *Journal of broadcasting & electronic media*, *59*(4), 539–555.
- Salem, H., & Stephany, F. (2020). Wikipedia: A Challenger’s Best Friend? Utilising Information-seeking Behaviour Patterns to Predict US Congressional Elections. *Utilising Information-seeking Behaviour Patterns to Predict US Congressional Elections (October 27, 2020)*.
- Scudder, M. F. (2020). *Beyond Empathy and Inclusion: The Challenge of Listening in Democratic Deliberation*. Oxford University Press.
- Steenbergen, M. R., Bächtiger, A., Spörndli, M., & Steiner, J. (2003). Measuring political deliberation: A discourse quality index. *Comparative European Politics*, *1*(1), 21–48.
- Steyaert, J. (2000). Local governments online and the role of the resident: Government shop versus electronic community. *Social Science Computer Review*, *18*(1), 3–16.
- Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. *Social Science Computer Review*, *38*(5), 503–516.

- Strandberg, K., & Grönlund, K. (2018). Online Deliberation. In *The Oxford Handbook of Deliberative Democracy*. Retrieved October 22, 2021, from <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780198747369.001.0001/oxfordhb-9780198747369-e-28>
- Strauss, N., Huber, B., & Gil de Zuniga, H. (2020). "Yes, I Saw It - But Didn't Read It ..." A Cross-Country Study, Exploring Relationships between Incidental News Exposure and News Use across Platforms. *Digital Journalism*, 8(9), 1181–1205. <https://doi.org/10.1080/21670811.2020.1832130>
- Sun, N., Rau, P. P.-L., & Ma, L. (2014). Understanding lurkers in online communities: A literature review. *Computers in Human Behavior*, 38, 110–117.
- Sunstein, C. R. (2002). The law of group polarization. *Journal of political philosophy*, 10(2), 175–195.
- Sunstein, C. R. (2018). *# Republic: Divided democracy in the age of social media*. Princeton University Press.
- Tewksbury, D., Weaver, A. J., & Maddex, B. D. (2001). Accidentally informed: Incidental news exposure on the World Wide Web. *Journalism & mass communication quarterly*, 78(3), 533–554.
- Thompson, D. F. (2008). Deliberative democratic theory and empirical political science. *Annu. Rev. Polit. Sci.*, 11, 497–520.
- Ugarriza, E., & Caluwaerts, D. (2014). *Democratic Deliberation in Deeply Divided Societies:: From Conflict to Common Ground*. Springer.
- Valeriani, A., & Vaccari, C. (2016). Accidental exposure to politics on social media as online participation equalizer in Germany, Italy, and the United Kingdom. *New Media & Society*, 18(9), 1857–1874.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Wallsten, K. (2007). Agenda setting and the blogosphere: An analysis of the relationship between mainstream media and political blogs. *Review of policy research*, 24(6), 567–587.

- Warnes, G. R., Bolker, B., Gorjanc, G., Grothendieck, G., Korosec, A., Lumley, T., MacQueen, D., Magnusson, A., Rogers, J., et al. (2015). *Gdata: Various R Programming Tools for Data Manipulation*. <http://CRAN.R-project.org/package=gdata>
- Warren, M. E. (2021). A bridge across the democracy–expertise divide.
- Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686.
- Wickham, H., & Bryan, J. (2018). *Readxl: Read Excel Files*. <https://CRAN.R-project.org/package=readxl>
- Wickham, H., & François, R. (2015). Dplyr: A Grammar of Data Manipulation.
- Wickham, H., Hester, J., & François, R. (2017). *Readr: Read Rectangular Text Data*.
- Wiklund, H. (2005). A Habermasian analysis of the deliberative democratic potential of ICT-enabled services in Swedish municipalities. *new media & society*, 7(2), 247–270.
- Wright, S. (2012). From “Third Place” to “Third Space”: Everyday Political Talk in Non-Political Online Spaces [Publisher: Routledge \_eprint: <https://doi.org/10.1080/13183222.2012.11009088>]. *Javnost - The Public*, 19(3), 5–20. <https://doi.org/10.1080/13183222.2012.11009088>
- Xie, Y. (2018). Knitr: A comprehensive tool for reproducible research in R. In *Implementing reproducible research* (pp. 3–31). Chapman; Hall/CRC.
- Yadamsuren, B., & Erdelez, S. (2010). Incidental exposure to online news. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1–8.
- Yang, T., Majó-Vázquez, S., Nielsen, R. K., & González-Bailón, S. (2020). Exposure to news grows less fragmented with an increase in mobile access. *Proceedings of the National Academy of Sciences*, 117(46), 28678–28683.
- Zeleny, D. (2021). Analysis of community ecology data in R. Retrieved October 11, 2021, from <https://www.davidzeleny.net/anadat-r/doku.php/en:start>

Zhu, H. (2019). KableExtra: Construct complex table with 'kable' and pipe syntax. *R package version, 1(0)*.

Ziegele, M., Quiring, O., Esau, K., & Friess, D. (2020). Linking news value theory with online deliberation: How news factors and illustration factors in news articles affect the deliberative quality of user discussions in SNS' comment sections. *Communication Research, 47(6)*, 860–890.

## Appendix A

### Methodological Details

#### Political website identification

In order to capture every potentially relevant website visit, an extensive dictionary of political topics, in particular topics within the public discourse in 2017, was compiled to classify the full sample of URLs. While most studies in the field of online media research operate on the domain or subdomain level (Guess, 2021), I use the full text of URLs for this classification step.

Keywords capturing potentially political topics were collected from a range of source websites, for example the federal center for political education (Federal Agency for Civic Education, 2017) and various news articles discussing relevant issues for the 2017 federal election in Germany (see SI D). I went through the full text of those articles and collected keywords until the dictionary appeared 'saturated' because with further keyword collection, majorly duplicates appeared<sup>4</sup>. Instead of limiting the analytical focus a-priorily, this procedure ensures that as many sites as possible that are potentially relevant for the public discourse are captured. The first version of the dictionary included 266 keywords<sup>5</sup>. Using this list, I ran the filter query to select the potentially political URLs. In a next step, grouped by domains, the output was checked manually for face validity on the website level (by visiting the URL). More specifically, I checked whether the websites were actu-

<sup>4</sup>I also tried scraping the full text of websites and processing it with various automated text mining approaches to extract keywords in a semi-automated way. However, those resulting keyword sets appeared much less useful which I attribute to substantive parts of text being embedded in graphics which cannot be scraped easily with an html-based text extraction process. Therefore, and facing a small number of dense websites, I opted for a manual approach of keyword extraction.

<sup>5</sup>Words including German Umlaute like 'ä', 'ö' and 'ü', I included as is, as 'ae', 'oe' and 'ue', and in UTF-8 hexadecimal encoding.

ally featuring political topics or discussions and if not, I tried to identify the keyword that mistakenly selected this website. I then excluded a considerable set of keywords from the dictionary, to sharpen the query. I also added a layer of exclusions to the query, for example I include URLs with the keyword ‘petition’ but exclude ‘competition’. I repeated this manual cross validation step repeatedly, until I did not find any more systematic mismatches (false positives). The final dictionary and the list of exclusions can be found in SI (D).

Furthermore, since many keywords (such as family, rent and finance) also appear in many non-political contexts, all website visits that were previously classified as irrelevant category<sup>6</sup> were excluded from the set of relevant URLs. Search engines were kept in the dataset as information-providers when people searched for political topics, even though they often only work as mediators and do not provide information themselves. However, political information seeking is largely considered as highly important digital feature of the overall democratic process, to the extent that it can be used to predict elections (Salem & Stephany, 2020).

### Latent Class Analysis

The latent class model probabilistically assigns each website to a ‘latent class’, which in turn produces expectations about how that website will score on each of the manifest deliberative criteria that constitute the six dimensions of deliberative potential (Linzer & Lewis, 2011). Accordingly, websites within the same latent class are similar on certain dimensions, while those in different latent classes are dissimilar from each other in some significant way.

The parameters of the latent class model,  $p_r$  and  $\pi_{jrk}$  as prior and class-conditional outcome probabilities that each individual belongs to each class, are iteratively estimated by an EM-Algorithm (estimation maximization). In a first step, class-membership probabilities are estimated, while in a second step those estimates are altered to maximize the likelihood-function

---

<sup>6</sup>irrelevant domain categories identified in the first round of top 1000 domain coding: astrology, banking, cashback, cooking, dating, file-hosting, function, gambling, mail, micro-job, porn, shopping, sport, survey, tracking, transport, travel, uni, work



$$\ln L = \sum_{i=1}^N \ln \sum_{r=1}^R p_r \prod_{j=1}^J \prod_{k=1}^{K_j} (\pi_{jrk})^{Y_{ijk}}. \quad (\text{A1})$$

With  $N$  representing the number of observations (websites) and  $J$  representing the manifest variables (deliberative criteria) with  $K$  possible outcomes (present or not present).  $R$  represents the outcome probabilities which are fixed prior to estimation.

Both steps are iteratively repeated until the global maximum is found (Ohlsen, 2015). I fitted the latent class models with the `poLCA::poLCA()` R function (Linzer & Lewis, 2011), which suits the categorical assessment of the criteria for deliberative potential. In this case, the number of classes in the sample is not theoretically defined but identified through the consideration of the smallest Bayesian Information Criterion (BIC). Ideally, the identified model should have the smallest number of classes possible while maintaining a low classification error Díaz and Koutra (2013).

I had no prior theoretical expectations about the number of latent classes and therefore, started to fit an ‘independence’ model with one latent class (Linzer & Lewis, 2011) and then iteratively increased the number of latent classes by one up to a number of 10 latent classes.

Considering various model fit criteria a model with three latent classes was selected (Lin & Dayton, 1997). The model fit criteria were very similar between a two-class and a three-class solution. I therefore also present results for the two class solution in SI but stick with the three-class solution in the main body of the text. I validated the model (a) using a split-sample approach, randomly cutting 50% of the sample with no considerable effects to the latent class solution (see SI C10) and (b) by making changes to the input criteria (Bacher & Vermunt, 2010). Dropping criteria from the model input that do not differ substantially within a class (in this case, age and educational inclusivity and party preference heterogeneity), does also not significantly change the results (see SI C9). The classes still show the same characteristics and relative sizes, with only minor changes in the individual class memberships of websites.

As another robustness check, the supply side criteria (information, communication and participation) and the demand side criteria (connectivity, inclusivity and heterogeneity)

were also considered separately in two LCA models, as they form two natural, methodological groups. For the computationally coded, demand side criteria, a simple two factor solution was suggested with one class including all websites with high probabilities of fulfilling each criterion and one class with overall very low scores for connectivity, inclusivity and heterogeneity in other words, high and low engagement websites. The model including only the manually coded infrastructural criteria of information, communication and participation possibilities suggested a more interesting pattern that is in line with the findings from the main model including all criteria. A first class contains websites with a strong information profile, including all public broadcasting pages. A second class contains websites with an especially strong forum component or communication profile with pages that also enable participation to some extent. The last class is rather a residual class including websites with overall low probabilities of fulfilling any criteria.

### Shannon-Wiener Index

Entropy-based diversity indices such as the here applied Shannon-Wiener index (Dixon, 2003; Grafton et al., 2012; Kiernan, 2014; Oksanen, 2013) are regularly used in the context of bio-diversity assessments, combining measures of richness (the number of categories) and evenness (similar frequencies of categories).

$$H = - \sum_{i=1}^S p_i \ln p_i \quad (\text{A2})$$

with  $p_i$  being the proportion of category  $i$ ,  $S$  being the number of categories so that  $\sum_{i=1}^S p_i = 1$  (Oksanen, 2013).

## Appendix B

### Information on the Survey and Tracking Data

#### Sampling

The survey was administrated by YouGov combining purposive sampling with a multi-stage sample-matching and weighting procedure (Rivers, 2006). First, the target population, the German online population, was defined. For the panel, data on demographic marginals (gender, age and educational attainment) from Best for Planning (2017)

were used, who conducted 30,000 face-to-face interviews to evaluate the German online population. A stratified sample was drawn from this frame and matched as closely as possible to YouGov's longstanding panel (with over a million members) (Munzert et al., 2021, same data).

The resulting target sample constitutes a representative set of respondents in terms of traditional sampling theory. However, respondents might be hard to contact because they either have never reported their contact details or do not agree to the terms of the survey. Hence, multi-stage matching was applied, combining the representative target sample with YouGov's longstanding panel of reliable respondents. From this panel, a sample of individuals was selected that matches as closely as possible the distribution of the target sample and had opted in to providing website visit data (see below). Through this procedure, YouGov guarantees not only a minimum of 1,000 respondents in the survey, but also the inclusion of hard-to-reach population subgroups (Munzert et al., 2021).

### **Survey Design and Fielding**

All data was gathered by YouGov from July 1 to December 9, 2017. The panel was made up of five waves. The survey covered a wide range of topics such as people's political preferences, political knowledge on several domains, their general attitudes towards politics, opinions on particular parties, and what people think of the election campaign (the federal elections were held on September 24, 2017) (Munzert et al., 2021, same survey).

### **Passive Metering Technology**

Wakoopa, the tracking software used by YouGov in this study, ran in the background of panelists' devices and collected anonymized visit data. There are no technological limits to the types of websites that can be included in the data. Moreover, the software tracks web traffic (passwords and financial transactions are ignored) for all browsers installed on a user's computer. The technology does not slow the performance of users' computers and is transparent about the data that is being sent: Panelists can see a list of the last several captured URLs and can also pause tracking for 15 minutes. Of course, they can also uninstall the software at any time. YouGov encourages its panelists to install the

software on as many devices as possible, including laptops, mobile phones, and tablets. The capabilities for mobile tracking are somewhat more limited for privacy reasons, but data on domain-level visits and app use are collected (Munzert et al., 2021, same data).

Panelists were recruited from YouGov’s traditional participant pool via incentives. The company reports fairly strong incentives: 4,000 “points” for signing up and downloading the Wakoopa software—roughly 8 times the number offered for a typical survey—and 1,000 additional points every month. Participants in online surveys can redeem these points for clothing, prepaid gift cards, and other merchandise. One consequence of this recruitment strategy is that YouGov Pulse users are a subset of the overall panel, making sampling somewhat more challenging (Munzert et al., 2021).

### **Privacy and Ethical Considerations of Data**

Combining survey data and digital trace data of the same respondents has substantive merits to understand the effects of online exposure on people’s attitudes and behavior. However, it entails challenging tasks for the protection of privacy and raises ethical questions, as users may not be aware of how their data are being used. Even with the consent of the participants, it still could be problematic because the account names and meta-information of social media accounts can be identifiable and linked to their survey responses (Stier et al., 2020). Thus, it is important to communicate these concerns as clearly as possible when collecting data (Menchen–Trevino, 2013). In every step of data collection, participants were informed about the scope of data collection, data management, confidentiality, and research purpose. Explicit and informed consent was obtained from all participants whose data was collected (Munzert et al., 2021, same considerations).

Regarding the web-tracking data, YouGov received the consent from the panel that their web browsing data can be linked to other survey items they have participated. They highlighted that participants have complete control over which data they share for research purposes. Participants can choose which information they want to share, pause the tracking app when they want, and withdraw their consent anytime. After data collection, YouGov removed any personally identifying information and sensitive data (e.g., financial transaction) and stripped-out geocoding information that is too specific before delivering the data

to researchers. The deliverables are de-identified and anonymized and fully comply with the EU's General Data Protection Regulation (GDPR) requirements (Munzert et al., 2021).

### **P3. Beyond the tip of the iceberg? Exploring Characteristics of the Online Public with Digital Trace Data.**

#### **Preprint Publication**

Oswald, L., Munzert, S., Barberá, P., Guess, A., and Yang, J. (2022).  
Beyond the tip of the iceberg? Exploring Characteristics of the Online Public with  
Digital Trace Data. <https://osf.io/preprints/socarxiv/yfmzh/>

#### **Supplementary Materials**

Supplementary information including documented code, available data,  
supplementary tables and figures and more details on data collection available under  
<https://osf.io/2k45y/>.

# Beyond the tip of the iceberg? Exploring Characteristics of the Online Public with Digital Trace Data

*Lisa Oswald*\*<sup>1</sup>, *Simon Munzert*<sup>1</sup>, *Pablo Barberá*<sup>2</sup>,  
*Andrew Guess*<sup>3</sup> and *JungHwan Yang*<sup>4</sup>

<sup>1</sup>Hertie School

<sup>2</sup>University of Southern California

<sup>3</sup>Princeton University


<sup>4</sup>University of Illinois at Urbana-Champaign

## Abstract

Who constitutes the online public, the citizens engaging with political content online? The digital extension of the public arena challenges conventional approaches examining public discourse. Besides mapping the political online ecosystem beyond news outlets, a crucial factor for understanding the online public discourse are the individuals involved. Using web tracking data in combination with survey data, we map out who engages with political content across various arenas of the complex political online ecosystem in Germany. We find that people with especially high political knowledge engage with political information online. We do not find evidence for strong selection along other demographic, attitudinal, and political variables but instead observe large overlaps among the online public. However, people with rather extreme political views tend to engage more with platforms that enable political expression and discussion than people with more mainstream political views. This finding highlights differences between active and passive political internet use and has implications for the inferences drawn from social media data that usually reflects a - likely skewed - tip of the iceberg.

*Keywords:* political information consumption, online deliberation, digital trace data, selection, latent profiles

---

\* corresponding author  <https://orcid.org/0000-0002-8418-282X>

Address: Friedrichstrasse 180, 10117 Berlin, Germany. Email: [l.oswald@phd.hertie-school.org](mailto:l.oswald@phd.hertie-school.org), Web: [lfoswald.github.io](https://lfoswald.github.io).

## Introduction

The complexity of online ecosystems that provide political information and diverse channels for political discussion challenges traditional public sphere concepts (Habermas, 2021; Habermas et al., 1974). The public sphere can be understood as intermediary system that mediates between civil society and the political system (Jungherr & Schroeder, 2021). The communication processes taking place within the public sphere, often referred to as ‘deliberation’, are supposed to help form public opinion, legitimize political decisions, and mitigate opinion polarization (Friess & Eilders, 2015; Wessler, 2008).

While evidence from experiments suggests a negative relationship between deliberation and partisan polarization (J. Fishkin et al., 2021), deliberative processes are not only shaped by the setting and the discussed issues but also by the individuals involved (Bächtiger & Hangartner, 2010; Esterling, 2018). Beauchamp (2020) describes the deliberativeness of online discussions as a function of structures and membership. This conceptualization emphasizes the importance of selection processes, especially because online political deliberation, narrowly defined as an ideal form of political discussion in which participants weight conflicting arguments based on high quality information, appears to be a rather rare phenomenon in relation to the overall volume of online communication. Experimental deliberation research tries to circumvent the problem of selection through random sampling and/or random assignment for the benefit of internal validity. However, while experimental evidence can clarify fundamental principles under ideal conditions, this evidence hardly generalizes to real-life online political discourse that underlies inherent selection processes.

When examining the online public discourse, a first important question is, which websites, platforms and forums belong to the online public sphere? From a systemic perspective, different arenas can fulfill different functions within the democratic system (information provision, opinion exchange and eventually political participation) and can therefore be characterized by varying, potentially complementary degrees of deliberative quality (Bächtiger & Parkinson, 2019). The description of the political online ecosystem of websites, forums and platforms has been the subject of the paper by Oswald (2022). The following, and potentially even more important question to understand the online public discourse is, who engages to what extent with which arenas of the online public sphere?



Contributing to the body of empirical research on deliberative systems (Bächtiger & Parkinson, 2019), this study uses a unique combination of digital trace data with linked survey data, including extensive information on demographics, knowledge and attitudes, in order to quantify patterns of selection into different arenas of the complex political online ecosystem.

We find that especially people with high political knowledge engage with political information online. However, we do not find evidence for strong selection along other variables but a large demographic and attitudinal overlap among the online public. However, people with more extreme political views tend to select more into online forums and other platforms that enable political expression and discussion than people with more mainstream political views.

Our findings imply important differences between passive and active forms of political online engagement. While in our study, political knowledge appears to be the only strong selection factor that predicts visits to websites that feature political content, other research that considers digital traces of active engagement such as user comments or retweets implies more pronounced selection patterns, for example, along political ideology (Barberá et al., 2015; Kim et al., 2021).

### **Political Engagement Online**

Even though online political deliberation may not be a mainstream phenomenon, it is central to public opinion formation (Beauchamp, 2020; Habermas, 2021). Drawing meaningful conclusions about the state of the online public discourse, requires the examination of differences between those who select into it and those who stay away from political information, communication and participation online.

The question of who engages with political content online has previously been addressed by media researchers with a particular focus on news media diets (e.g. Guess, 2021). In this course, the quality and ideological framing of online information that is consumed can be considered as one important driver of public opinion.

Facing concerns about increasing political polarization online (Pariser, 2011; Sunstein, 2002), some researchers identify an increased availability of slanted information

through the active engagement of partisan activists in new online media outlets, that continue to depart further from the moderate majority (Lelkes, 2016; Prior, 2013). Not only the information provision but also the consumption of partisan information appears to increase with the digitization of media (Peterson et al., 2021). However, a growing body of literature suggests a different picture. Despite a homophilic structure of online social networks, the exposure to cross-cutting content is, by far, no exception in online environments (Bakshy et al., 2015; Barberá et al., 2015; Bright et al., 2020; Lorenz-Spreen et al., 2021; Oswald & Bright, 2022; Vaccari & Valeriani, 2018) and most users continue to receive political information from their favorite, mostly mainstream, news outlets (Flaxman et al., 2016; Guess, 2021; Oswald, 2022). It is, of course, difficult to define a benchmark of an ‘ideal level’ of opinion heterogeneity or ‘ideal’ diversity in news consumption. However, the magnitude of empirical effects of selective exposure are modest and mainly driven by a minority of highly partisan individuals (Barberá et al., 2015; Bright et al., 2019; Guess, 2021).

Measuring the extent of selective exposure to news media is certainly one important component to assess the potential of the online sphere for a functioning democracy. Previous research, for example, examined news access through social media sites (Bakshy et al., 2015; Barberá et al., 2015; Eady et al., 2019), tends to focus on mainstream media outlets and often used self-reports in surveys (Boxell et al., 2017; Dubois & Blank, 2018; Lelkes, 2016). Instead, web tracking data provides unique insights into real-life online behavior. For example, Guess (2021) illuminates important aspects of online news media diets but does not capture the entire ecosystem of political online environments, which, besides news, also provides forums for political discussion and opportunities for online political participation. In contrast, most deliberation research focuses exclusively on the dimension of communication. Therefore, building upon the findings in Oswald (2022), a more holistic measurement is used in this project to address the question how the online public is exposed to the ecosystem of political environments.

From a methodological point of view, issues of selection can pose substantial challenges to the interpretation of empirical findings. For example, comparing the deliberative quality of discussions in different online environments risks selection bias in the inferences made about, for example, platform design effects (Coe et al., 2014; Esau et al., 2020; Zhang

et al., 2013). In general, using observational communication data for the comparative analysis of online deliberation, for example by sampling posts from social media or comments from a news media site, usually neglects the composition of users in terms of demography, deliberative capacities, attitudes and motivation (Springer et al., 2015).

Within the field of deliberation research, the issue of selection bias has most explicitly been acknowledged in works on motivations to participate in fact-to-face deliberative settings and the discussion of representativity (J. S. Fishkin, 2018; Jacquet, 2017). In a rather early paper, Stromer-Galley (2003) note that Wilhelm (2000) analyzed Usenet groups with ‘atypically homogeneous participants’ (Wright & Street, 2007, p. 853) as a validity issue. However, what is the ‘typical’ composition of users across a complex online public sphere?

Meanwhile, the issue of selection was extensively discussed in the literature on political participation (Conway, 2001; Persson, 2015; Prewitt, 1965; Putnam, 1993; Quintelier & Hooghe, 2012; Stoker & Jennings, 1995; Uslander, 2002). Consistent with our approach of examining diverse sites within a complex political online ecosystem, research on the influence of political information on offline political participation considered different offline settings of information exchange, such as family, work or leisure contexts. While those traditional offline settings are often structurally imposed on an individual (Mutz, 2006), in the absence of physical constraints, online settings are more actively constructed as individuals have great freedom to self-select into different online environments (Feezell, 2016).

The research field of online political engagement and the effects of online activities on offline political behavior is growing (e.g. Boulianne, 2020; Oser & Boulianne, 2020). For example, Feezell (2016) found that, after correcting for selection bias of overall internet usage, online political participation is not well explained by resource-related predictors. In her study, income and age, usually important predictors for traditional political participation, were unrelated to online political participation. Instead, the exposure to attitude-supporting political information was related to higher levels of online political participation when compared to the exposure to opposing or neutral information. Other research suggests that digital media literacy is associated with online political engagement (Kahne et al., 2012) and that political online group membership, for example on Facebook, is related to

offline political participation (Conroy et al., 2012). Remarkably, despite increased offline participation, they did not find increased political knowledge among political online group members, which they attribute to low quality group discussions (Conroy et al., 2012).

Taking up recent theoretical developments in deliberation research, including systemic perspectives on the online public discourse (e.g. Bächtiger & Parkinson, 2019; Mansbridge et al., 2012; Parkinson et al., 2020), with this project, we aim to contribute to the, yet small, quantitative field of evidence. The primary objective of this paper is the exploration of selection factors through the description of individual characteristics between differently engaged citizens online. In other words, at the core of this paper, we ask: who engages with which arenas of an online public sphere? In this project, the term political ‘engagement’ refers to visits to websites where citizens usually engage with political content and does not imply publicly visible or active behaviors such as writing comments on social media or ‘liking’ content.

This paper builds on the systematic, data driven definition of the political online environment in Germany by Oswald (2022). In this study, we use the same data to examine patterns of selection across different sites and sub-spheres to define a typology of the online public. We are, at the time of the study and to our best knowledge, the first who add evidence from digital trace data to the study of online deliberation.

## Methods

In the following sections, we outline the data, summarize the approach taken by Oswald (2022) that maps the online ecosystem of political websites and explain the approaches of this study that measure selection and to structure the online public along different user profiles.

### Data

Our dataset is a unique combination of web tracking data and linked survey responses which allows us to observe real-life online behavior along with granular data on users’ demographics, attitudes and self-reported political behavior on an individual level (Oswald, 2022). The data was collected via the German YouGov Pulse panel with survey

questions fielded to 1,500 respondents that largely mirror the German online population with respect to gender, age and, education based on the marginals of Best for Planning (2017). Respondents were asked to install a URL-tracking software that uses passive metering technology to record detailed browser histories on an opt-in basis. These digital trace data include more than 56 million website visits, to almost 200,000 different domains, by 1,282 different individuals between July 2017 and December 2017. As this paper draws on data collected in a prior research project, details on the sampling procedure, the deployment of the passive metering software as well as privacy and ethical considerations can be found in SI and Oswald (2022).

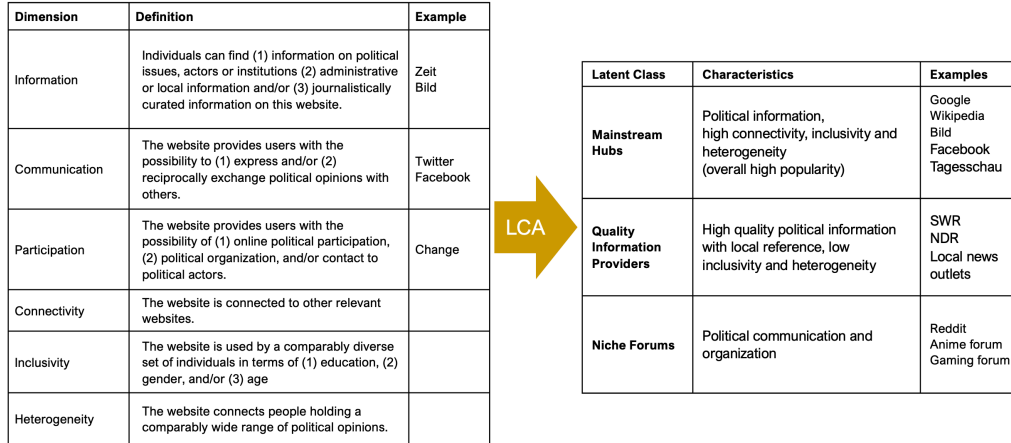
Our data structure allows us to observe the behavior of both, individuals who engage with sites featuring political content and those disengaged with politics online. The description of differences between these groups lets us identify variables potentially influential for the selection processes that separates the ‘online public’ from the politically disengaged rest of the sample. A summary of variables in the dataset can be found in SI 02.

### **Mapping the Political Online Ecosystem**

This paper builds on the work by Oswald (2022) that goes beyond the usual paradigm of the a priori nomination of sites relevant for the assessment of information exposure or the comparison of deliberation, but mapped the political website landscape in Germany, 2017, starting from the bottom-up, namely political topics in the online public discourse. Using a semi-automated dictionary approach, they filtered the full corpus of URLs down to political website visits only. Therefore, in the results of this paper which uses the same data, engagement with, for example, Facebook only captures *political* Facebook usage. Furthermore, Oswald (2022) considered different functions of websites - information provision, political communication and participation - to account for the complementary nature of different sites in a deliberative system. Domains were characterized along website design features, such as the provision of a discussion forum, using a standardized coding scheme. Three additional characteristics were considered as relevant for the deliberative potential of a website - connections to other relevant platforms, the platform’s demographic inclusivity and opinion heterogeneity among the users of a platform. Detailed informa-

**Figure 1**

*Clustering the political online ecosystem along deliberative characteristics*



*Note.* Approach taken by Oswald (2022). Left: dimensions of deliberative potential as defined in Oswald (2022). Clustering process using latent class analysis (LCA). Right: three latent classes of websites as result of the clustering process.

tion on the classification of political sites as well as the clustering of sites with latent class analysis can be found in Oswald (2022); overview of dimensions in Fig. 1, left.

Oswald (2022) identified a class of mainstream informational hubs, including major online news outlets, search engines as information intermediaries and the major social media platforms, showing high engagement levels and connections to other websites. Furthermore, a class of quality information providers, such as public broadcasting sites and local newspapers was identified, that was, however, not characterized by a diverse user base. Surprisingly, a third class of niche online forums hosted political discussions among more small and specific online communities, such as committed anime fans (see an overview of latent classes in Fig. 1, right.). While the mainstream information hubs in the sample attracted a much larger volume of clicks, on average, users spend relatively more time consuming political information on quality information sites as well as on niche online forums to engage with politics online (Oswald, 2022).

### **Selection into the online public sphere**

Following the description of the political online ecosystem in Oswald (2022), in this paper, we turn towards the online public, citizens that select into engaging with political issues in online environments.

Our study is subject to two potential selection processes: first, participants select into the survey and then second, among these participants, some select into engaging with political content online. The first selection problem is difficult to address, since, despite having large sample that largely mirrors the German online population<sup>1</sup>, we cannot observe the online behavior of those who did not agree to take part in the study.

Selection processes in online political contexts can generally be considered from two different angles which has implications for measurement and operationalization. First, one could consider differences in the media diets or usage patterns of individuals, as for example in Guess (2021). Second, one could consider differences of audience distributions between websites. Both approaches, with foci either on users or on websites, are representations of the same selection process. In this project, we follow the second approach and map selection from the statistical angle that we predict website engagement with individual characteristics of users.

Even though it would ultimately be interesting to examine selection into concrete communicative acts of online deliberation even more explicitly, using web browsing histories, we can realistically only observe the exposure to websites that provide political information, enable online political expression, discussion and participation.

Concerning different types of online engagement, we consider engagement with political information providers and engagement with political communication sites (see rows 1 and 2 of Fig. 2). The outcome measure of political information is a summary score of three distinct elements: whether the website features information on political issues, actors or institutions, information on administrative or local procedures and whether the website is a primary source of journalistically curated information or information that underwent

---

<sup>1</sup>678 individuals identified as male, 603 as female. We find a normal age distribution for the German online public (see SI 09), and diverse educational backgrounds (see SI 06). The distribution of the self-reported political orientation of the sample can be found in Fig. 010. Also geographically, online activity patterns in our sample distributes about evenly across Germany (see 011).

some form of fact checking. The communication dimension is a summary score of two criteria: whether the site enables political expression in the form of comments (or ratings of comments or primary content) and whether the site enables reciprocity in communication through open replies to other comments (see definitions in Fig. 1).

We further consider engagement with three latent classes of sites identified in Oswald (2022): informational ‘mainstream’ hubs, journalistic news outlets with local reference, and niche online forums that host political discussions (see Fig. 1, right). We also predict engagement with single domains as prototypical examples of functional categories: ‘Zeit’ and ‘Bild’ as online newspapers, ‘Facebook’ and ‘Twitter’ as social media platforms and ‘Change.org’ as online petitioning platform.

### Latent Profile Analysis of Users

While the selection models consider different forms of online political engagement separately, we then turn to more complex engagement profiles to describe interpersonal overlaps in combinations of political internet usage patterns. In other words, in order to structure the online public, users are clustered along their complex engagement patterns. A latent profile analysis (LPA, details see SI ) was performed on various distinct engagement measures, a method that aligns closely with deliberative systems ideas (Bächtiger & Parkinson, 2019). We used the standardized, log scaled scores of 9 different engagement measures<sup>2</sup> as input variables in the Latent Profile Analysis: the overall online engagement (including apolitical content), the overall political engagement, the engagement with public broadcasting sites, the engagement with political discussions in niche forums, and the political engagement with five different specific websites (Zeit, Bild, Facebook, Twitter and Change).

## Results

The presentation of results will be structured along the following research questions:

1. How is engagement distributed between different sites of the political online environment? (Section ‘Online Activity’ and Fig. 2)

---

<sup>2</sup>Measures that correlated less than 0.6 (see Fig. 05)



2. Who selects into engaging with different websites featuring political topics? (Section ‘Selection’ and Fig. 3)
3. How does the latent structure of different engagement profiles look like? (Section ‘User Typology’ and Fig. 4)

Users’ engagement with sites was measured with regard to the duration of (political<sup>3</sup>) engagement with a domain, broken down to the unit of minutes/week. We decided to focus on the duration as measure of engagement instead of the number of website visits to resolve the issue of duplicate clicks, for example, when refreshing a page. In addition to the presentation of results in the main text, we report all results also for website visits as metric of engagement in the SI.

### Online Activity

After various iterations of automated dictionary classification and manual cross validation, a sample of 69 domains (493,714 URLs) was identified featuring political topics that were relevant for the German online public discourse in 2017<sup>4</sup>.

Cumulative engagement metrics on the user level are displayed as density plots in Fig. 2. The log-scaled engagement distributions exclude non-users (users that never visited a specific domain or group of websites during the six-month observation period), percentages of users and non-users are displayed as barplots.

The overall majority of URLs in our dataset is dedicated to shopping, entertainment, work etc., while only a small fraction (about 1%) featured political content. Out of the, on average, 12 hours online activity per user per week, only about 10 minutes were identified as engagement with political issues online. However, at the same time, the majority of the sample (1,190 out of the original sample of  $N = 1,282$ ) engaged with at least some political content during the six-month observation period (Oswald, 2022).

Considering the major infrastructural setup of a website, we find that users spend, on average, 9 minutes per week on 39 different sites predominantly providing political

---

<sup>3</sup>‘Political’ clicks are those URLs that feature at least one political keyword. More details on the political website identification, see (Oswald, 2022)

<sup>4</sup>More detailed information on the procedure of website identification and classification can be found in Oswald (2022).

information and 8 minutes on 33 different sites that enable political discussions through the provision of comment sections and forums.

Regarding five selected domains as prototypical examples of different functional groups in the online public sphere, we find that most political online engagement was identified for the online outlets of the national newspapers, Zeit and Bild. While a third of our sample engages with the tabloid newspaper Bild, only a quarter of the sample engages with the online outlet of the weekly newspaper Zeit. The asymmetry in usage is even more pronounced when considering the average engagement duration. Among those who used the respective website, we find only 1.9 minutes per week spent on Zeit, while 3.9 minutes per week are spent on Bild.

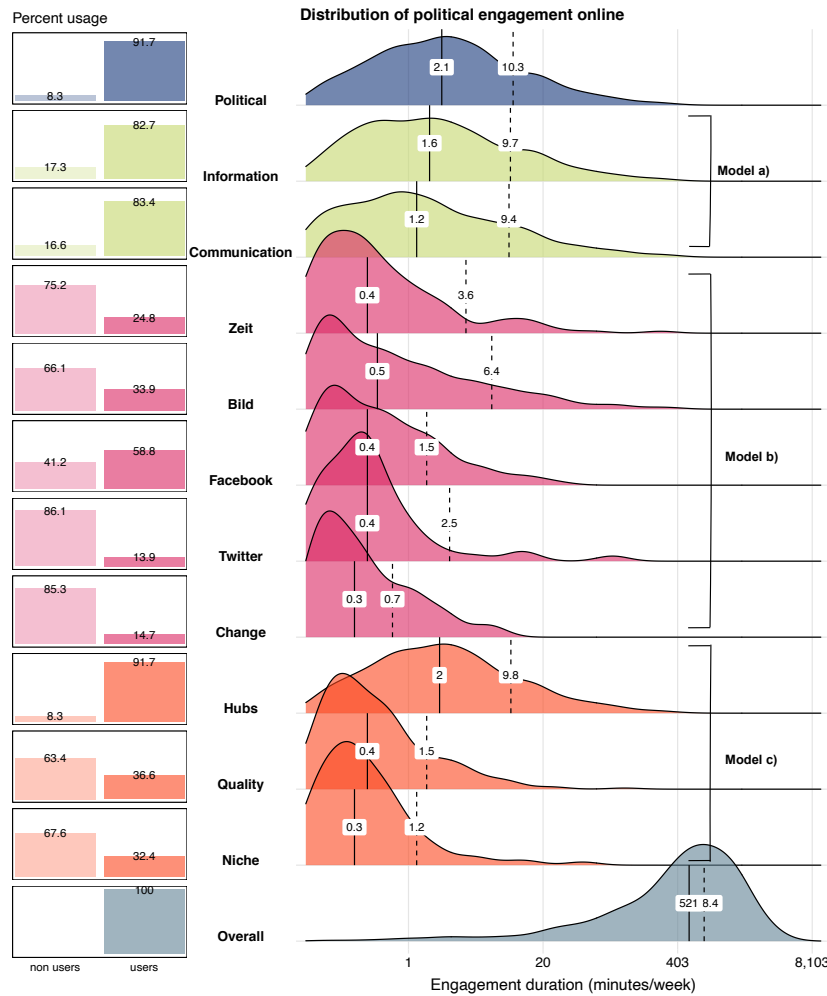
Considering two major social media platforms, we find a large asymmetry among German users between Facebook and Twitter in 2017. While 60% of the sample was exposed to content we identify as political on Facebook at some point during the six-month observation period in 2017, only 14% engaged with Twitter. However, among those who do use Twitter, we find slightly more engagement (0.9 min/week) compared to those using Facebook (0.7 min/week).

Finally, considering the online petitioning platform Change.org, only 15% of our sample engaged with Change at all and among those, we only find an engagement of 0.4 minutes per week. Despite the push for stronger inclusion of civic perspectives and more political participation in decision making (Bundestag, 2020; Commission, 2021; OECD, 2021), other, smaller or more specific, petitioning platforms, such as the one launched by the German Bundestag, do not appear in our dataset at all.

Building upon the latent structure of political websites identified in Oswald (2022), we also considered the political engagement, broken down for different latent classes of websites within the online public sphere. Within the 10 minutes of political engagement per week, the vast majority (8.9 min/week) were identified as engagement with political issues on 34 different informational hubs like Google, Wikipedia and Facebook, the most important national newspapers (Bild, Zeit, Spiegel, Welt, etc.) and, facing the national elections 2017 within the tracking period, voting advice applications like Wahl-o-mat. Users' engagement with 15 different journalistic outlets with local reference, for example online outlets of public

**Figure 2**

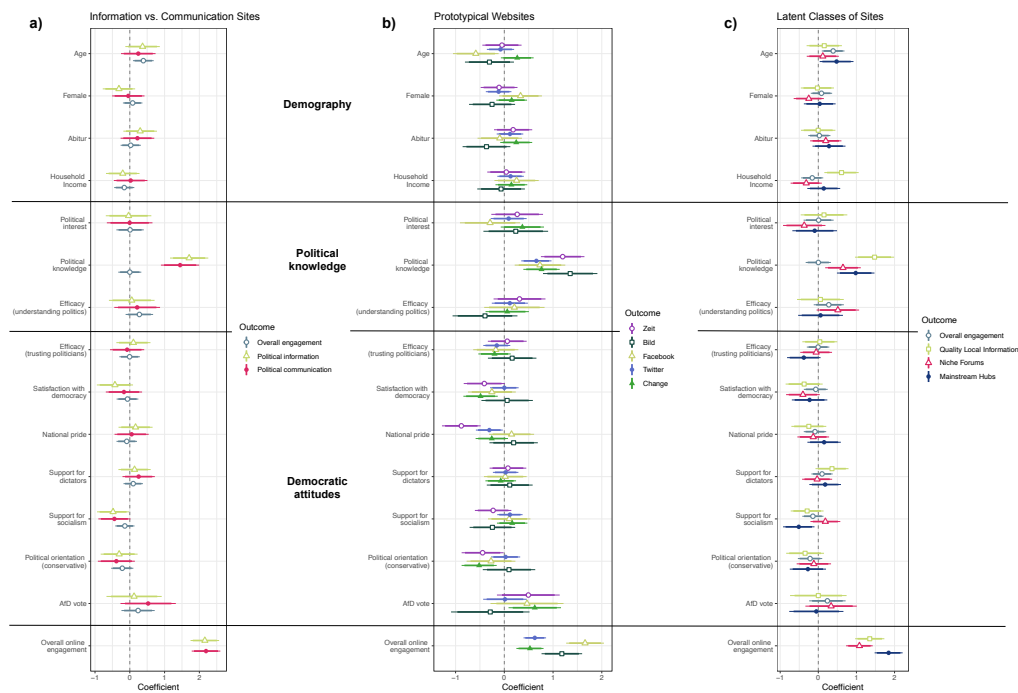
*Political engagement online*



*Note.* Density plots exclude non-users. Full line representing the median of the distribution, dashed line representing the mean. Political: all political engagement, information and communication: functional categories identified according to the sites' infrastructures (overlap possible); single domains: Zeit, Bild, Facebook, Twitter, Change; Latent classes of websites identified in Oswald (2022): Informational mainstream hubs, quality journalistic news outlets with local reference and niche online forums hosting political discussions. Engagement measured in minutes/week. Overall engagement (including political and apolitical content) corresponds to approximately 12h/week.

broadcasting services, only corresponds to an average engagement of one minute per week. The smallest engagement class (0.6 min/week), is engagement with political discussions in 20 different niche online forums.

## Selection



**Figure 3**

Multiple regression models summarizing determinants of selection into engagement with different sites or groups of sites. Error bars represent 90% and 95% confidence intervals. Overall engagement as baseline model, including non-political engagement. Outcome metric: engagement duration. For more details see SI 02, 03, and 04.

Regarding our primary research question, who selects into engaging with different websites featuring political topics, we calculated various multiple linear regression models including variables on demographics, political knowledge and democratic attitudes to predict the engagement duration for different functional categories of political sites (see Fig. 3). The multivariate nature of the model allows the interpretation of comparisons between individuals with different expressions of a certain characteristic, ‘holding all other variables

equal'. However, this multivariate nature sometimes complicates the interpretation of regression coefficients when related variables are included in the model. Even though we do not see general issues with collinearity in our model, we provide a matrix of bivariate correlations in SI 05.

As baseline model, we considered users' overall engagement with political as well as a-political websites. In the following models, considering engagement with political information, political communication sites and the three latent classes of sites, we followed the approach of Feezell (2016) and further conditioned on the overall online engagement by including it as covariate.

Fig. 3 summarizes the beta-coefficients with 95% and 90% confidence intervals for all engagement models. In the following, we discuss patterns of selection for three groups of models (a, b and c, see Fig. 2), starting with the two functional categories of political information providers and political communication platforms (see Fig. 3a).

We find that political knowledge appears to be by far the strongest predictors for political engagement online. We cannot find statistically significant differences in the political engagement for any demographic variable. However, when comparing individuals with the highest and second highest German secondary education degree (Abitur and Realschulabschluss) to those with the lowest (Hauptschulabschluss or no degree), we find that those with higher education are more active on political websites compared to those with the lowest or no degree (see SI 02).

Despite a large overlap between websites that provide political information and opportunities for communication, we find an interesting pattern when considering AfD (German far right) voters in comparison to individuals expressing other party preferences. While AfD voters tend not to be particularly engaged with sites that provide political information, they are more active on communication sites (such as social media, or sites hosting forums and comment sections).

To consider selection in a less aggregate form and to directly address some limitations that we identify in comparative work in the field, we calculated five further multiple linear regression models including the same demographic and political predictive variables but the engagement with specific websites as outcomes (see Fig. 3b). All sites are typ-

ical representatives of a respective functional category: Bild and Zeit as German online newspapers with political information function in the online public sphere, Facebook and Twitter representing two major communication platforms, and Change.org as typical online petitioning platform as form of online political participation. Just like in the models above, we only counted engagement with sites that featured political topics. For example, we only counted the duration of a Twitter-visit when the URL featured a political keyword from our dictionary of topics in the public discourse in Germany 2017. When selecting the five exemplary websites, we ensured that at least 150 different individuals (out of our original sample of 1282 participants) engaged with the respective website.

When comparing people with different degrees of political knowledge and holding everything else equal, again, those with higher political knowledge show significantly higher engagement with Zeit, one of the most prominent weekly newspapers in Germany. We find, indeed, a similar pattern when considering the engagement with Bild, the most prominent tabloid newspaper, as outcome variable. However, for Zeit we find a more nuanced pattern of selection: especially those expressing low agreement with items of national(ist) pride, holding everything else equal, spend more time on Zeit online.

The older the participant in our sample, the lower is their respective time spent on engaging with political issues on Facebook, everything else held constant. Engagement with Twitter for political topics is related to a considerably different selection pattern: Those with higher political knowledge, and low national pride show more engagement with political Twitter in Germany.

Concerning the only frequented participation platform in our sample, Change.org, we find that, holding everything else equal, those with higher political knowledge and political interest but lower satisfaction with democracy, and, in comparison to all other voters, especially participants with a preference for the right-wing AfD were more engaged with the petitioning platform Change.org.

Furthermore, in addition to the two groups of models, we calculated three further regression models to predict the engagement with different latent classes of sites (see Fig. 3c) that were identified in Oswald (2022) that mapped the online public sphere's infrastruc-

ture<sup>5</sup>. Overall, these models echo the findings from the previous analysis to the extent that political knowledge and education remain the most dominant predictors for political online engagement - causal direction, of course, to be determined (Esterling, 2018).

The user base of different national and regional outlets of (mostly) public broadcast media is shaped by high income individuals with high levels of political knowledge. Considering the engagement with political discussions on niche online forums, we find that individuals with lower household income and lower levels of political interest and satisfaction with democracy but at the same time, comparably high levels of political knowledge, tend to be most engaged. However, these differences are not statistically significant, holding all other factors equal.

### User Typology

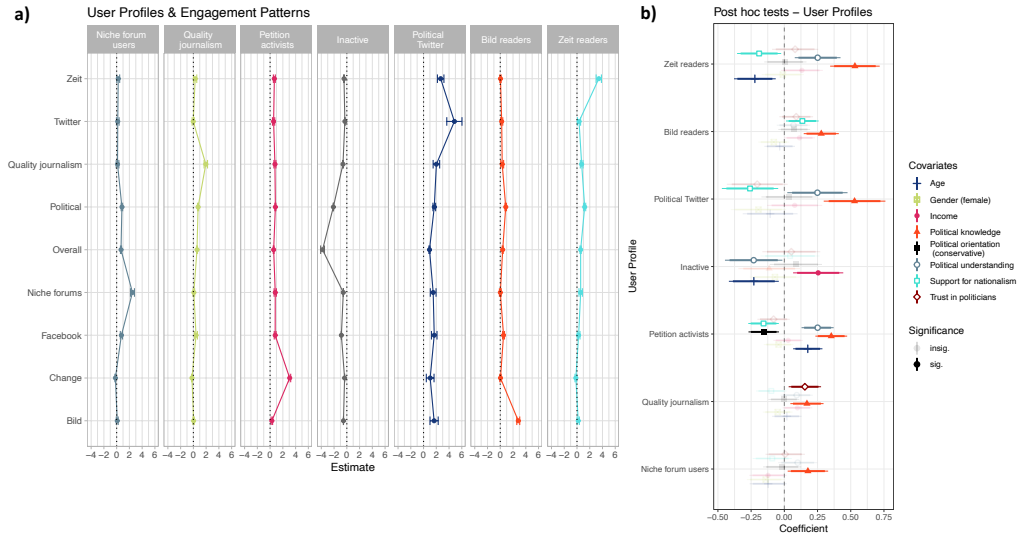
To further examine more complex, multidimensional user-profiles among the online public, we used a latent clustering approach. The analytic hierarchy process built into the `tidyLPA` R package (Akogul & Erisoglu, 2017; Rosenberg et al., 2019), based on various fit indices, suggests the selection of a model with varying means, equal variances, covariances fixed to 0, and 10 latent profiles (Fig. 4a summarizes the characterization of the profiles visually).

We conducted Post hoc tests to analyze whether the determined classes, using online engagement measures as input, also differ on other variables that were not included in the clustering process. In summary, these post hoc tests showed that classes did indeed appear considerably distinct with regard to demographic and attitudinal variables (see Fig. 4b). A summary of profiles with relative group sizes is shown in Table 01. Profiles 3, 5 and 7 (see 01) show overall below-average engagement but did not appear considerably different from individuals assigned to the ‘average’ profile 10, and were therefore dropped from Fig. 4.

For the post-hoc tests, we used the ‘average engagement’ profile as reference cat-

---

<sup>5</sup>This previous latent class segmentation analysis on the websites within the German online public sphere in 2017 revealed three clusters of websites: (1) informational hubs that are used as common source of political information by a large fraction of the sample, (2) a cluster of public broadcasting online outlets that feature high quality information, often with a specific local focus and (3) niche online forums that are dedicated to specific topics or communities outside the political context but that feature in-depth political discussions in certain corners of their forums. Following the segmentation of the deliberative infrastructure of websites (Oswald, 2022), we now focus on the users of these clusters of websites.



**Figure 4**

**a)** Latent Profiles of Users. Y-axis labels represent different forms of engagement (corresponding to Fig. 2) as input for clustering process. Grey labels represent profile names as results of clustering process. Selective presentation of substantively interesting profiles. Complete set of profiles: 1. Niche online forum users, 2. Quality (local) journalism readers, 4. Petition activists, 5. Inactive, 6. Political internet users - especially on Twitter, 8. Bild readers, 9. Zeit readers. Profiles 10. Average users, as well as profiles 3. and 7. (both showing little engagement) were excluded from the plot because of little informational value. Engagement levels are scaled, 0 representing the overall average. **b)** Post hoc tests. Coefficients (y-axis) labels correspond to latent profile names in a) and coefficients indicate the degree to which profile membership predicts the expression of the covariates (technically outcome variables in this model). Differences between latent profiles (profile 10 (average users) as reference category). Significance of prediction according to conventional threshold of  $p > .05$ .

egory. Parallel to the results of the multiple regression models, the ‘niche forum users’ tend to be younger, have lower income but high political knowledge. Users assigned to the ‘public broadcasting profile’ also show high political knowledge but especially high trust in politicians and tend to have higher income. People assigned to the ‘inactive’ profile are, according to the post-hoc analysis significantly younger, have higher income but experience very low efficacy of understanding politics.

In contrast, the ‘petition activists’ appear significantly older, indicate left leaning political orientation and low sympathy for nationalist ideas. They show very high political knowledge and efficacy in understanding political issues but, in tendency, indicate to have little trust in politicians. We find similar expressions of demographic and attitude variables



for the ‘political internet users’, individuals showing a profile with overall above average engagement, especially on Twitter. However, in contrast to the ‘petition activists’ they appear significantly younger than the average.

The ‘Bild readers’, individuals particularly engaged with Germany’s most prominent tabloid newspaper, show above average political knowledge, but also significantly more sympathy for nationalism than the average. Finally, we find a complementary pattern for the ‘Zeit readers’, individuals particularly engaged with a prominent high quality weekly newspaper in Germany, who show especially high political knowledge but significantly lower sympathy for nationalism. Interestingly, the ‘Zeit readers’ appear significantly younger than individuals assigned to the ‘average engagement’ profile. This could be a sign that older Zeit readers may still hold a subscription for the print version instead of engaging with Zeit online.

Overall, the results support our findings from the regression models that regarded political engagement separately between engagement forms. For example, we find political knowledge to stand out for all profiles with clear political engagement pattern and we also find large overlaps in online political engagement between different profiles of users. However, again, nationalist ideology seems to be an underlying factor that may drive selection into different corners of the online public sphere. For example, while support for nationalism predicts membership in the ‘Bild reader’ profile, it is the opposite for the ‘Zeit readers’.

The major findings of the LPA are robust to a 50% split-sample validation (see SI 07 and 08). Even though the engagement profiles do not exactly mirror the original solution, the characteristics in the post-hoc tests confirm results from the regression models.

## Discussion

Deliberative processes can occur in various settings online and offline with different arenas offering different potentials for democracy. In this paper, we used a combination of digital trace data and survey data to answer the question of *who* constitutes the online public, and in particular, who engages with the complex German political online ecosystem.

In line with large segments of the offline political behavior literature (e.g. Boulianne, 2015; De Vreese & Boomgaarden, 2006; Galston, 2001), political knowledge, which in our

sample, is highly correlated with political interest and the efficacy to understand politics, is a strong predictor for the engagement with political content online. While this stands in contrast to some findings on the relationship between social media use and political knowledge (e.g. Conroy et al., 2012), we explicitly focus on the political context of social media use. Our findings also imply that, in contrast to Feezell (2016), who used survey data to examine predictors of online political participation, resource related predictors such as education and political knowledge do appear to be relevant, even if controlled for overall online engagement. Apparently, the internet does not leverage all barriers to political engagement but shows path dependencies and selection processes similar to the offline world in which those most interested in politics also consume most political information and are the most politically engaged (Moy et al., 2005).

Selection along this path dependency is potentially most pronounced when considering established journalistic media. Especially the engagement with quality journalism, often with local reference such as in public broadcasting outlets in Germany, we see a pattern of selection for a group of individuals with high income and high education. This finding is in line with the entropy based measures of inclusivity and heterogeneity that were previously found to be low for public broadcasting online services (Oswald, 2022).

However, our results provide some reason for optimism: almost everyone in the sample engaged with some political content some time during the observation period. Furthermore, in line with Guess (2021) there is a considerable overlap between users in their engagement with political content online. We already know from Oswald (2022) that large information hubs are a common source of information for the vast majority of people who engage with political content online.

We do see some differential engagement patterns along ideological lines when it comes to the consumption of specific online outlets. While the tabloid newspaper Bild appears to be particularly attractive for politically right-leaning individuals with sympathy for nationalism, the prominent weekly newspaper Zeit appears more attractive for anti-nationalist citizens. Petitions on Change.org seem to attract particularly left-leaning individuals even though, if compared to all other German parties, more radical right-wing AfD voters are significantly more active on the online petitioning platform.

Even though political knowledge is, by far, the most important factor to predict political engagement online across all types of media, we find that political discussions on communication platforms such as Facebook are most visited (and potentially driven) by individuals with more radical political orientation. For example, we find that in comparison to mainstream conservative CDU voters, people holding preferences for the right-wing AfD, show especially high engagement, a pattern that we do not see for any other party.

In contrast to online political communication research that focuses on digital traces left from active participation, for example comments in online forums or Tweets, the nature of our data allows us to also observe the large majority of more ‘passive’ internet users. While prior research on active users largely shows pronounced patterns of selection, for example along ideological extremity and demographic variables such as gender and socio-economic status (Barberá & Rivero, 2014; Hargittai, 2020; Mukerjee et al., 2022), passive usage seems to be much more equally distributed. While the vast majority of our representative sample of the German online public engages with some political content online, we speculate that only few end up leaving visible digital traces. This has implications for empirical research, but also for the state of the online public discourse, if the visible posts and comments are contributed by a politically extreme minority with a distinct demographic profile (Kim et al., 2021).

When considering the literature on deliberative systems, we have to acknowledge that we can, of course, only speak to the digital component of the public discourse. Furthermore, an issue that becomes more and more problematic with increasing mobile device uptake is that we consider desktop use only.

A logical next step for further research is now, to also consider the communicative acts of online political deliberation. As a first step, our data allowed the examination of the exposure to political content in various arenas of the complex political online ecosystem in Germany but, for example, no conclusions about whether individuals actually used forums to actively discuss political issues instead of reading comments. However, we know that most online content is produced by a minority of active users whereas the majority of passive ‘lurkers’ (Davis, 2013; Tucker et al., 2018) develops not only their political opinions but also social norms from observing others’ active engagement with political content online.

Considering the results of the latent profile analysis that clusters individuals along various engagement patterns, political orientation and political knowledge parameters, it was difficult to determine the exact number of profiles and post hoc tests showed mostly demographic differences between users' profiles, beyond the differences in political knowledge. This further indicates that there is a large overlap of in media exposure between different segments of the online public, speaking against the notion of severe ideological 'echo chambers' in the German online public sphere in 2017. However, one may speculate that large information hubs could potentially still entail different niches (potential internal 'echo chambers') but at the same time also offer the opportunity for incidental exposure to cross cutting content.

Regarding the ongoing academic debate on the role of Twitter for political polarization among Western democracies, it is interesting to see that the latent profile of Twitter users in our sample showed a unique pattern of unusually high engagement with political issues across different arenas which is, in Germany, absolutely not representative of the overall online public (Göbel, 2021).

## **Conclusion**

While existing research on online political activity that is based on actively produced digital traces, such as the comments, likes, and shares on social media platforms, suggests a heavy overrepresentation of extreme views, our findings on the individual profiles of web browsing behaviour draw a different picture.

We find that especially people with high political knowledge engage with political information online but no evidence for strong selection along other variables, such as political ideology. However, we do find that people with rather extreme political views tend to engage more with platforms that enable political expression and discussion than people with more mainstream political views. In other words, despite the fact that the moderate majority regularly consumes political information online, the potentially less well informed minority of people with more radical political orientation diverts to the communication and participation spaces - a finding that supports previous research on the misrepresentation of political opinions on social media platforms.

The suggested discrepancy between passive information consumption and active political expression online has important implications for the inferences drawn from social media data that likely can only reflect a heavily skewed tip of the iceberg of the public discourse.

Furthermore, we clearly see that not all social media platforms are created equal. While the German political Twitter user base is characterized by high political knowledge and low nationalism, the German political Facebook crowd seems considerably different. This is only one example in which a direct comparison of online discussions between platforms would be heavily biased by selection issues, which confounds assumptions about the state of the online public discourse and political polarization.

Our findings concerning online petitioning platforms and other forms of online political participation are, from a democratic perspective, probably the most disenchanting. Overall, we do not find much engagement with online petitions and other participation platforms in 2017. The existing engagement is mostly driven by people with rather extreme political orientation and low satisfaction with democracy. In other words, despite the push for more inclusion of civic perspectives in democratic decision making, those few who take up the opportunity to participate, are either those with very high political interest and knowledge and/or those with the lowest belief in democracy.

Finally, our overall consistent finding of political knowledge as strongest common characteristic for those engaging with political content in the online public sphere highlights the importance of selection issues for deliberation research. While experimental evidence on deliberation, as for example collected in large-scale deliberative polling events (J. Fishkin et al., 2021) can highlight the *potential* of deliberation under ideal conditions, observational evidence on real-life deliberation in existing online environments is crucial to assess the current *state* of the online public discourse and to construct tangible strategies to support a constructive, open public debate.

### **Acknowledgements**

Data collection for this research was funded by a grant from the Volkswagen Foundation Computational Social Science Initiative, reference 92 143.

We thank all participants of the 2022 MEOF workshop in Princeton for the thorough discussions and the invaluable feedback to the manuscript.

LO thanks the German National Academic Foundation for funding her PhD and all participants of the 2022 writing retreat in Alfter for facilitating the revision of the manuscript.

### **Supplementary material**

All supplementary materials to this project, including code for all analyses can be found in the related OSf repository under <https://osf.io/xb3cy/>. An online appendix, including supporting tables and figures, more details on sampling, data collection and ethical considerations of the data, the dictionary used to select politically relevant websites and a software statement can be found under <https://osf.io/f5g7v>.

## References

- Akogul, S., & Erisoglu, M. (2017). An approach for determining the number of clusters in a model-based cluster analysis. *Entropy*, *19*(9), 452.
- Bächtiger, A., & Hangartner, D. (2010). When deliberative theory meets empirical political science: Theoretical and methodological challenges in political deliberation. *Political Studies*, *58*(4), 609–629.
- Bächtiger, A., & Parkinson, J. (2019). *Mapping and measuring deliberation: Towards a new deliberative quality*. Oxford University Press.
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132.
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science*, *26*(10), 1531–1542.
- Barberá, P., & Rivero, G. (2014). Understanding the Political Representativeness of Twitter users. *Social Science Computer Review*, 1–18.
- Beauchamp, N. (2020). Modeling and Measuring Deliberation Online. In *The Oxford Handbook of Networked Communication*.
- Best for Planning. (2017). *Berichtsband b4p 2017*. Gesellschaft für integrierte Kommunikationsforschung.
- Boulianne, S. (2020). Twenty Years of Digital Media Effects on Civic and Political Participation. *Communication Research*, *47*(7), 947–966. <https://doi.org/10.1177/0093650218808186>
- Boulianne, S. (2015). Social media use and participation: A meta-analysis of current research. *Information Communication & Society*, *18*(5), 524–538. <https://doi.org/10.1080/1369118X.2015.1008542>
- 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.
- Bright, J., Bermudez, S., Pilet, J.-B., & Soubiran, T. (2019). Power users in online democracy: Their origins and impact. *Information, communication & society*, 1–16.
- Bright, J., Marchal, N., Ganesh, B., & Rudinac, S. (2020). Echo Chambers Exist!(But They're Full of Opposing Views). *arXiv preprint arXiv:2001.11461*.
- Bundestag, D. (2020). Deutscher Bundestag - Bundestag beschließt neue Form der Bürgerbeteiligung: Bürgerrat... Retrieved September 28, 2022, from <https://www.bundestag.de/webarchiv/presse/pressemitteilungen/2020/701614-701614>
- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658–679.
- Comission, E. (2021). How to make citizens active participants in the governance of EU cohesion policy. Retrieved September 28, 2022, from [https://ec.europa.eu/regional\\_policy/en/policy/how/improving-investment/citizens-participation/](https://ec.europa.eu/regional_policy/en/policy/how/improving-investment/citizens-participation/)
- Conroy, M., Feezell, J. T., & Guerrero, M. (2012). Facebook and political engagement: A study of online political group membership and offline political engagement. *Computers in Human behavior*, *28*(5), 1535–1546.

- Conway, M. M. (2001). Women and political participation. *PS: Political Science and Politics*, 34(2), 231–233.
- Davis, R. (2013). *Politics online: Blogs, chatrooms, and discussion groups in American democracy*. Routledge.
- De Vreese, C. H., & Boomgaarden, H. (2006). News, political knowledge and participation: The differential effects of news media exposure on political knowledge and participation. *Acta Politica*, 41(4), 317–341.
- Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729–745.
- Eady, G., Nagler, J., Guess, A., Zilinsky, J., & Tucker, J. A. (2019). How many people live in political bubbles on social media? Evidence from linked survey and Twitter data. *Sage Open*, 9(1), 2158244019832705.
- Esau, K., FleuSS, D., & Nienhaus, S.-M. (2020). Different Arenas, Different Deliberative Quality? Using a Systemic Framework to Evaluate Online Deliberation on Immigration Policy in Germany. *Policy & Internet*.
- Esterling, K. (2018). Deliberation and Experimental Design. In *The Oxford Handbook of Deliberative Democracy*. Retrieved October 22, 2021, from <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780198747369.001.0001/oxfordhb-9780198747369-e-13>
- Feezell, J. T. (2016). Predicting online political participation: The importance of selection bias and selective exposure in the online setting. *Political Research Quarterly*, 69(3), 495–509.
- Fishkin, J., Siu, A., Diamond, L., & Bradburn, N. (2021). Is Deliberation an Antidote to Extreme Partisan Polarization? Reflections on America in One Room. *American Political Science Review*, 115(4), 1464–1481. <https://doi.org/10.1017/S0003055421000642>
- Fishkin, J. S. (2018). *Democracy When the People Are Thinking P: Revitalizing Our Politics Through Public Deliberation*. Oxford University Press.
- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, 80(S1), 298–320.
- Friess, D., & Eilders, C. (2015). A systematic review of online deliberation research. *Policy & Internet*, 7(3), 319–339.
- Galston, W. A. (2001). Political knowledge, political engagement, and civic education. *Annual review of political science*, 4(1), 217–234.
- Göbel, S. (2021). Voting and Social Media-Based Political Participation. *Journal of Quantitative Description: Digital Media*, 1.
- Guess, A. M. (2021). (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets. *American Journal of Political Science*.
- Habermas, J. (2021). Überlegungen und Hypothesen zu einem erneuten Strukturwandel der politischen Öffentlichkeit. In *Ein neuer Strukturwandel der Öffentlichkeit?* (pp. 470–500). Nomos Verlagsgesellschaft mbH & Co. KG.
- Habermas, J., Lennox, S., & Lennox, F. (1974). The public sphere: An encyclopedia article (1964). *New German Critique*, (3), 49–55.



- Hargittai, E. (2020). Potential biases in big data: Omitted voices on social media [Publisher: SAGE Publications Sage CA: Los Angeles, CA]. *Social Science Computer Review*, 38(1), 10–24.
- Jacquet, V. (2017). Explaining non-participation in deliberative mini-publics. *European Journal of Political Research*, 56(3), 640–659.
- Jungherr, A., & Schroeder, R. (2021). Digital Transformations of the Public Arena [ISBN: 9781009064484 9781009065542 Publisher: Cambridge University Press]. *Elements in Politics and Communication*. <https://doi.org/10.1017/9781009064484>
- Kahne, J., Lee, N.-J., & Feezell, J. T. (2012). Digital media literacy education and online civic and political participation. *International Journal of Communication*, 6, 24.
- Kim, J. W., Guess, A., Nyhan, B., & Reifler, J. (2021). The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity [Publisher: Oxford University Press]. *Journal of Communication*, 71(6), 922–946.
- Lelkes, Y. (2016). Mass polarization: Manifestations and measurements. *Public Opinion Quarterly*, 80(S1), 392–410.
- Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2021). Digital Media and Democracy: A Systematic Review of Causal and Correlational Evidence Worldwide.
- Mansbridge, J., Bohman, J., Chambers, S., Christiano, T., Fung, A., Parkinson, J., Thompson, D. F., & Warren, M. E. (2012). A systemic approach to deliberative democracy. In J. Parkinson & J. Mansbridge (Eds.), *Deliberative Systems* (pp. 1–26). Cambridge University Press.
- Moy, P., Manosevitch, E., Stamm, K., & Dunsmore, K. (2005). Linking dimensions of Internet use and civic engagement [Publisher: SAGE Publications Sage CA: Los Angeles, CA]. *Journalism & Mass Communication Quarterly*, 82(3), 571–586.
- Mukerjee, S., Jaidka, K., & Lelkes, Y. (2022). The Political Landscape of the US Twitterverse [Publisher: Taylor & Francis]. *Political Communication*, 1–31.
- Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press.
- OECD. (2021). Innovative Citizen Participation - OECD. Retrieved September 28, 2022, from <https://www.oecd.org/governance/innovative-citizen-participation/>
- Oser, J., & Boulianne, S. (2020). Reinforcement effects between digital media use and political participation : A meta-analysis of repeated-wave panel data. *Public Opinion Quarterly*, 84(S1), 355–365. <https://doi.org/10.1093/poq/nfaa017>
- Oswald, L. (2022). More than News! Mapping the Deliberative Potential of a Political Online Ecosystem with Digital Trace Data. <https://doi.org/10.31235/osf.io/68udg>
- Oswald, L., & Bright, J. (2022). How Do Climate Change Skeptics Engage with Opposing Views Online? Evidence from a Major Climate Change Skeptic Forum on Reddit [Publisher: Routledge \_eprint: <https://doi.org/10.1080/17524032.2022.2071314>]. *Environmental Communication*, 0(0), 1–17. <https://doi.org/10.1080/17524032.2022.2071314>
- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- Parkinson, J., De Laile, S., & Franco-Guillén, N. (2020). Mapping Deliberative Systems with Big Data: The Case of the Scottish Independence Referendum. *Political Studies*, 0032321720976266.

- Persson, M. (2015). Education and political participation. *British Journal of Political Science*, 689–703.
- Peterson, E., Goel, S., & Iyengar, S. (2021). Partisan selective exposure in online news consumption: Evidence from the 2016 presidential campaign [Publisher: Cambridge University Press]. *Political science research and methods*, 9(2), 242–258.
- Prewitt, K. (1965). Political socialization and leadership selection. *The Annals of the American Academy of Political and Social Science*, 361(1), 96–111.
- Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, 16, 101–127.
- Putnam, R. (1993). *Making Democracy Work: Civic Traditions in Modern Italy* Princeton: Princeton Univ. Press.
- Quintelier, E., & Hooghe, M. (2012). Political attitudes and political participation: A panel study on socialization and self-selection effects among late adolescents. *International Political Science Review*, 33(1), 63–81.
- Rosenberg, J. M., Beymer, P. N., Anderson, D. J., Van Lissa, C., & Schmidt, J. A. (2019). tidyLPA: An R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *Journal of Open Source Software*, 3(30), 978.
- Springer, N., Engelmann, I., & Pfaffinger, C. (2015). User comments: Motives and inhibitors to write and read. *Information, Communication & Society*, 18(7), 798–815.
- Stoker, L., & Jennings, M. K. (1995). Life-cycle transitions and political participation: The case of marriage. *American political science review*, 421–433.
- Stromer-Galley, J. (2003). Diversity of political conversation on the Internet: Users' perspectives. *Journal of Computer-Mediated Communication*, 8(3), JCMC836.
- Sunstein, C. R. (2002). The law of group polarization. *Journal of political philosophy*, 10(2), 175–195.
- Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)*.
- Uslaner, E. M. (2002). *The moral foundations of trust*. Cambridge University Press.
- Vaccari, C., & Valeriani, A. (2018). Dual screening, public service broadcasting, and political participation in eight western democracies. *The international journal of press/politics*, 23(3), 367–388.
- Wessler, H. (2008). Investigating deliberativeness comparatively [Publisher: Taylor & Francis]. *Political communication*, 25(1), 1–22.
- Wilhelm, A. G. (2000). *Democracy in the digital age: Challenges to political life in cyberspace*. Psychology Press.
- Wright, S., & Street, J. (2007). Democracy, deliberation and design: The case of online discussion forums. *New media & society*, 9(5), 849–869.
- Zhang, W., Cao, X., & Tran, M. N. (2013). The structural features and the deliberative quality of online discussions. *Telematics and informatics*, 30(2), 74–86.

## **P4. Automating the Analysis of Online Deliberation? A Comparison of Manual and Computational Measures Applied to Climate Change Discussions.**

### **Preprint Publication**

Oswald, L. (2022). Automating the Analysis of Online Deliberation? A Comparison of Manual and Computational Measures Applied to Climate Change Discussions. <https://osf.io/preprints/socarxiv/qmzwx/>

### **Supplementary Materials**

Supplementary information including documented code, codebook for manual content analysis and supplementary tables and figures available under <https://osf.io/d8ncj/>.

# Automating the Analysis of Online Deliberation?

## A comparison of manual and computational measures applied to climate change discussions

Lisa Oswald  
Hertie School

### Abstract

Facing the importance and the sheer volume of online discussions, reliable computational approaches to assess the deliberative quality of online discussions at scale would open a new era of deliberation research and create opportunities for deliberation centered platform regulation. But to what extent is it possible to automate the assessment of deliberative quality? I compare automated measures of text features and discussion thread structures to established manual content analysis with an application to online discussions on ‘Reddit’ that deal with the 2020 wildfires in Australia and California. I further compare discussions between two ideologically opposite online communities, one featuring discussions in line with the scientific consensus and one featuring climate change skepticism. While no single computational measure can capture the multi-dimensional concept of deliberative quality, I find that (1) automated measures of structural complexity capture engagement and participation as preconditions for deliberation, (2) automated toxicity scores predict manual measures of respect, and (3) the length of comments in combination with the absence of toxicity predicts manual measures of argumentation. While the presented computational approaches cannot replace in-depth content coding, the findings imply that selected automated measures can be useful, scalable additions to the measurement repertoire for specific dimensions of online deliberation. I discuss implications for platform regulation and communication research and suggest interdisciplinary collaboration to synthesize past content coding efforts using machine learning.

*Keywords:* Online discussions, deliberation, automated assessment, content analysis, network analysis, climate change skepticism, Reddit

## Introduction

Within countless arenas, from the comment sections of mass media news outlets, over mainstream social media platforms to niche online forums, people exchange opinions and arguments on political, societal and scientific issues. These discussions vary systematically in complexity and deliberative quality across different arenas of public debate (Esau et al., 2020). A high degree of deliberative quality is often assumed to enhance democratic outcomes, for example by mitigating opinion polarization (Grönlund et al., 2015). However, the current state of the online public discourse is a contested issue. While concerns about online incivility accumulate (see e.g. Coe et al., 2014; Frimer et al., 2022; Hmielowski et al., 2014), it remains challenging to empirically operationalize deliberative quality in unregulated communication processes online (Beauchamp, 2020; Habermas, 2021). Facing the importance and the sheer volume of online discussions, reliable computational approaches to assess their deliberative quality at scale would open a new era of deliberation research and open opportunities for platform regulation such as deliberation-centered algorithm optimization. But is it possible to automate the assessment of deliberative quality?

The deliberative quality of a discussion has traditionally been assessed with in-depth manual content coding approaches that require time and human effort and that cannot be scaled appropriately to the vast availability of online communication data (Esau et al., 2017; Friess et al., 2020; Steenbergen et al., 2003). Gonzalez-Bailon et al. (2010) developed an approach to automate the analysis of online discussion threads with regard to their structural complexity. This measure was soon adopted by other researchers as a proxy for online deliberation (see e.g. Aragón et al., 2017). While in the past years various automated measures of online deliberation, based on simple to detect features, were developed and critiqued (Beauchamp, 2020), a formal quantitative comparison with an established benchmark - for example, the Discourse Quality Index (DQI, Steenbergen et al., 2003) - is still missing.

After decades of refining manual content coding approaches to carve out a set of core deliberation dimensions accompanied by a plethora of additional facets (Friess & Eilders, 2015; Graham, 2012; Steenbergen et al., 2003) it remains an open question to what extent the established dimensions of deliberation can be measured using computational approaches

and how specific dimensions relate to different computational measures.

In this paper, I analyze the structural complexity and deliberative quality of online discussion threads on ‘Reddit’. Focusing on a specific set of discussions dealing with the wildfires in Australia and California, I juxtapose two methodological approaches, (1) the established method of manual content coding to assess the deliberative quality of a discussion, and (2) the automated analysis of discussion threads regarding their structural complexity and simple textual features. Finally, I discuss the potentials of machine learning approaches for the synthesis of prior classification efforts.

The process and outcomes of deliberation also depend on the source and quality of evidence that is used to justify arguments (e.g. Fishkin, 2018). Besides the methodological comparison that is the focus of this paper, it is not yet clear how deliberative quality relates to the degree to which communication is evidence-informed. In consequence, I question how deliberative quality appears in the context of polarized issues in which at least one camp leaves the ground of established facts. In light of the catastrophic wildfire season of 2020, I compare discussions on climate change with regard to the wildfires in two forums: one forum aligning with the scientific consensus (Cook et al., 2016) and one climate change skeptic forum on ‘Reddit’.

In summary, holding the theoretical framework of deliberative democracy and the analytical depth of manual content coding as benchmark, and focusing on an important real-life example, this study aims to assess the potential of simple computational approaches to better understand the quality of online discussions at scale.

### **Measuring Deliberation**

Deliberation can be measured on various levels: from a macroscopic perspective that takes into account a whole system, for example a country, like the deliberative democracy indicator in the Varieties of Democracy Project (Bühlmann et al., 2012; Coppedge et al., 2020; FleuSS & Helbig, 2020), or from a micro perspective that analyses the deliberative quality of certain communicative acts within a specific arena of discourse. Today, by far, most research efforts have been taken on the micro level (e.g. Coleman & Moss, 2012; Esau et al., 2020; Follesdal, 2010; Janssen & Kies, 2005; Martin, 2013; Mendonça et al., 2020;

Roos et al., 2020). This project focuses on procedural measures, also called communicative “throughput” (Friess & Eilders, 2015), at the heart of deliberation (Beauchamp, 2020).

Automated measures of deliberation are still rare but growing (Beauchamp, 2020) and most approaches center either around text or around network structures (e.g. Brundidge, 2010; Choi, 2014; Eveland & Kleinman, 2013; Himmelboim, 2011). In the following, I will elaborate in more detail on selected approaches relevant for this project. A systematic overview of these approaches with each group of approaches aiming to measure different dimensions of deliberation can be found in Table 1.

**Content coding.** The traditional approach to assess the deliberative quality of a discussion, either face-to-face or online, is through systematic content coding approaches applied to discussion threads or speech transcriptions (Friess & Eilders, 2015). A widely used instrument is an overarching coding scheme, the Discourse Quality Index established by Steenbergen et al. (2003), that can be applied to various contexts. The DQI relies on coding categories that closely follow the principles of Habermas’ discourse ethics, that is, among others, the degree and content of justification, the degree of respect, and the constructive value of an individual speech act (Steenbergen et al., 2003). Similar coding schemes are used until today (Friess et al., 2020).

Discussions can be characterized according to various criteria for deliberative quality (Friess & Eilders, 2015). While Strandberg and Grönlund (2018) consider reciprocity, respect, reflection and justification, Graham (2012) for example used a more inclusive conceptualization of deliberation including the concepts of rational-critical debate, coherence and continuity, reciprocity, reflexivity, empathy and discursive equality. More recent research appreciates this development as it seems to be better suited for real-life online interactions and the classification of spontaneous deliberation in non-institutional settings, such as online forums or social media platforms.

While such manual content coding approaches can capture various aspects of deliberative quality by acknowledging the depth of human communication and potentially allowing for some degree of interpretation, they take a lot of time and effort which makes it hard to keep up with the speed of data generation in online communication spaces.

**Table 1***Measuring Online Deliberation*

Approach	Data basis	Dimensions captured	Operationalization	Exemplary publications
<b>Manual</b>	<b>Content</b>	Justification, respect, constructive value	<b>Manual consideration of individual speech acts</b> (e.g. comments)	Steenbergen et al. (2003)
		Rationality, constructiveness, politeness, civility, reciprocity		Friess et al. (2020)
		reciprocity, respect, reflection and justification		Strandberg and Grönlund (2018)
		Rational-critical debate, coherence and continuity, reciprocity, reflexivity, empathy and discursive equality		Graham (2012)
<b>Automated Structure</b>		Representation and argumentation	<b>Discussion thread width and depth</b>	Aragón et al. (2017), Gómez et al. (2008), and Gonzalez-Bailon et al. (2010)
		Participation, deliberation and equality	Network analytic approach in combination with time series data	Shin and Rask (2021)
		Veracity of rumors; different types of media	Average depth-to-breadth ratio of retweet networks; diffusion cascades	Goel et al. (2016), Juul and Ugander (2021), and Vosoughi et al. (2017), Vosoughi et al. (2018)
<b>Automated Text</b>		Respect	<b>Automated toxicity classification</b>	Davidson et al. (2017) and Vidgen et al. (2019)
		Respect	Patterns of interruption, expressions of doubt	Gold et al. (2017) and Zyma (2014)
		Reciprocity	<b>Number of replies</b>	Janssen and Kies (2005) and Schneider (1997)
		Argumentation	<b>length of comments</b> , the ratio of typos, the complexity of vocabulary	Jennstål (2019)
		Argumentation	Causal connectors such as because	Gold et al. (2017)

*Note.* Empirical examples and suggestions from the literature. References exemplary. Bold face indicates measurement approaches that were tested in this paper.



**Discussion thread structure.** A family of automated approaches that does not require a consideration of single comments but operates on the discussion thread level are analyses of network structures. Gonzalez-Bailon et al. (2010) focus on structural aspects of online discussions when analyzing forums on ‘Slashdot’ with hierarchical network analysis. They found that online political discussion networks are generally more complex, namely wider and deeper than discussions about non-political topics, such as gaming. The criterion of ‘width’ captures the number of people engaged in the discussion, while the ‘depth’ represents the number of comment levels throughout the discussion cascades. Gonzalez-Bailon et al. (2010) understand width and depth as necessary conditions for deliberation, with width as measure for representation and depth as measure for argumentation, based on the Madisonian model by Ackerman and Fishkin (2008).

Gómez et al. (2008) also analyze discussion threads on Slashdot using a network analytical approach and propose an adapted version of the H-index (Hirsch, 2005) as indicator for the subjective controversy provoked by a post but did not explicitly link their findings to the concept of deliberation.

Aragón et al. (2017) examine the effect of the modification of a platform feature on the level of deliberation on the Spanish social media platform ‘Menéame’ using the same structural metric as Gonzalez-Bailon et al. (2010). They found that over a course of five years, threads progressively become deeper but less wide. This trend changes discontinuously in January 2015, when the comment view changes from a linear display of comments to a hierarchical view of comment threads. Following this change in design features, the average thread width remains stable while the average depth grows, especially with the introduction of a second version of the hierarchical view in which the maximum visual depth was increased. Even though Aragón et al. (2017) observed communication dynamics on the platform as they occurred, their findings provide interesting implications for platform design governance. However, it remains unclear how the structural change related to various dimensions of deliberative quality, such as incivility or argumentation.

Shin and Rask (2021) use an alternative network analytic approach in combination with time series data in order to automate the assessment of deliberative quality along the dimensions of participation, deliberation and equality, as proposed by Fishkin (2011).

However, they call for future research to develop automated indicators that complement qualitative investigation (Shin & Rask, 2021).

Finally, in an influential project on the spread of rumors on Twitter, the average depth-to-breadth ratio of retweet-networks was used as information on the nature of the diffusion of rumors, which, according to the authors, can be used to predict the veracity of rumors (Vosoughi et al., 2017; Vosoughi et al., 2018). However, this approach did not remain uncontested (Juul & Ugander, 2021). While true news and false news do not appear to differ structurally but only in size, the diffusion cascades of different types of media (video, images, petitions and news) appear to underlie structurally different diffusion processes (Goel et al., 2016). Even though retweet-networks differ in some regards from comment cascades on Reddit or Slashdot, these projects also relate structural features, the width and depth of discussion networks, to qualitative principles of online communication.

To this day, an evaluation of network-structural measures through a systematic comparison with established methods of deliberative quality assessment is, at the time of writing and to my best knowledge, lacking. This paper aims to fill this gap.

**Textual features of online discussions.** Referring back to the modern theoretical conception of deliberation, it is certainly not easy to develop comprehensive indicators of deliberative quality in an automated manner (FleuSS et al., 2018; Janssen & Kies, 2005; Jonsson & Åström, 2014). However, some automated measures of specific dimensions of online deliberative quality have been suggested in the literature.

The *civility* of a discussion, including aspects such as respectful listening, respect for groups, respect towards the demands of others and respect towards the counterarguments of others (Janssen & Kies, 2005; Steenbergen et al., 2003) might be captured with methods of automated toxicity or hate-speech classification (Davidson et al., 2017; Vidgen et al., 2019). Furthermore, patterns of interruption can be linguistic markers of lacking respect (Gold et al., 2017). Previous research also identified rhetorical questions that contain focus particles such as *even* and *ever* to express doubts of the overall competence of a conversational counterpart as indicators of disrespect. However, despite the conceptual sophistication, those markers of incivility do not appear to be reliably detectable with automated approaches (Gold et al., 2017; Zymla, 2014).

*Reciprocity* appears to be a dimension of deliberative quality that is comparably straight-forward to identify with automated measures. When trading the qualitative nature of the interaction for simplicity, a common approach is to count the number of replies a post receives (Janssen & Kies, 2005; Schneider, 1997). While even manual content coding approaches do not necessarily capture the valence of reciprocity between comments, automated measures based on comment counts, of course, neither capture valence nor do they distinguish between reciprocal comments and isolated statements.

Approaches to assess the level of *argumentation* are manifold. One could simply consider the length of comments, the ratio of typos or the complexity of vocabulary - similar to the complexity-of-thinking approach (Jennstål, 2019). However those rather conservative approaches are likely confounded with the sophistication of speech and therefore the socio-economic background of the speaker. The justification of a statement by arguments, personal opinion or experience, facts and figures, group interests or the common good (Jensen, 2003; Steenbergen et al., 2003) appear to be more difficult to assess in an automated manner. However, links to external resources such as blog posts, news articles or research papers might be one attempt to operationalize one dimension of justification in online communication (Oswald & Bright, 2021). Furthermore, causal connectors such as because and the expression of common ground could be used to measure argumentation and justification. However, these markers face several challenges related to the ambiguity of indicators, for example in the German language (Gold et al., 2017).

While many text based automated approaches are currently suggestions from the literature without extensive examples of practical implementation, they clearly aim to operationalize specific established dimensions of deliberative quality.

### **Approach & Scope**

With automated measures of deliberation emerging while manual content coding approaches involve enormous efforts and time, two research questions arise in the context of this example case. First, *can automated measures such as the structural complexity of online discussion threads and simple textual features be used to assess online deliberation?* Second, regarding the conceptual definition of deliberation (in)forming public opinion, *are discus-*

*sions aligning with the scientific consensus on climate change showing a higher deliberative quality compared to discussions featuring climate change skepticism?*

Building upon the approaches of Gonzalez-Bailon et al. (2010) and Gómez et al. (2008), I analyze the discussion threads of two different online communities on Reddit, dealing with the 2020 wildfires in Australia and California, regarding their structural complexity as indicator for deliberation. Furthermore, I test the features of comment length, the number of replies per post and an automated measure of toxicity (Votta, 2019) as computational proxies of argumentation, reciprocity and respect. To provide an established benchmark for comparison, I analyze the deliberative quality of submissions and comments using a systematic content coding approach (Friess et al., 2020) that includes the criteria of argumentation, reciprocity and respect. Regarding the difference between the two online communities, I also consider the content of the discussions with regard to the attribution of the fires to climate change.

I argue that the apparent superficiality of automated measures can neither be claimed nor productively discussed if not explicitly—not only conceptually but also empirically—compared to established content coding, applied to the same material. This paper tries to add a first empirical piece to this discussion.

## Methods

### Data

Data was collected from *Reddit*, the social news aggregation, content rating and discussion platform. In contrast to social media platforms such as Facebook or Twitter, Reddit is a self-governed community space which makes it especially interesting for the study of deliberation.

On Reddit, users are registered under a pseudonym and post text, links to blog posts and news articles, images, or videos onto any of thousands of subreddits, pages dedicated to certain topics. Meanwhile, other users read, upvote or downvote, and comment on the submissions. Reddit is a highly frequented website and most content is publicly available which makes it an interesting resource of organic content for social scientists (Amaya et al., 2019). However, similar to social media platforms, the sociodemographic profile of users

is not comparable to the general population. During the registration process, no personal information is collected and the demographics of the 330 million active users can only be estimated within empirical survey investigations. Their results suggest that users are more likely to be male and younger than the general population (Duggan & Smith, 2013; Singer et al., 2014). The subreddits in focus feature English content and discussions but do not allow inferences on the users' nationalities.

Submissions and comments from two ideologically opposite subreddits, *r/climatechange* and *r/climateskeptics* were accessed through Reddit's API<sup>1</sup>. With more than 29,000 users each, both subreddits are the largest forums on Reddit for the respective topics of climate change and climate change skepticism. The communities are open to everyone, but the content is moderated by four to five users that are highly engaged with their forum. They have the right to delete submissions and comments, and to ban users from the subreddit.

The discussions in focus take place in ideologically opposing forums on Reddit with different positions regarding the scientific, factual baseline. The majority of scientists clearly link the 2020 Australian wildfires, which burned more than 500,000 hectares and killed at least 29 people in one of the worst wildfire seasons in Australian history to progressing climate change (Bruce-Lockhart & Romei, 2020). This is equally the case for the 2020 Californian wildfires, burning more than 1.6 million hectares, damaging or destroying over 10,000 houses, and causing at least 31 fatalities until November 2020 (Jones et al., 2020; State of California, 2020). Despite a vast scientific consensus (Cook et al., 2016), especially in the US, climate change is one of the most polarized issues in the public discourse that is potentially delaying urgently necessary political action (Douglas et al., 2017; Häkkinen & Akrami, 2014; Poortinga et al., 2011).

In order to limit the set of discussions to a comparably narrow topic, to be collected, the original submission of the discussion threads had to include the word 'fire'. The dates

---

<sup>1</sup>Reddit provides support for script-based use through an API, which is wrapped in the Python package PRAW. No limit was set to the data collection, but Reddit imposes strict limitations on request frequency in its API (no more than two requests per second) and on the number of objects the API returns for a single request (100-1500 requests depending on the type of object and user level). Therefore, the data collected for this analysis is a sample from Reddits listing of the comments to the hot submissions for the respective month.

**Table 2**

*Sample description*

subreddit	wave	comments	threads
Climate change	Australia	567	28
	California	169	18
Climate skeptics	Australia	95	13
	California	611	48
total		1,442	107

*Note.* Full sample, excluding discussions not dealing with the 2020 wildfires.

for submissions and comments for the Australian Wildfires range from November 2019 to March 2020, which corresponds to the acute period of the Australian wildfires, including some discussion in the aftermath. Respectively, the collection period for the Californian wildfires was between August and October 2020. Deleted comments were kept in the data set to analyze the original structure of the discussion thread. The distribution of comment and submission counts for both, subreddit and waves, can be found in Table 2. Threads and comments are not regarded separately between waves, however, it is interesting to note that, within the time frame, the ‘first wave’ of wildfires in Australia in 2020, was mainly discussed within the r/climatechange community, whereas the ‘second wave’, about half a year later in California, was mostly discussed within the subreddit of climate change skeptics.

The overall data set consists of 1,543 comments, written by 572 different authors, and belonging to 115 threads. In total, 7 threads did not contain discussions on the 2020 wildfires, but the word ‘fire’ was, for example, used in a metaphorical way and one original submission was deleted, which is why the information about the structure of one thread is incomplete. Comments belonging to these 8 discussion threads were excluded from the analyses. The sample only consisting of comments that refer to submissions dealing with the 2020 wildfires includes  $N = 1,442$  comments written by 539 different authors. These comments belong to 107 different threads that were triggered by submissions from 58 different authors. Out of the 107 original submissions, a subset of 20 submissions opposes the mainstream opinion of the respective subreddit, for example, climate change skeptics posting misinformation in r/climatechange or people trying to oppose climate change skeptics

and convince them of the scientific consensus in r/climateskeptics.

## Measures

### *Manual Content Coding - Deliberative Quality*

The content of 107 threads (the full sample), including the original submission and all related 1,442 comments was manually coded using a systematic manual content analysis approach (Herring, 2009; Mayring, 2014). To assess the deliberative quality of a discussion thread, an adapted, context independent, coding scheme on the basis of Friess et al. (2020)<sup>2</sup> was used, which follows a similar scheme to the established and widely used Discourse Quality Index (DQI) by Steenbergen et al. (2003). Similar to the approach of Esau et al. (2020) all core criteria of deliberative quality (argumentation, respect and reciprocity) are included but also the additional criteria of empathy, emotion and humor were coded. Furthermore, the overall expressed stance on the relationship between climate change and the 2020 wildfires was coded for each comment to allow the exploration of the relation between deliberative quality and evidence-based communication. For an extensive description of categories, see the codebook in SI ???. The coded categories with summary statistics for both subreddits are displayed in Table 4.

In line with the content coding procedure of Esau et al. (2020), the full sample of 1,442 comments was coded by the principal investigator. The inter-rater reliability was computed for all variables within a subset of 300 comments from both communities after the consultation of a trained second coder. There are overall very high agreement rates, however, some criteria come with highly skewed base rates<sup>3</sup>. To display agreement rates transparently in context, the agreement rates, base rates and a robust indicator of inter-coder agreement for skewed base rates, as suggested by Xu and Lorber (2014) are presented in Table B2.

---

<sup>2</sup>Unpublished coding scheme provided by the authors.

<sup>3</sup>Agreement for categories with very low frequency of appearances is heavily penalized by agreement indicators such as Krippendorff's Alpha and Cohen's Kappa. For example, both raters do not find a single example of racism (inverted sub-criterion for respect) which gives us an agreement rate of 1 (100% agreement) with a zero-base-rate of 1 (100% zero scores) and a Krippendorff's Alpha of 1 (perfect agreement). Instead, in one out of 300 comments, one rater finds an example of sexism, which the other rater does not agree with (this rater does not find any examples of sexism). While still having a very high agreement rate (above 0.99, 99% agreement), in this case, Krippendorff's Alpha drops to 0, due to a highly skewed base rate (1:299).

For the further analyses, I focus on the three core components of argumentation, reciprocity and respect which were constructed as summary scores from the raw categories. More details on the content coding procedure and the construction of scales can be found in the SI.

### *Computational approaches to assess deliberation*

**Structural complexity.** As outlined previously, Gonzalez-Bailon et al. (2010) assess patterns of deliberation in an automated form, using hierarchical network analysis. Their efficient structural measure of deliberative quality based on the dimensions of width and depth of discussion threads was also used and refined by other researchers (Aragón et al., 2017; Gómez et al., 2008).

In order to determine the width of discussions, the maximum width was calculated as the maximum of the sums of in-degrees of comments on each depth level within the discussion thread, which corresponds to the approach of Gonzalez-Bailon et al. (2010) (see Figure B3). A high in-degree is an indicator for many replies and the involvement of many different people in the discussion. In order to identify the depth of discussion threads, the longest path metric was used within each thread. The relation of depth and width across both subreddits and waves, taking into account the overall size of a discussion thread, is displayed in Figure B1. The combined measure of discussion complexity, the multiplication of width and depth was used.

**Textual features.** To construct an alternative and particularly easy-to-use computational measure to assess the deliberative quality of online discussion, I focused on simple textual features of the discussions. Building upon the literature on online deliberation and following previous recommended operationalizations of online deliberative quality (Janssen & Kies, 2005; Jennstål, 2019; Schneider, 1997), I chose the number of comments a submission receives as measure of reciprocity, the mean length of comments within a thread as - arguably very rough - measure of argumentation, and an easy to use toxicity classifier that is implemented in Google's Perspective API (Votta, 2019) as inverse measure of civility or respect.

While the measures appear to be very rough proxies, I assume that they have two



advantages over the structural measures of deliberation. First, they do not require the reply-to structure of discussion threads which is often not explicitly retained when collecting discussion data from the web, especially if the data was initially collected for the purpose of manual content coding. Second, the three measures operationalize three distinct dimensions of deliberative quality and are therefore more closely aligned with the theoretical conceptualization of online deliberation than a summary measure of discussion complexity.

### **Analytical Strategy**

The major research question regards how different computational measures of deliberation relate to established content coding criteria. To explore the between-paradigm relationships, first, bivariate correlations were calculated separately for all measures to provide first insight into the strength and direction of the relationship between each computational measure and their corresponding manual measure, irrespective of any additional information.

Furthermore, to explore how different automated measures can, in combination, predict manual criteria of deliberative quality, various multiple regression models were calculated. The predictor variables can correlate and contribute in different degrees to the prediction of the outcome variable. Therefore, these models can provide information on whether certain computational measures carry redundant information or whether the explanation of an outcome benefits from the combination of different computational measures.

To prevent overfitting<sup>4</sup> and prioritize parsimony of the models, lasso regression models were calculated in addition to conventional linear regression models. Applying lasso (least absolute shrinkage and selection operator) regularization to regressions aims to improve the prediction accuracy and interpretability of regression models while reducing the set of variables in a model. Lasso prevents overfitting by forcing the sum of the absolute value of coefficients below a constant to shrink certain coefficients to zero and exclude them from the model (Tibshirani, 1996). The optimal penalty parameter lambda was selected using cross-validation.

The predictive performance of a regression model is usually reported by summariz-

---

<sup>4</sup>Including too many variables into a model that can then well predict the data but will hardly generalize to other data.

ing how close on average predictions are to their expected values using various error metrics (Brownlee, 2021). Beyond the regression coefficients and the overall coefficient of determination ( $R^2$ ), the RMSE (root-mean-square error) as measure of the differences between predicted and observed values is reported based on a training data set that contains 80% of the data. Finally, the MSE (mean square error) is reported based on a test data set containing 20% of the data, to allow an out-of sample evaluation of the lasso models.

When it comes to the comparison of deliberative quality and structural measures, there is a trade-off concerning the unit of analysis. While deliberative quality is assessed on the comment level using systematic content analysis, structural measures can only be assessed on the discussion thread level. The following analyses are performed on both levels of measurement after the measures were transformed respectively (a) by computing mean scores of quality criteria over the comments of a discussion thread, and (b) by using the structural measures of the thread a comment belongs to. For consistency, all analyses in the main text use comment level data while the respective analyses using thread level data are reported in the SI.

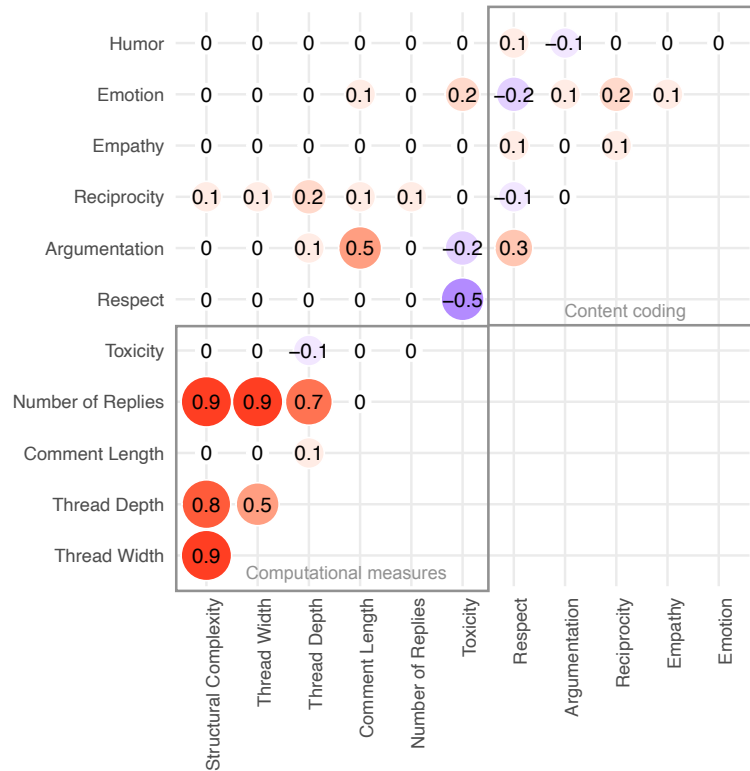
## Results

I start by describing the relationship between automated and manual criteria in the form of bivariate, multivariate and regularized models. In a second step, I examine differences between discussions of opposing ideological camps.

Concerning the within paradigm relationships, I find partially very high correlations between structural and textual features, whereas correlations within the manually coded criteria are comparably low. However, I do find two strong correlations between paradigms: a positive bivariate association between argumentation (measured using content coding) and comment length ( $r = 0.5$ ) and a negative association between respect (content coding) and toxicity (Perspective classifier, Votta (2019),  $r = -0.5$ ). Furthermore, reciprocity (measured with content coding) is with  $r = 0.2$  correlated with discussion thread depth. A more detailed description of bivariate correlations between the automated measures and the complete set of raw coding categories of deliberative quality dimensions can be found in the SI (see Fig. B4).

**Figure 1**

*Bivariate correlations between computational and content coding measures*



*Note. Bivariate Correlations between variables; structural complexity, thread width, thread depth, comment length, number of replies and toxicity are computational measures; respect, argumentation, reciprocity, empathy, emotion and humor are measures from manual content coding; comment level data.*

When considering different automated measures in combination (see Table 3), the unrestricted multiple linear regression models as well as the regularized lasso models largely mirror the initial correlative results, especially regarding the prediction of respect using the toxicity classifier (Votta, 2019). However, when considering the argumentation dimension of deliberative quality, the prediction seems to benefit from the inclusion of toxicity in addition to the length of comments. In other words, the comment lengths seems to be particularly indicative of deliberative argumentation in the absence of toxicity. Another set of models that includes interactions between the variables is reported in the supplementary information (see SI B1). The models mirror the previously described results and do not

**Table 3***Linear and Regularized Regression Models*

	Respect	Respect Lasso	Reciprocity	Reciprocity Lasso	Argumentation	Argumentation Lasso
Thread Depth	-0.02 (0.04)		0.33*** (0.04)	0.26.	0.07* (0.03)	0.03
Thread Width	0.03 (0.06)		0.16** (0.06)		0.05 (0.05)	
Comment Length	0.04 (0.03)	0.04	0.06* (0.03)	0.05	0.53*** (0.04)	0.52
Toxicity	-0.50*** (0.03)	-0.46.	0.05* (0.03)	0.04	-0.18*** (0.02)	-0.17
Number of Comments	-0.03 (0.07)		-0.28*** (0.08)	-0.06	-0.07 (0.06)	
Opposing	-0.03 (0.02)		0.05 (0.03)	0.04	0.01 (0.02)	
R <sup>2</sup>	0.25	0.26	0.06	0.06	0.32	0.36
Adj. R <sup>2</sup>	0.25		0.06		0.31	
Num. obs.	1442	1442	1442	1442	1442	1442
RMSE	0.87	0.87	0.97	0.97	0.83	0.79
MSE		0.72		0.98		0.80

\*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$

*Note.* Linear regression models estimated with HC2 robust standard errors using `lm_robust()` based on comment level data. Lasso models estimated with `glmnet()` to implement regularization. R<sup>2</sup> and RMSE in lasso models calculated based on training data (80%), MSE based on test data (20%).

show any significant interactions.

Considering the summary measures of the models, in this sample of discussions, the automated measures can predict different criteria of deliberative quality to different extents. While the automated measures in combination, but especially the automated toxicity score, explains about a quarter of the variance in respect, about a third of the variance in argumentation can be explained by automated measures, especially the length of comments (see Table 3 and B1). However, taking together all automated measures, including the depth of comment threads, not even 10% of the variance in reciprocity can be explained. The out-of-sample model evaluation metric of the MSE mirrors this pattern.

To address the secondary research question of whether discussions aligning with the scientific consensus on climate change show stronger indicators of deliberation compared to discussions among climate change skeptics, I exploited the community structure of Reddit in which, on average, users self select into groups that align with their ideologies and interests (Duguay, 2021; Jungherr et al., 2021).

According to the hand coded measure of deliberative quality, discussions in r/climateskeptics are characterized by a lower degree of deliberative quality compared to

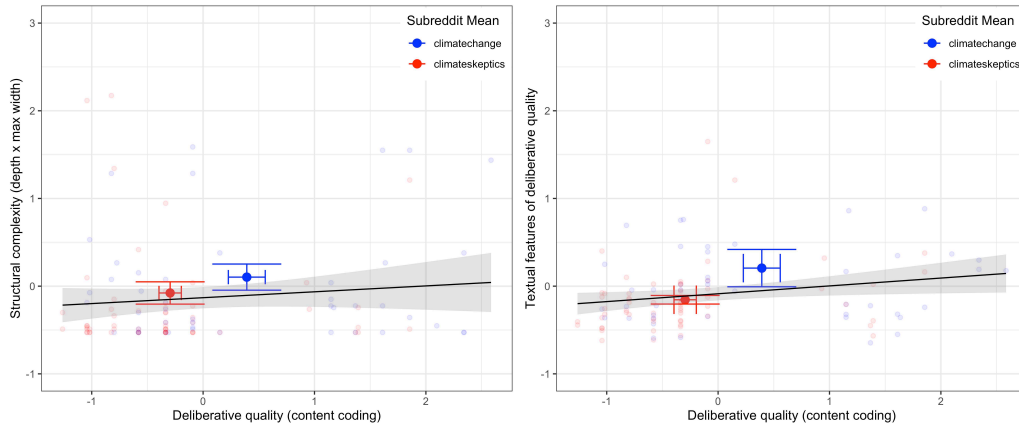
discussions in r/climatechange. Considering the structural measure with the comments as unit of analysis I come to opposite findings: discussions among people who self selected into the climate change skeptic community are characterized by higher structural complexity than discussions in r/climatechange, the community aligning with the scientific consensus on climate change (see Fig. 2). However, when considering the threads as units of analysis, the direction aligns with the finding of the manual content coding measure. This sign change in the community effects for structural measures between the levels of measurement is not surprising because large and complex threads contain many comments and are therefore weighted more strongly than less complex threads with fewer comments. However, this finding shows that researchers using the measure of structural complexity for the assessment of online discussions have to be very cautious with the level of analysis (see Figure B2).

In other words, in the sample of discussions on the 2020 wildfires and their relation to climate change, the hand coded measure of deliberative quality rather appreciates the discussions in r/climatechange as being more deliberative, whereas the structural measure rather assesses the discussions among climate change skeptics as meeting criteria of deliberative structure. However, using textual features (comment length, number of replies and toxicity) as automated measure instead, the finding that discussions in r/climatechange show signs of higher deliberative quality aligns with the established content coding measure.

Based on the more fine grained manual measures, I further explored how specific elements of deliberative quality compare between the communities of opposing ideology when it comes to the community members' stance on climate change. More specifically, the question is whether discussions in line with the scientific consensus on climate change are, for example, on average better justified and more respectful than discussions among climate change skeptics. On average, discussions among individuals exchanging views within r/climatechange are characterized by a higher level of argumentation, respect and reciprocity, the three core components of deliberative quality (see Table 4). The more inclusive criteria of deliberative quality all suffered from very low base rates. Keeping this limitation in mind, there are statistically significant differences for empathy and the expression of negative emotion, both being higher in r/climatechange.

**Figure 2**

*Online Deliberation Measurement Comparison*



*Note.* For better comparability, metrics were standardized along mean and standard deviation. Points with error bars (SE) indicate group centers of scatterplots, small points indicate comments. Linear fit line with standard errors based on comment level data. Textual features measure as mean score of three features: comment length, number of replies and toxicity (latter with negative sign)

**Table 4**

*Differences in Deliberative Quality Criteria between Communities*

	Deliberation	Argumentation	Respect	Reciprocity	Humor	Empathy	Emotion	N
<b>r/climatechange</b>								<b>736</b>
Mean	0.56	0.20	0.96	0.51	0.02	0.01	0.06	
Std	0.17	0.15	0.07	0.50	0.13	0.06	0.15	
<b>r/climateskeptics</b>								<b>706</b>
Mean	0.47	0.15	0.94	0.32	0.01	0.00	0.01	
Std	0.17	0.13	0.08	0.47	0.08	0.02	0.06	
t-value	9.80	6.23	4.96	7.66	1.83	2.88	7.93	
p-value	<.01	<.01	<.01	<.01	0.07	<.01	<.01	
df	1439	1421	1399	1439	1254	880	999	

*Note.* Comment level analysis: N = number of comments per subreddit.

### Discussion

In this study, I compared structural and simple textual features to an established content coding scheme, while applying all measures to the same material, climate change discussions dealing with the 2020 wildfires on Reddit.

As Gonzalez-Bailon et al. (2010) suggested, the structural complexity of online discussions is an efficient measure to capture certain preconditions or meta-variables for deliberation: participation or engagement. Nevertheless, the analysis of discussions on the 2020 wildfires and their relation to anthropogenic climate change suggests that, even though very scalable to the vast availability of online discussion data, structural measures do not capture classic criteria of deliberative *quality*. Instead, certain simple textual features, the length of comments and an out-of-the-box toxicity classifier, could be related to specific dimensions of deliberative quality, namely argumentation and respect. The absolute levels of variance determination are difficult to interpret in the absence of meaningful comparisons. However, keeping in mind the enormous paradigmatic distance of the measurement approaches and the extreme simplicity of most automated measures, it is encouraging to observe R-squared values between 0.26 for the complex deliberative quality dimension of respect and 0.36 for deliberative argumentation. Another advantage of the applied simple text based measures is that they place lower demands on discussion data, as they neither require any information about the network structure, nor users' methodological sophistication in handling network data.

The theory of deliberative democracy values in-depth discussions and emphasizes the value of sound justification of arguments. However, when purely relying on structural measures of online deliberation, controversial discussions around false information or conspiracy theories—as it is the case for climate change skepticism—can be overrated in terms of their deliberative value. The manual content coding approach, using components of deliberative quality from established coding schemes (Friess et al., 2020; Steenbergen et al., 2003), allows for some human interpretation of, for example, incivility components like sarcasm and irony as well as some degree of source quality. This approach appears to distinguish discussions of misinformation from discussion aligning with the scientific consensus on climate change better than structural measures. However, it is still debatable whether

deliberative quality, a procedural measure (Friess & Eilders, 2015), even if coded with a sophisticated in-depth manual content coding scheme, captures whether a discussion has ‘democratic value’, is close to the current state of scientific knowledge or even approaches socially desirable solutions. These questions are rather a matter of input and output analysis, not a matter of the deliberative process itself. Therefore, it is in turn difficult to accuse simple computational measures of failing to make this distinction.

Quantifying complex characteristics such as argumentation and justification, but especially respect and persuasion, with suitable automated markers is arguably difficult. While this work, in a way, demonstrates the simplicity of relating simple features to qualitative criteria, I would like to emphasize my concern, echoing other critiques (see e.g. Beauchamp, 2020), that limiting the analysis to easily detectable features of online discussions may risk biases in addition to the often ambiguous nature of automated measures. Therefore, when applying automated measures as primary screening for the quality of online discussion, a certain degree of caution and nuance is necessary. Evidence from this project clearly suggests that the suitability of automated measures should be considered separately between different dimensions of deliberation. It appears unrealistic to capture ‘online deliberation’ or deliberative quality as overarching construct that can be operationalized in one automated summary measure. While structural measures may map on participation and engagement, the discussed text based measures are better suitable to capture other dimensions of deliberative quality, namely argumentation and respect.

### Conclusion

Considering the currently dominating aim of digital media platforms—keeping users engaged—this analysis of discussion threads on Reddit suggests that discussions on the basis of controversial information are, at least equally lively as discussion on the basis of scientific facts, a finding that echoes previous research (see e.g. Bright et al., 2020; Vosoughi et al., 2018). This dynamic arguably limits media companies’ motivation to pro-actively work against misinformation and other harmful content. In consequence, research that monitors the state of the public discourse online, for example to inform decision makers about how to regulate problematic content, has to carefully consider issues of measurement and go



beyond engagement focused metrics.

While communication research often requires the in-depth analysis of online discussions, for regulatory purposes, it is impossible to rely on manual content coding. If regulators and/or platforms decide to optimize online environments for deliberation instead of blind engagement, assessing dimensions such as (1) how many different users participate in discussions, (2) for how many replies does a discussion continue, (3) do comments go beyond very short opinion expressions or pure emoji and (4) the level of toxicity, could be a valuable, scalable starting point to shape online environments to better serve democracy.

While Tucker et al. (2018) mention the focus on the deliberative tradition in the analysis of online communication for the lack of precise estimates for specific quantities of interest, I argue that keeping the deliberative tradition as benchmark while systematically assessing the value of automated approaches appears as promising way to enable meaningful conclusions from large-scale online communication data.

Going forward, I would like to propose a strategy to automate the assessment of online deliberative quality that literally unites established content coding with computational approaches. In particular, a synthesis of past research efforts of manual content coding through a comprehensive machine learning approach could, in a way, recycle the countless hours of work put into text labeling by trained coders. In principle, this approach is neither a novel, nor a bold idea as Fournier-Tombs and MacKenzie (2021) have already demonstrated a great proof of concept on their corpus of parliamentary speech acts and Ziegele et al. (2018) were similarly successful in their automated analysis of online incivility using hand coded social media and news comments as input for a machine learning model. Given that, for example, the coding scheme by Steenbergen et al. (2003) was used, with slight adaptations, on a great variety of platforms and contexts (e.g. Esau et al., 2020; Fournier-Tombs & Di Marzo Serugendo, 2020; Friess et al., 2020), depending on the size and quality of the labeled material, this common text corpus from previous projects of deliberative quality analysis could, in principle, serve as input for machine learning models to classify online discussion data along the core deliberative dimensions of argumentation, reciprocity, respect and constructiveness. Such a trained model could then, in turn, be made openly available for future assessments of the state of the public discourse online. In the course of

a call for such unified research efforts, I would like to echo Shin and Rask (2021) in that “automated indicators are not the end of democratic assessment but the start of collective learning” (p. 1204).

### **Acknowledgments**

### **Supplementary material**

Data collection and data pre-processing was conducted using *Python*, empirical analyses were conducted in *R*. Any code that was used in the project is provided in the OSF repository of the project under <https://osf.io/d8ncj/>. Supplementary information, including supporting tables and figures can be found under <https://osf.io/4276x>.

### **Conflict of interest**

The author declares that there is no conflict of interest.

### **References**

- Ackerman, B., & Fishkin, J. S. (2008). *Deliberation day*. Yale University Press.
- Amaya, A., Bach, R., Keusch, F., & Kreuter, F. (2019). New data sources in social science research: Things to know before working with Reddit data. *Social Science Computer Review*, 0894439319893305.
- Aragón, P., Gómez, V., & Kaltenbrunner, A. (2017). Detecting platform effects in online discussions. *Policy & Internet*, 9(4), 420–443.
- Beauchamp, N. (2020). Modeling and Measuring Deliberation Online. *The Oxford Handbook of Networked Communication*.
- Bright, J., Marchal, N., Ganesh, B., & Rudinac, S. (2020). Echo Chambers Exist!(But They're Full of Opposing Views). *arXiv preprint arXiv:2001.11461*.
- Brownlee, J. (2021). Regression Metrics for Machine Learning. Retrieved October 12, 2022, from <https://machinelearningmastery.com/regression-metrics-for-machine-learning/>

- Bruce-Lockhart, C., & Romei, V. (2020). *Australia's deadly wildfires in numbers*. Financial Times London. <https://www.ft.com/content/c068339e-3c55-11ea-b232-000f4477fbca>
- Brundidge, J. (2010). Encountering difference in the contemporary public sphere: The contribution of the Internet to the heterogeneity of political discussion networks. *Journal of Communication*, 60(4), 680–700.
- Bühlmann, M., Merkel, W., Müller, L., & WeSSels, B. (2012). The democracy barometer: A new instrument to measure the quality of democracy and its potential for comparative research. *European Political Science*, 11(4), 519–536.
- Choi, S. (2014). Flow, diversity, form, and influence of political talk in social-media-based public forums. *Human Communication Research*, 40(2), 209–237.
- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, 64(4), 658–679.
- Coleman, S., & Moss, G. (2012). Under construction: The field of online deliberation research.
- Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R. L., Verheggen, B., Maibach, E. W., Carlton, J. S., Lewandowsky, S., Skuce, A. G., Green, S. A., Nuccitelli, D., Jacobs, P., Richardson, M., Winkler, B., Painting, R., & Rice, K. (2016). Consensus on consensus: A synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*, 11(4), 048002. <https://doi.org/10.1088/1748-9326/11/4/048002>
- Coppedge, M., Gerring, J., Glynn, A., Knutsen, C. H., Lindberg, S. I., Pemstein, D., Seim, B., Skaaning, S.-E., & Teorell, J. (2020). *Varieties of Democracy: Measuring Two Centuries of Political Change*. Cambridge University Press. Retrieved July 21, 2021, from <https://www.cambridge.org/core/books/varieties-of-democracy/3E23D3D8E0A47E7A6C9537C37C6D3726>
- Dahl, D. B. (2016). *Xtable: Export Tables to LaTeX or HTML*.

- Davidson, T., Warmusley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *Proceedings of the International AAAI Conference on Web and Social Media*, 11, 512–515.
- Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017). The Psychology of Conspiracy Theories. *Current Directions in Psychological Science*, 26(6), 538–542. <https://doi.org/10.1177/0963721417718261>
- Duggan, M., & Smith, A. (2013). 6% of online adults are reddit users. *Pew Internet & American Life Project*, 3, 1–10.
- Duguay, P. A. (2021). Read it on Reddit: Homogeneity and Ideological Segregation in the Age of Social News. *Social Science Computer Review*, 08944393211001053. <https://doi.org/10.1177/08944393211001053>
- Esau, K., FleuSS, D., & Nienhaus, S.-M. (2020). Different Arenas, Different Deliberative Quality? Using a Systemic Framework to Evaluate Online Deliberation on Immigration Policy in Germany. *Policy & Internet*.
- Esau, K., Friess, D., & Eilders, C. (2017). Design Matters! An Empirical Analysis of Online Deliberation on Different News Platforms. *Policy & Internet*, 9(3), 321–342. <https://doi.org/https://doi.org/10.1002/poi3.154>
- Eveland, W. P., & Kleinman, S. B. (2013). Comparing general and political discussion networks within voluntary organizations using social network analysis. *Political Behavior*, 35(1), 65–87.
- Fasnacht, L. (2018). Mmappickle: Python 3 module to store memory-mapped numpy array in pickle format. *Journal of Open Source Software*, 3(26), 651.
- Fishkin, J. S. (2011). *When the people speak: Deliberative democracy and public consultation*. Oxford University Press.
- Fishkin, J. S. (2018). *Democracy When the People Are Thinking P: Revitalizing Our Politics Through Public Deliberation*. Oxford University Press.
- FleuSS, D., & Helbig, K. (2020). Measuring Nation States Deliberativeness: Systematic Challenges, Methodological Pitfalls, and Strategies for Upscaling the Measurement of Deliberation. *Political Studies*, 0032321719890817.

- FleuSS, D., Helbig, K., & Schaal, G. S. (2018). Four parameters for measuring democratic deliberation: Theoretical and methodological challenges and how to respond. *Politics and Governance*, 6(1), 11–21.
- Follesdal, A. (2010). The place of self-interest and the role of power in the deliberative democracy. *Journal of political philosophy*, 18(1), 64–100.
- Fournier-Tombs, E., & Di Marzo Serugendo, G. (2020). DelibAnalysis: Understanding the quality of online political discourse with machine learning. *Journal of Information Science*, 46(6), 810–822.
- Fournier-Tombs, E., & MacKenzie, M. K. (2021). Big data and democratic speech: Predicting deliberative quality using machine learning techniques. *Methodological Innovations*, 14(2), 20597991211010416.
- Friess, D., & Eilders, C. (2015). A systematic review of online deliberation research. *Policy & Internet*, 7(3), 319–339.
- Friess, D., Ziegele, M., & Heinbach, D. (2020). Collective Civic Moderation for Deliberation? Exploring the Links between Citizens Organized Engagement in Comment Sections and the Deliberative Quality of Online Discussions. *Political Communication*, 0(0), 1–23. <https://doi.org/10.1080/10584609.2020.1830322>
- Frimer, J., Aujla, H., Feinberg, M., Skitka, L., Aquino, K., & Willer, R. (2022). Incivility is rising among American politicians on Twitter.
- Gamer, M., Lemon, J., Gamer, M. M., Robinson, A., & Kendall's, W. (2012). Package irr. *Various coefficients of interrater reliability and agreement*.
- Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2016). The structural virality of online diffusion. *Management Science*, 62(1), 180–196.
- Gold, V., El-Assady, M., Hautli-Janisz, A., Bögel, T., Rohrdantz, C., Butt, M., Holzinger, K., & Keim, D. (2017). Visual linguistic analysis of political discussions: Measuring deliberative quality. *Digital Scholarship in the Humanities*, 32(1), 141–158.
- Gómez, V., Kaltenbrunner, A., & López, V. (2008). Statistical analysis of the social network and discussion threads in slashdot. *Proceedings of the 17th international conference on World Wide Web*, 645–654.

- Gonzalez-Bailon, S., Kaltenbrunner, A., & Banchs, R. E. (2010). The structure of political discussion networks: A model for the analysis of online deliberation. *Journal of Information Technology*, 25(2), 230–243.
- Graham, T. (2012). Beyond political communicative spaces: Talking politics on the Wife Swap discussion forum. *Journal of Information Technology & Politics*, 9(1), 31–45.
- Grönlund, K., Herne, K., & Setälä, M. (2015). Does enclave deliberation polarize opinions? *Political Behavior*, 37(4), 995–1020.
- Habermas, J. (2021). Überlegungen und Hypothesen zu einem erneuten Strukturwandel der politischen Öffentlichkeit. *Ein neuer Strukturwandel der Öffentlichkeit?* (pp. 470–500). Nomos Verlagsgesellschaft mbH & Co. KG.
- Hagberg, A., & Conway, D. (n.d.). *NetworkX: Network Analysis with Python*.
- Häkkinen, K., & Akrami, N. (2014). Ideology and climate change denial. *Personality and Individual Differences*, 70, 62–65.
- Herring, S. C. (2009). Web content analysis: Expanding the paradigm. *International handbook of Internet research* (pp. 233–249). Springer.
- Himmelboim, I. (2011). Civil society and online political discourse: The network structure of unrestricted discussions. *Communication Research*, 38(5), 634–659.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572.
- Hmielowski, J. D., Hutchens, M. J., & Cicchirillo, V. J. (2014). Living in an age of online incivility: Examining the conditional indirect effects of online discussion on political flaming. *Information, Communication & Society*, 17(10), 1196–1211.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *IEEE Annals of the History of Computing*, 9(03), 90–95.
- Janssen, D., & Kies, R. (2005). Online forums and deliberative democracy. *Acta política*, 40(3), 317–335.
- Jennstål, J. (2019). Deliberation and complexity of thinking. Using the integrative complexity scale to assess the deliberative quality of minipublics. *Swiss Political Science Review*, 25(1), 64–83.

- Jensen, J. L. (2003). Public spheres on the Internet: Anarchic or government-sponsored? A comparison. *Scandinavian political studies*, 26(4), 349–374.
- Jia, W., Shaw, K. A., & Martonosi, M. (2012). Stargazer: Automated regression-based GPU design space exploration. *2012 IEEE International Symposium on Performance Analysis of Systems & Software*, 2–13.
- Jones, M. W., Smith, A., Betts, R., Canadell, J. G., Prentice, I. C., & Le Quéré, C. (2020). Climate change increases risk of wildfires. *ScienceBrief Review*.
- Jonsson, M. E., & Åström, J. (2014). The challenges for online deliberation research: A literature review. *International Journal of E-Politics (IJEP)*, 5(1), 1–15.
- Jungherr, A., Posegga, O., & An, J. (2021). Populist Supporters on Reddit: A Comparison of Content and Behavioral Patterns Within Publics of Supporters of Donald Trump and Hillary Clinton. *Social Science Computer Review*, 0894439321996130. <https://doi.org/10.1177/0894439321996130>
- Juul, J. L., & Ugander, J. (2021). Comparing information diffusion mechanisms by matching on cascade size. *Proceedings of the National Academy of Sciences*, 118(46).
- Martin, J. (2013). A feeling for democracy? Rhetoric, power and the emotions. *Journal of Political Power*, 6(3), 461–476.
- Mayring, P. (2014). Qualitative content analysis: Theoretical foundation, basic procedures and software solution.
- McKinney, W. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 445, 51–56.
- McKinney, W. (2011). Pandas: A foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, 14(9), 1–9.
- Mendonça, R. F., Ercan, S. A., & Asenbaum, H. (2020). More than Words: A Multidimensional Approach to Deliberative Democracy. *Political Studies*, 0032321720950561.
- Ooms, J. (2018). *Writexl: Export Data Frames to Excel 'xlsx' Format*. <https://CRAN.R-project.org/package=writexl>
- Oswald, L., & Bright, J. (2021). How do climate change skeptics engage with opposing views? Understanding mechanisms of social identity and cognitive dissonance in an online forum. *arXiv preprint arXiv:2102.06516*.

- Poortinga, W., Spence, A., Whitmarsh, L., Capstick, S., & Pidgeon, N. F. (2011). Uncertain climate: An investigation into public scepticism about anthropogenic climate change. *Global environmental change, 21*(3), 1015–1024.
- R Core Team. (2018). R: A language and environment for statistical computing.
- Roos, C. A., Postmes, T., & Koudenburg, N. (2020). The microdynamics of social regulation: Comparing the navigation of disagreements in text-based online and face-to-face discussions. *Group Processes & Intergroup Relations, 23*(6), 902–917.
- Schneider, S. M. (1997). *Expanding the public sphere through computer-mediated communication: Political discussion about abortion in a Usenet newsgroup* (PhD Thesis). Massachusetts Institute of Technology.
- Shin, B., & Rask, M. (2021). Assessment of Online Deliberative Quality: New Indicators Using Network Analysis and Time-Series Analysis. *Sustainability, 13*(3), 1187.
- Singer, P., Flöck, F., Meinhart, C., Zeitfogel, E., & Strohmaier, M. (2014). Evolution of reddit: From the front page of the internet to a self-referential community? *Proceedings of the 23rd international conference on world wide web*, 517–522.
- State of California. (2020). Stats & Events. Retrieved October 23, 2020, from <https://www.fire.ca.gov/stats-events/>
- Steenbergen, M. R., Bächtiger, A., Spörndli, M., & Steiner, J. (2003). Measuring political deliberation: A discourse quality index. *Comparative European Politics, 1*(1), 21–48.
- Strandberg, K., & Grönlund, K. (2018). Online Deliberation. *The Oxford Handbook of Deliberative Democracy*. Retrieved October 22, 2021, from <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780198747369.001.0001/oxfordhb-9780198747369-e-28>
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1996.tb02080.x>]. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). Social media, political polarization, and political disinformation: A



- review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)*.
- Vidgen, B., Harris, A., Nguyen, D., Tromble, R., Hale, S., & Margetts, H. (2019). Challenges and frontiers in abusive content detection. *Proceedings of the Third Workshop on Abusive Language Online*, 80–93. <https://doi.org/10.18653/v1/W19-3509>
- Vosoughi, S., Mohsenvand, M. N., & Roy, D. (2017). Rumor Gauge: Predicting the Veracity of Rumors on Twitter. *ACM Transactions on Knowledge Discovery from Data*, 11(4), 50:1–50:36. <https://doi.org/10.1145/3070644>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Votta, F. (2019). peRpective: A wrapper for the Perspective API. *R*. <https://github.com/favstats>
- Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H. (2017). The tidyverse. *R package ver, 1(1)*, 1.
- Wickham, H., & Bryan, J. (2018). *Readxl: Read Excel Files*. <https://CRAN.R-project.org/package=readxl>
- Wickham, H., & Francois, R. (2015). *Dplyr: A Grammar of Data Manipulation*.
- Wickham, H., Hester, J., & Francois, R. (2017). *Readr: Read Rectangular Text Data*.
- Xie, Y. (2018). Knitr: A comprehensive tool for reproducible research in R. *Implementing reproducible research* (pp. 3–31). Chapman; Hall/CRC.
- Xu, S., & Lorber, M. F. (2014). Interrater agreement statistics with skewed data: Evaluation of alternatives to Cohens kappa. *Journal of Consulting and Clinical Psychology*, 82(6), 1219.
- Zhu, H. (2019). KableExtra: Construct complex table withkableand pipe syntax. *R package version, 1(0)*.
- Ziegele, M., Daxenberger, J., Quiring, O., & Gurevych, I. (2018). Developing automated measures to predict incivility in public online discussions on the facebook sites of established news media. *Proceedings of the 68th Annual Conference of the International Communication Association (ICA)*.

Zymla, M.-M. (2014). *Extraction and Analysis of non-canonical Questions from a Twitter-Corpus* (PhD Thesis). University of Konstanz.

## Appendix A

### Details on data structure and collection

The comment text was collected together with the reply-to structure (identified through comment ids and parent-ids). Using this procedure, a network structure of nodes (comments) and edges (reply structure) was built. Discussion (sub)threads were identified using the network metric of weakly connected components. Within one API call, PRAW collects either comments or submissions, not both belonging together. Therefore, these discussion threads are considered as sub-threads, starting with root comments, comments directly commenting a submission. Another API call collected the submissions that triggered the root-comments. They were traced back using the parent-id of the root comments. In other words, the mothers for all the small comment threads had to be identified in order to merge sibling-threads together. In a manual step, the thread-sub-thread matching was validated.

To enable the assessment of deliberative quality of the discussion threads on the 2020 wildfires with network analytical methods, the submissions and comments were not only collected in table format, but also saved as separate graph objects for each subreddit with comments as nodes and the reply structure as directed edges. Discussion trees on Reddit are an example of directed acyclic graphs (DAGs). The degree of a node is the number of connections it has to other nodes. In this case, it describes how many comments replied to a submission or comment. Usually, most of the comments get few to no replies while some comments can have a high degree, therefore, greater visibility on the platform and attract even more replies. This mechanism can result in the typical pattern of power-law or fat-tail distributions in network data. The degree of a node within an acyclic discussion tree, corresponds to the partial width of the discussion relating to one comment, or the number of engaged individuals.

## Appendix B

### Details on Content Coding and Scale Construction

Respect was computed as summary score of the ten inverted incivility items (screaming, insults, sarcasm, lie accusations, stereotypes, sexism, racism, vulgarism, paternalism and threats). The component of argumentation was computed as summary score of the ‘ra-

tionality' items (argument, serious question, topic reference, rational, moral, balanced or deliberative argument, additional knowledge and personal experience). This component also captures whether a comment has a topic reference, a concept that for example Esau et al. (2020) regarded as separate component. Concerning the reciprocity of a discussion, it was coded whether comments referred to a specific user, all other users, some third party like a media outlet or researchers whose material was cited in a post or comment, and whether there was positive or negative reciprocity (agreement or disagreement with what was posted previously). In line with agreement scores of the reciprocity sub-components, only the item capturing reciprocity between participants of a discussion was used for the computation of a deliberative quality score. Besides the core components of deliberative quality, an empathy value was computed as mean of emotional empathy and cognitive perspective taking and an emotion component, capturing the expression of negative emotion (fear, anger and sadness). All variables were scaled (centered around the mean and scaled by their standard deviation) to allow a more convenient comparison in the statistical models. The descriptive statistics of the deliberative criteria, as well as the deliberation quality summary score, are presented in Table B3.

## **P5. Effects of Preemptive Empathy Interventions on Reply Toxicity among Highly Active Social Media Users**

### **Preprint Publication**

Oswald, L. (2023). Effects of Preemptive Empathy Interventions on Reply Toxicity among Highly Active Social Media Users.

<https://osf.io/preprints/socarxiv/evdxy/>

### **Supplementary Materials**

Supplementary information including documented code, data, pre-analysis plan, experimental materials and supplementary tables and figures available under

<https://osf.io/379hz/>.

# Effects of Preemptive Empathy Interventions on Reply Toxicity among Highly Active Social Media Users

Lisa Oswald  
Hertie School

## Abstract


Can we reduce toxicity in online communication using behavioral interventions? Following the theoretical turn towards the role of empathy for democratic discourse, this preregistered survey experiment tests various changes in an online discussion environments' user interface (nudges) and an educative intervention targeting users' motivation to engage in empathy (boost) to reduce reply toxicity. Participants were recruited via Facebook Ads in the United States and Germany to acquire a sample of  $N = 2,154$  highly active social media users. Based on their reported political attitudes, participants were asked to comment on social media statements they likely disagree with. Compared to the control group, neither the empathy or perspective taking nudges nor the friction placebo reduced reply toxicity. Boosting decreased reply toxicity to some degree but the effect was not robust against the inclusion of covariates. However, all nudging interventions significantly increased the length of replies. Overall, toxicity was highest among people with high online activity. While reducing toxicity online does not seem to work via simple changes in the user interface (nudging), more complex interventions (boosting) appear more promising to preemptively reduce toxicity before important voices are forced out of the public discourse online.

*Keywords:* Political communication, toxicity, empathy, perspective taking, survey experiment, nudging, boosting

## Introduction

The public discourse is essential to public opinion formation and therefore, a crucial element of any democratic system. However, the current state of the online public discourse has given reason for concern, including phenomena of affective polarization and online hate speech (Lewandowsky et al., 2020; Lorenz-Spreen et al., 2022, e.g.).

---

Lisa Oswald  <https://orcid.org/0000-0002-8418-282X>

Correspondence concerning this article should be addressed to Lisa Oswald, Hertie School, Friedrichstraße 180, 10117 Berlin, Germany. E-mail: [l.oswald@phd.hertie-school.org](mailto:l.oswald@phd.hertie-school.org)

The prevalence and severity of toxicity in online discussions varies between platforms, contexts and measurement approaches (Coe et al., 2014) but there is large consensus that discussion toxicity can have severe adverse effects on political trust and worsen perceptions of the political outgroup, thus fostering existing political divides (Bail, 2021; Chen, 2017; Mutz, 2016; Rossini, 2019). Furthermore, high levels of toxicity in online discussions have the potential to skew the online public discourse and give more room to extreme viewpoints, while the moderate majority and already marginalized people are discouraged from discussing politics online (Bor & Petersen, 2022; Kim et al., 2021). Despite great heterogeneity in definitions, many researchers conceptualize toxicity as rude, disrespectful, or unreasonable speech that is likely to make people leave a discussion (Google, 2022).

This study tests interventions to increase the quality of online discussions by reducing reply toxicity. It builds upon deliberative democratic theory that emphasizes the value of a reflective citizenry, capable of navigating conflicting arguments in political judgements (Bächtiger et al., 2018; Chambers, 2003; Dryzek, 2002; Muradova, 2020). Following the theoretical turn towards the role of empathy for democratic deliberation (Morrell, 2007, 2010), this study focuses on the effects of affective empathy and cognitive perspective taking.

Using behavioral science interventions, this preregistered survey experiment ‘nudges’ (Thaler & Sunstein, 2009)—by implementing changes to the user interface—and ‘boosts’ (Hertwig & Grüne-Yanoff, 2017)—by targeting individuals’ motivation—two of the most central criteria of constructive political discussions, respect and argumentation. Respect is operationalized using the established machine classifier of online toxicity (Google, 2022), whereas argumentation is approximated using the length of comments (Oswald, 2022).

Participants were recruited via Facebook Ads in the United States and Germany to acquire a sample of  $N = 2,154$  highly active social media users. Based on their reported political attitudes, participants were asked to comment on social media statements they likely disagree with.

Compared to the control group, neither the empathy or perspective taking nudges, nor the friction placebo reduced reply toxicity. Boosting decreased reply toxicity to some degree but the effect was not robust against the inclusion of covariates. However, all nudging interventions significantly increased the length of replies. Overall, toxicity was highest

among people with high online activity. While reducing toxicity online does not seem to work via simple changes in the user interface, more complex interventions appear more promising to preemptively reduce toxicity before important voices are forced out of the public discourse online.

## Theory

### Toxicity in online environments

Definitions and operationalizations of comment toxicity (often also referred to as incivility<sup>1</sup>) vary between studies. A unifying core element is the expression of disrespect towards others (Chen, 2017; Coe et al., 2014; Sydnor, 2018) that can be understood as a continuum from impoliteness as mild form to hate speech on the unacceptable end of the scale (Chen, 2017). For example, toxicity can include expressions of disrespect to others by using insulting, racist, sexist and xenophobic language or personal attacks. In this study, toxicity is measured using an established machine learning classifier (Google’s Perspective API) which is trained on online comment data to detect rude, disrespectful, or unreasonable comments that are likely to make people leave a discussion.

Incivility in comments to political messages can also have an entertaining function, evoking more positive emotions than raw political information online (Kosmidis & Theocharis, 2020). However, toxicity in political discussions among citizens can polarize political opinion (Anderson et al., 2014) and decrease open-mindedness and political trust (Borah, 2014). Therefore, changing public discourse norms towards less hostility and polarizing rhetoric, especially on social media, was identified as important way to reduce partisan animosity, a recently growing phenomenon with the potential to erode societal trust and support for democracy (Hartman et al., 2022).

The salient public perception that the online political discourse is more hostile than face-to-face discussions, is captured under the term ‘hostility gap’ Bor and Petersen (2022). However, there is no evidence for the hypothesis that online environments can make anyone a ‘troll’<sup>2</sup>, for example through platform characteristics like anonymity. Instead, Bor

---

<sup>1</sup>The terms toxicity and incivility are used interchangeably throughout this manuscript.

<sup>2</sup>A person who frequently posts inflammatory, insincere, digressive or off-topic messages on social media, in online forums or comment sections (Wikipedia, 2022)



and Petersen (2022) found that specific individuals with more extreme viewpoints deliberately decide to engage in hostile online discussions. Apparently, they behave equally hostile online and offline but gain more visibility through the public nature of online spaces. Accordingly, the fact that aggressive discursive strategies can be pursued with low cost and low possibility of sanctions online specifically attracts status-seeking individuals with hostile communication habits into online environments (Bor & Petersen, 2022).

Kim et al. (2021) found similar patterns of selection into toxic commenting on social media. More specifically, they found skewed participation in online political discussions: those who select into commenting are, on average, more interested in politics, have more polarized opinions and use more toxic language. Furthermore, they provide evidence for an inherently problematic logic of social media, namely that toxicity drives engagement on social media. For example, toxic comments attract more ‘likes’ on Facebook.

### **Empathy and perspective taking from a psychological perspective**

In short, empathy lets individuals share emotions while perspective taking is a cognitive mechanism of understanding others (Stietz et al., 2019). While empathy and perspective taking are, strictly speaking, two separate psychological processes, they are often conceptualized as two sides of the same coin.

Social psychological work on inter-group perspective taking identifies a broad set of positive effects of actively imagining others’ experiences, perspectives and feelings: more favorable implicit and explicit intergroup evaluations, positive non-verbal behaviors, reduced reliance on stereotype-maintaining mental processes, heightened recognition of intergroup disparities as well as the reduction of intergroup conflict (Klimecki, 2019; Todd & Galinsky, 2014). They emphasize that the psychological mechanisms can be both, affective by evoking empathy, and cognitive in the form of shifts in attributional thinking and merging of mental representations of the self and the other.

Epley and Caruso (2009) highlight that perspective taking is rarely automatic but needs to be activated explicitly in most cases. They identify three critical barriers to effective perspective taking: activating the ability of mental simulation, adjusting an egocentric default, and accessing accurate information about others.

Empathy can be understood as natural antidote to aggression (Lim et al., 2015). However, face-to-face settings might evoke empathy and perspective taking to a greater extent than online settings (Baek et al., 2012). Coupled with the observations of Kim et al. (2021) on social media dynamics that structurally foster toxicity, this implies room and relevance for designing online interventions.

### **Empathy and perspective taking from a public discourse perspective**

In line with their desirable effects on intergroup relations, the concepts of empathy and perspective taking have also been receiving growing attention from political theorists as important elements in a healthy public discourse. Already Hannah Arendt argued that perspective taking is essential for democratic opinion formation.

By considering a given issue from different viewpoints, by making present to my mind the standpoints of those who are absent. ... The more people's standpoints I have present in my mind while I am pondering a given issue, and the better I can imagine how I would feel and think if I were in their place, the stronger will be my capacity for representative thinking and the more valid my final conclusions, my opinion. (Arendt, 1968, p. 237)

In order to bridge psychology and political theory, Muradova (2020) proposes a theory of perspective taking in deliberation. Similar to personality psychologists, she considers perspective taking as both, dispositional characteristic of a person (trait) and as state, which can be activated in specific situations. Institutional features beneficial for perspective taking in political discussions are a diversity of viewpoints as well as the combination of fact-based argumentation and storytelling. Using interviews and survey data, she finds a positive relationship between dispositional perspective taking and reflective political judgements and suggest perspective taking as mechanism by which deliberation produces more reflective political judgements. However, the experimental setup did not allow explicit conclusions about this triangular relationship because it was unclear whether deliberation actually induced the mediating perspective taking process.

Seemingly conflicting findings on the effects of exposure to opposing political views through digital media (Bail, 2021; Levy, 2021) may, in part, be explained by individual differences in perspective taking abilities. Mutz (2006) found increased political tolerance for those with high levels of perspective taking ability when confronted with cross-cutting

views, while she found decreased tolerance among those with low perspective taking ability (Mutz, 2006).

Monti et al. (2022) find that messages including empathetic expressions of support (broadly defined as ‘giving emotional or practical aid and companionship’) are effective at inducing opinion change in online discussion forum on Reddit in which people explicitly try to change each other’s opinion<sup>3</sup>. While political discussions among citizens are mostly considered as processes of reasoning with a focus on cognition and communication, also internal processes of reflection are fundamentally necessary, for example, to establish tolerance, respect, openness, and reciprocity between individuals (Krause, 2008; Morrell, 2010).

Previous research considered perspective taking and empathy mostly as outcomes of public deliberation (Grönlund et al., 2017). This study focuses on the procedural component, namely the effects of empathy and perspective taking interventions on the deliberative quality of communication with a particular focus on reply toxicity and the length of comments as proxy for argumentation.

### **Empathy and toxicity in prior experimental work**

Morrell (2007) argues that increasing citizens’ empathy should be an important part of democratic education. Fortunately, previous research demonstrates that empathy can be developed through interventions (Batson, 2010; Taylor et al., 2019). However, most experimental evidence on the link between empathy and democracy focused on the effects of empathy interventions on a specific set of democratic outcomes.

For example, in the context of a mega study on interventions to reduce out partisan animosity (Voelkel et al., 2022), an intervention targeting the belief in the utility of cross-partisan empathy by Santos et al. (2022) was among the most effective interventions to reduce partisan animosity. In their study, the authors provided an educational text about the benefits of empathizing with the partisan outgroup. Then, participants were asked to write about how empathy can be useful in competitive contexts and how they could be more empathetic going forward in their own lives. Also when considering other projects within

---

<sup>3</sup><https://www.reddit.com/r/changemyview/>

the multi-intervention study by Voelkel et al. (2022) ‘two variables stood out as statistically mediating the effects of many interventions on partisan animosity: empathy toward, and perceived similarity to, outpartisans’ (p. 8).

Regarding previous attempts to reduce online toxicity, experimental evidence is very limited (Windisch et al., 2022). Overall, more work has been done on more severe forms of online hate speech. For example, Munger (2017) developed a strategy to sanction racist harassment on Twitter using counter speech that was deployed with the help of bots. Hangartner et al. (2021) identified Twitter users who had engaged in xenophobic hate speech and randomly assigned them to three counterspeech strategies: empathy, warning of consequences, and humor, or to a control group. They find that only empathy-based counterspeech messages could consistently increase the retrospective deletion of xenophobic hate speech and significantly reduce the prospective creation of xenophobic hate speech over a 4-week follow-up period.

Existing content moderation strategies can be broadly split into two categories: *ex ante* (preemptive interventions before content is accepted to be posted) and *ex post* (moderation or counter speech after publication), with the latter dominating current moderation practices (Grimmelmann, 2015). Likewise, most existing experimental interventions focus on counter speech (Hangartner et al., 2021; Munger, 2017), reacting after hate speech took over the conversation. However, (Kim et al., 2021) demonstrate experimentally that exposure to toxic language in comments increases the toxicity of subsequent comments—a finding that highlights the value of developing preemptive interventions.

Few researchers explored the potential of *ex ante* interventions but some found that posting community rules increased rule compliance in Reddit forums (Matias, 2019), that the sentiment of posts can be enhanced by CAPTCHAs containing stimuli to prime positive emotions (Seering et al., 2019), that users write more constructive feedback when treated with an empathy-based intervention (Wu & Bailey, 2021) and that prompting the consideration of both pro and con perspectives can contribute to a healthier debate online (Kriplean et al., 2012).

A mid-level strategy are interventions at the time of content creation such as prompts to reconsider offensive messages during their creation. Such interventions require real-time

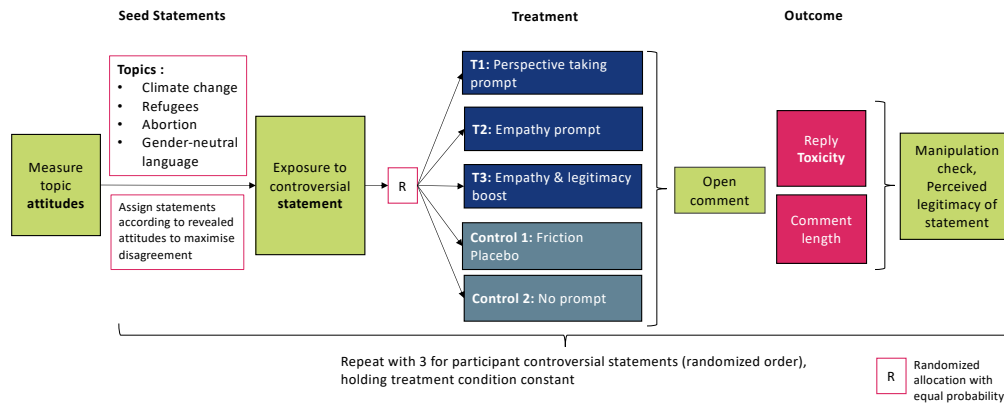
analysis of text and immediate feedback (Katsaros et al., 2022) and are based on the assumption that people are in highly emotional state when writing offensive messages online. While already adopted by several social media companies, such real-time interventions are subject to severe novelty effects. Even though effective for those first exposed to the intervention, the rate of message revisions and cancellations decreases for each additional exposure to the intervention (Katsaros et al., 2022). This highlights the potential of educative boosting interventions that target individuals' motivation and cognitive competencies with more persistent effects across time and contexts (Hertwig & Grüne-Yanoff, 2017).

According to the literature on descriptive social norms (Cialdini & Trost, 1998), an effective preemptive intervention strategy might also have spillover effects (Thøgersen, 1999) to the overall online public discourse. The interventions might not only affect those who directly receive them but also those who observe a more respectful online discourse and adjust the toxicity of their contributions. Taken further, this logic counteracts the exclusionary dynamic of online toxicity that pushes already marginalized individuals out of the discourse. Considering injunctive norms, for example, if community rules are announced in online forums, not only the behavior of established users can change, but also more civil individuals are encouraged to participate actively Matias (2019).

In this project, I develop a first preemptive intervention strategy to reduce toxicity in political online discussions. Connecting to the literature on behavioral science interventions, in this project I juxtapose 'nudging' interventions that focus on simple but rather intransparent changes in the online environment to 'boosting' interventions that target individuals' motivation and competences in a more transparent and arguably more sustainable way (Hertwig & Grüne-Yanoff, 2017). Furthermore, as some forms of friction nudges, interventions to slow down communication, have already been piloted on social media platforms (Binder, 2020), the empathy and perspective taking interventions are tested against a simple friction placebo as second control condition.

Building upon the existing literature on empathy and perspective taking in the online public discourse, I test four main hypotheses (See further non-directional and exploratory hypotheses in the Pre-Analysis Plan (PAP) under <https://osf.io/ym3b2>).

- H1: Subjects assigned to an intervention write comments with lower levels of toxicity

**Figure 1***Study Design.*

than the control group.

- H2: Subjects assigned to an intervention write longer comments than the control group.
- H3: The effect of the intervention is moderated by the attitudinal distance to the statement.
- H4: The comments of people with high self-reported online activity are more toxic than the comments of people who report to be less active online.

## Materials and Methods

### Study design

This study was designed to test interventions to reduce the toxicity of online communication, either through small changes in the online environment, or through the provision of educational information on the potential of empathy for constructive public discourse. The study was conducted in the form of a preregistered online survey experiment including various covariates. For example, interpersonal differences in commenting were explored along demographics, attitudes, and other reported online behaviors.

Participants were confronted with controversial statements (3 out of 6 statements, representing rather extreme positions on 3 different policy relevant topics that bear conflict potential within the German and/or US public discourse: climate change, refugees, gender-neutral language (Germany only) or abortion (US only), see SI Fig. 5), as they could typically be found on social media platforms with the potential to trigger escalations of toxic comments. By deliberately allocating participants to statements they likely disagree with, based on their previously stated attitudes, the aim was to ensure disagreement between the view presented in the statement and the opinion of the participant.

Before participants commented on the statements, one group was prompted to take the perspective of the author of the opposing post and to think about possible reasons the author thinks this way (cognitive perspective taking), one group was prompted to try to create a sense of empathy with the author and reflect on how the author might feel (affective empathy), see Fig. 2. The control group did not receive a prompt while a fourth group received a friction placebo, a prompt to take a deep breath before answering. Prior to reading the provocative posts as a form of educative boosting intervention (Hertwig & Grüne-Yanoff, 2017), a fifth group read a short educative text about the potential of empathy for democratic discourse and the distinction between legitimate opinion and harmful content online. A distraction task (selecting their favorite emoji and the most offensive emoji from a list, amusing results see SI Fig. A21) to separate the boosting intervention from the outcome measure, was presented to all participants. Participants were assigned to one of the treatment or control groups with equal probability, 1/5. Furthermore, it was tested whether the effect of the intervention depends on the attitudinal distance to the statement. To assess the effectiveness of the interventions for those individuals who would, in real life, likely select into engaging with opposing views, individuals' self-reported tendency to engage with conflicting information online was assessed.

A concept that is theoretically independent of political issue attitudes but in practice closely related, is the perceived legitimacy of political statements. The general perception of a statement as, in principle, legitimate political opinion can be understood as precondition for fruitful engagement with it in the public discourse. If one perceives a statement as illegitimate, for example, because it contains hate speech or is perceived as purely offensive,

**Figure 2***Intervention Materials.*

**Control 1: Friction Placebo**

Before you comment – Take a deep breath.

**Treatment 1: Perspective Taking**

Before you comment – Take a moment to imagine yourself in the perspective of the author of this post. Why does the person think this way?

**Treatment 2: Empathy**

Before you comment – Take a moment and try to create a sense of empathy with the author of this comment. How does the person feel?

**How to have better conversations**

We all have disagreements from time to time. Sometimes it is hard to respect the position of others, especially when we fundamentally disagree with them. In such situations, it can be helpful to take a step back and try to understand the perspective of the other. This does not mean that we need to accept the view of the other person, but it helps to continue a respectful conversation.

However, some opinions may be harmful or even illegal to say. It is therefore important to distinguish generally legitimate opinions from harmful, illegitimate claims, to not “feed the trolls” and unintentionally foster hate online.

An easy way to implement this in every-day life is:

1. Avoid trolls, people who intentionally upset others online.
2. Whenever you perceive a statement as generally legitimate, but it still makes you angry or upset, take a step back before you reply. Try to create a sense of empathy and take the perspective of the other, to keep a respectful and constructive conversation going.

**Treatment 3: Boost**

it may be the better communicative strategy to disengage with the sender of the statement and move on. The heuristic of ‘not feeding the trolls’ is mentioned as important cognitive strategy of ‘critical ignoring’, the competence of digital citizens to choose where to invest one’s limited attention capacity online (Kozyreva et al., 2022). Therefore, in addition to issue attitudes that are included as pre-treatment covariates, the perceived legitimacy of the statements was measured post treatment and explored descriptively.

**Measurement**

The primary outcome variable of this study is the toxicity of comments, computationally assessed using the toxicity measure within Google’s Perspective API (see <https://perspectiveapi.com/how-it-works/> for details) which is trained on online comment data to detect rude, disrespectful, or unreasonable comments that are likely to make people leave a discussion.

The secondary outcome variable is the length of comments, serving as proxy for the level of argumentation taking place within one comment. Justifying the personal view in response to a statement takes space, which should be reflected in the length of a comment.

These two outcome measures were selected based on previous work that identified those two specific computational measures to correspond best to two core dimensions of



deliberative quality: respect and argumentation (Oswald, 2022). While sign counts are an open ended measure, toxicity ranges between 0 (non-toxic) to 1 (toxic). Automated toxicity assessments were inspected manually with specific focus to the ends of the distribution, comments with specifically high or low toxicity (see SI for a sample of comments with toxicity scores).

To what extent empathy and perspective taking were induced through the treatments was measured in a manipulation check consisting of the questions ‘For a moment, I could share the feelings of the author.’ (affective empathy) and ‘I can understand the perspective of the author.’ (cognitive perspective taking) with answering options ranging from ‘strongly agree’ to ‘strongly disagree’. Trait empathy was measured using a 7-item battery with items taken from SITES (Konrath et al., 2018) and the EQ (Baron-Cohen & Wheelwright, 2004). An internally consistent scale that serves as a summary covariate was constructed using item analysis and exploratory factor analysis (see SI Figs. A19 and A20 for more details).

More details on the measurement of other variables in the questionnaire and experiment can be found in the supplementary information and the PAP under <https://osf.io/ym3b2>.

### **Sampling and data collection**

According to the selective exposure literature, most people avoid information that challenge their pre-existing views online (Sunstein, 2002). Meanwhile, online toxicity seems to be driven by a loud and active minority of users that does not shy away from conflict (Bor & Petersen, 2022; Bright et al., 2019; Kim et al., 2021). To maximize external validity in the sample, namely, to test interventions on a realistic target population, people who are highly active on social media, data was collected using advertisement on Facebook (Neundorf & Öztürk, 2021a, 2021b).

Ads were delivered through a conversion campaign without targeting. Data was collected between July 28, 2022 and October 10, 2022. Two ad-sets were used: (1) German speaking participants located in Germany and (2) English speaking participants located in the United States. Each ad-set had a daily Budget of 35€. The survey was filled out by

1,114 participants in Germany and 1,040 in the United States which resulted in a total of 5,481 comments written within the survey experiment in response to the opposing political seed statements. The final sample size was determined by the conversion of the Facebook ads campaign, because the research budget was fixed.

All participants viewed the same ad (see SI Fig. 8). To keep the appearance of the ad consistent, comments to the ad that were written by other Facebook users were hidden from participants. Once participants clicked ‘learn more’ on the advertisement, they were directed to the survey, hosted on a Qualtrics page. Before starting the survey, participants were informed about the length, objective and risks of the survey and provided written consent. The study has been evaluated in accordance with the Hertie School’s Research Ethics Review Policy and approved by the Hertie School’s Research Ethics Officer on July 11, 2022.

### **Analytical strategy**

For all experimental analyses, the differences in means are reported as post-hoc tests from OLS regression models. One-way ANOVA and covariate adjusted one-way ANCOVA results are presented as omnibus tests for the overall equality in means. I use HC2 robust standard errors in all OLS models and report p-values from two-sided tests. The tests were calculated using both outcomes: toxicity and comment length.

For covariate-adjusted models, covariates for inclusion were selected using lasso regression with default options in `glmnet`. This selection was run separately for both outcomes. The list of pre-treatment covariates for possible inclusion was the following: Age, gender, education, language, political interest, political orientation, topic of seed statement, valence of seed statement, topic attitudes, attitudinal distance, online activity, comment frequency, tendency to engage with opposing views online, and trait empathy.

Participants assigned to the boosting condition who skipped the boosting text (submitted the page after less than 5 seconds) were considered together with the control group as preregistered. This concerns a total of 5 individuals, reducing the boosting condition to  $N_{boost} = 357$  ( $N_{control} = 411$ ,  $N_{empathy} = 380$ ,  $N_{friction} = 376$ ,  $N_{perspective} = 409$ ). All analyses are reported for the pooled sample with language included as covariate in the

adjusted models. Descriptives about the subsamples are presented in the supplementary information.

To test the moderation hypothesis 3, the attitude distance regarding the topic of the post (ranging between 1 and 3) was considered as moderator. To test the descriptive hypothesis 4, an online activity index was constructed, consisting of the self-reported use of social media, the frequency of writing comments online and the frequency of engaging with opposing views online.

Post-treatment variables such as the disagreement with a statement, perspective taking, perceived empathy and the perceived legitimacy of the statement were considered as a form of manipulation check but were not included in the hypothesis testing models to avoid the induction of post-treatment bias.

Several diagnostic checks, as well as analyses of effect heterogeneity were performed and are reported in the supplementary information. More details on the analytical strategy can be found in the PAP under <https://osf.io/ym3b2>.

### **Deviations from the PAP**

The PAP states that participants with survey taking times below 1/3 of the mean survey taking time are excluded from the sample. However, this is arguably a very strict threshold that leads to the exclusion of a considerable part of the sample (11.6%).

Instead, to maximize transparency, experimental regression coefficients for the adjusted and unadjusted models are presented for various versions of the data: the raw data, the sample using the survey taking time threshold defined in the PAP, as well as the subset of respondents who pass the manipulation check—who disagree with the statement in the first place and/or indicate to have empathized with the author of the statement or to have taken the perspective of the author of the statement if assigned to a treatment group. Results presented in text are based on the raw data.

## Results

### Sample characteristics and data quality

Since a rather novel data collection method was used in this project, Facebook ads without targeting in contrast to established social science panels and platforms, it is worth reporting on a set of sample characteristics.

While there is relative balance regarding the distributions of gender (see Table 1) and education (see SI Fig. A14), the age distribution of the sample is skewed towards older participants ( $M = 66.18$ ,  $SD = 12.42$ ). This aligns with a recent finding that older people are more active on Facebook, especially when it comes to sharing political content (Moretto et al., 2022). Reported political orientation on a left-to-right scale is not skewed to any side but reported political interest is comparably high (see SI Fig. A15). While in Germany, most participants place themselves on the mid-point of the left-to-right scale, political orientation reported by participants in the US appears more polarized towards both ends of the scale.

**Table 1**

*Sample*

Gender	Germany	USA	Total
Male	480	602	1082
Female	603	429	1032
Other	10	3	13
NA	21	6	27
Total	1114	1040	2154

Compared to average responses obtained from frequent survey takers in established social science panels or survey taking platforms, data quality is comparably high. For example, the mean comment length in the open text field in response to the political seed statements was  $M = 152.5$  signs ( $SD = 167.7$ ). As this field contains the main outcome of this survey experiment, the data collection strategy seems to be well suited to the research objective. Furthermore, survey drop-out, NA-response and speeding was minimal (see SI Fig. A17).

There was particular interest to maximize external validity of the experiment by

testing the interventions at the actual target population, individuals who are highly active on social media and who frequently write comments. Fortunately, the participant recruitment strategy, involving Facebook Ads without targeting, resulted in a sample in which most users spend ‘multiple hours daily’ online, use social media ‘multiple times per day’ and write comments online ‘almost daily’ (see Fig A6). Furthermore, most participants report to encounter opposing views ‘often’ online, and ‘sometimes’ engage with them in the form of careful reading (‘often’) or commenting (‘sometimes’). On average, the sample tends to enjoy countering opposing views online ‘a little’ (see SI Figs. A11 and A12).

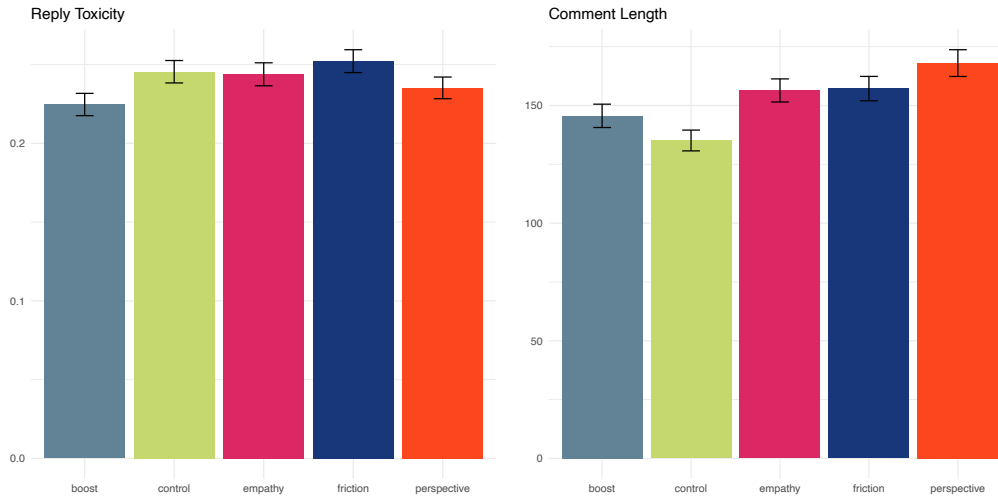
More details on other covariates, such as reported trait empathy (SI Figs. A7), agreement with the political seed statements (SI Fig. A13), perceived legitimacy of the seed statements (SI Fig. A8), as well as information on the manipulation check (SI Figs. A10 and A9) can be found in the supplementary information.

In summary, the sample consists of highly active social media users, many of them older, with high political interest but from both political camps, who opt into the survey without external incentives and show high response compliance.

### Effects of interventions

The main directional hypotheses regard the quality of comments in response to three treatments: a) receiving an emotional empathy prompt, b) a cognitive perspective-taking prompt, or c) a boost. In all three cases, the hypotheses anticipated that treating participants by assigning them to the empathy or perspective-taking condition would cause them to write comments with lower levels of toxicity and longer comments than the control groups.

**Effects on toxicity.** There is no significant main effect of the treatments on toxicity (H1), neither in the simple model,  $F(4, 5480) = 2.17, p = .07, \eta^2 = .002$ , nor in the covariate adjusted model,  $F(4, 5428) = 2.09, p = .08, \eta^2 = .001$ . For differential effects of treatment conditions and a summary of covariates see Fig. 4. While the unadjusted regression model shows a statistically significant negative effect for the boosting condition in the raw data ( $\beta = -0.04, p = .038$ ), with people writing 8% less toxic comments compared to the control group, this effect is not robust to the inclusion of covariates (see Fig. 4).



**Figure 3**

*Raw means of toxicity and comment length across experimental conditions. Note: error bars represent standard errors. Toxicity ranging between 0 (non toxic) and 1 (toxic), comment length represented as signs per comment. For means split between topics (climate change, feminism and migration) see SI Fig. 11, for boxplots see SI Figs. 7 and 9*

In the US sample, treatment effects appear as clear null,  $F(4, 2822) = 1.48$ ,  $p = .21$ ,  $\eta^2 = .002$ , whereas I do find a significant main effect on toxicity in the German subset of the data,  $F(4, 2653) = 2.78$ ,  $p = .025$ ,  $\eta^2 = .004$  (see also SI Fig. A5). Here, participants in the boosting condition write 8% less toxic comments than the control group.

As additional robustness checks, Fig. 4 also presents effects for different variations of manipulation check passes. However, caution should be taken when interpreting these effects, as they can include post-treatment bias. For example, within a reduced sample in which everyone disagreed with the political statement but those who were in a treatment condition agreed to have experienced empathy or perspective taking to some degree, there is a significant main effect on toxicity, both, in the simple model,  $F(4, 2182) = 4.08$ ,  $p = .003$ ,  $\eta^2 = .007$ , and in the covariate adjusted model,  $F(4, 2158) = 4.61$ ,  $p = .001$ ,  $\eta^2 = .008$ . Again, comments were less toxic in the boosting condition compared to the control group.

**Effects on comment length.** Concerning the secondary outcome variable, comment length, as proxy for argumentation (H2), there is a significant main effect of the

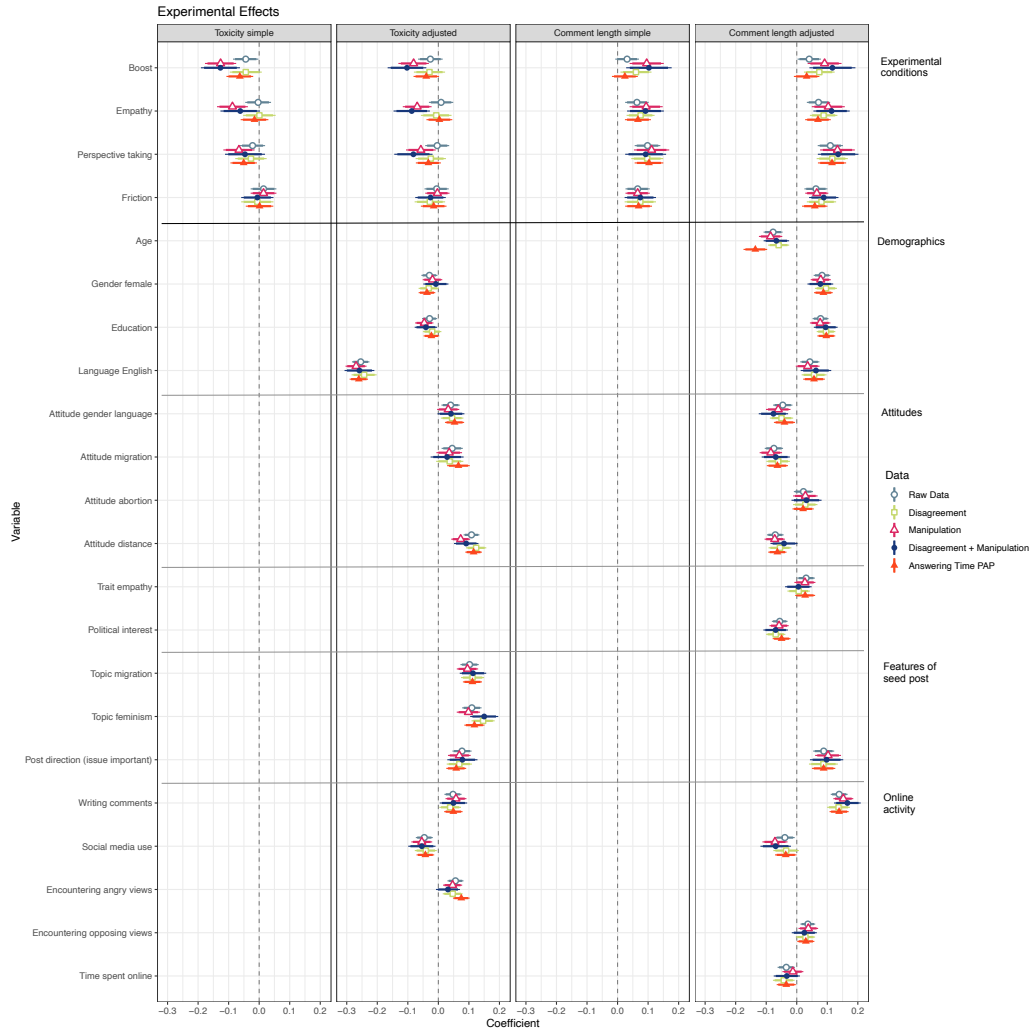
treatment, both, in the unadjusted model,  $F(4, 5480) = 6.29$ ,  $p < .001$ ,  $\eta^2 = .005$ , and in the covariate adjusted model,  $F(4, 5300) = 6.76$ ,  $p < .001$ ,  $\eta^2 = .005$ . Post-hoc tests suggest that all interventions that were delivered as prompts immediately after reading the statement significantly increased comment length. Compared to the control group, comment length in the friction group increased by 22 signs (95% CI [2.7; 41.5]), comments in the empathy group were on average 21 signs longer (95% CI [1.8; 40.7]), and comment length in the perspective taking group increased by 33 signs (95% CI [13.7; 52.1]). For more details on the adjusted and unadjusted models regarding comment length, different subsamples based on manipulation check passes and survey times, as well as a summary of included covariates, see Fig. 4. More information on regression diagnostics as well as effect heterogeneity can be found in the supplementary information.

**Attitudinal distance.** There is no evidence that the effects of the interventions are moderated by the attitudinal distance to the statement (H3),  $F(4, 5475) = 0.77$ ,  $p = .54$ ,  $\eta^2 < .001$  (see also SI Fig. 12) but there is a significant main effect of attitude distance,  $F(1, 5475) = 21$ ,  $p < .001$ ,  $\eta^2 = .004$ . Those with more strongly opposing views against the presented political statement write more toxic comments compared to those with smaller attitudinal distance to the statement. However, it is important to note that attitude distance was not randomly assigned.

**Online activity.** Regarding hypothesis 4, there is descriptive evidence that the comments of people with high self-reported online activity are more toxic than the comments of people who report to be less active online,  $\beta = 0.05$ ,  $p < .001$  (see also SI Fig. 12). For example, when applying a binary split to the online activity scale, people with self-reported online activity levels above the scale midpoint write 4% more toxic comments compared to those with activity levels below the scale midpoint. Comparing the lowest and highest quantile individuals of the online activity distribution, there is a 9% difference in reply toxicity.

### Further descriptive analyses

Toxicity levels were significantly higher in responses to political statements on feminist issues and migration compared to climate change related comments,  $F(2, 5482) =$



**Figure 4**

*Regression Models of Toxicity and Comment Length across Experimental Conditions and Data. Note: continuous regression coefficients were scaled by dividing by two standard deviations as suggested by (Gelman, 2008) to allow direct comparison between binary and continuous variables. Legend represents comparison of different subsets of the data: raw data, ‘disagree’ - subset of replies for which individuals expressed to have disagreed with the seed statements, ‘manipulation’ - subset of replies for which individuals, if assigned to a treatment condition, have indicated to have empathized with or taken the perspective of the author of the seed statement, ‘disagree + manipulation’ - stricter subset requiring both of the above conditions, ‘answering time PAP’ - subset of individuals with survey taking times above 1/3 of the mean survey taking time.*



19.28,  $p < .001$ ,  $\eta^2 = .007$  (see also SI Fig. A4).

Furthermore, toxicity of German comments was assessed to be significantly higher than the toxicity of US comments written in English,  $t(4919) = 21.997$ ,  $p < .001$  (SI Fig. A1). However, it is unclear whether this is a true difference between participants' responses in the US vs. Germany, or a model and/or language specific difference within Google's Perspective API. To explore this question, I machine translated random samples of 1000 German and 1000 English comments to the respective other language using the DeepL API (Zumbach & Bauer, 2021). While the machine translation absorbed toxicity to a great degree in both samples, the substantive toxicity difference between the original German and English comments disappeared after translation and even reversed ( $t_{original}(1915) = 11.8$ ,  $p < .001$ ,  $t_{translated}(1765) = -6.7$ ,  $p < .001$ , see SI Fig. A3). This suggests that the toxicity gap between the original German and English comments is rather language specific to Google's Perspective API but does not reflect a substantive difference which makes it advisable to include language as control variable when examining treatment effects<sup>4</sup>.

Concerning attitudes toward political topics, respondents show clear tendencies about feminist issues: pro-abortion attitudes in the US sample and anti-gender-neutral language attitudes in the German sample. However, the issues migration and climate change that were answered in both countries show greater variance. While Germans rather agree with the statement 'Immigrants should stay out of our country.', participants in the US sample predominantly disagree. Climate change appears to be a more polarizing issue in the US sample, while the attitudes for Germany are mixed on the statement 'Climate change is one of the greatest threats to humanity.' (SI Fig. A13).

The perceived legitimacy of statements varies between issues (see SI Fig. A8). For example, most Germans who are presented with the gender-neutral language devaluing statement<sup>5</sup> perceive it as fully legitimate. However, the opposite is the case for abortion in the US. Both, the issue devaluing statement<sup>6</sup> and the statement framing the issue as

<sup>4</sup>and, potentially, to consider implications of the language specific assessment gap for automated content moderation using Google's Perspective API

<sup>5</sup>"I don't give a shit about gender-neutral language – this whole thing is completely exaggerated! Why should I change the way I speak now when we've gotten along all these years?"

<sup>6</sup>"I don't give a shit about abortion rights – this whole thing is completely exaggerated! If women are careless, it's their problem to deal with the consequences."

priority<sup>7</sup> are predominantly perceived as ‘not at all legitimate’ (see SI Fig. A8). When limiting the sample to cases in which participants perceived the statements to be generally legitimate, there is no significant main effect of the treatment on toxicity,  $F(4, 2262) = 2.18, p = 0.07$ .

### Discussion

Understanding the role of empathy and perspective taking for a healthy public discourse is an ambitious endeavor. This study contributes first evidence on the potential of interventions to preemptively reduce toxicity in online political discussions.

Apparently, it is not easy to change the tone of online comments with the help of small changes in the user interface. I find consistent null effects on toxicity for all interventions that were delivered in the form of prompts or ‘nudges’ (Thaler & Sunstein, 2009). This aligns with recent findings on publication bias in the research on nudges in various domains (Maier et al., 2022).

However, the picture is less clear for the more complex and transparent educative boosting intervention that targets individuals’ understanding of the value of empathy in discussions involving conflicting arguments, and therefore, their motivation to engage in empathy and perspective taking. The fact that boosting effects increase consistently with more strict requirements (e.g. longer survey taking time and passing the manipulation check) is in line with the notion that motivation is inherently necessary for boosting interventions to be effective (Hertwig & Grüne-Yanoff, 2017).

Significant effects in the boosting condition could imply that the intervention changed participants’ understanding of disagreement in conversations and the role of empathy and perspective taking to continue constructive discussions which led them to apply the strategies proposed when replying to the provocative comments. However, boosts may also simply have an advantage over nudges due to greater length and therefore, higher treatment dosage.

Regarding the implied mechanism, it is important to note that the specific boost used in this study mainly targeted users’ motivation to engage in empathy, rather than training

---

<sup>7</sup>"I really don't get how people can be so ignorant about abortion rights! We have to support women with unwanted pregnancies – provide information and free, unbureaucratic abortion care everywhere!"

the cognitive competences necessary to engage in empathy. Developing these competences would be the logical next step in future projects.

In Germany, the boosting condition reduced reply toxicity compared to the control group. This marginal finding coincides with the fact that political orientation appeared less polarized in Germany. However, at the same time, overall toxicity levels assessed post-treatment by Google's Perspective API were higher in Germany than the US. However, this country specific toxicity gap appears to be rather language model specific to Google's Perspective API rather than reflecting a substantive toxicity difference between the German and US comments.

Generally, toxicity levels should not be considered in absolute terms because of an artificial survey setting and because previous research demonstrated that exposure to toxic comments fosters further toxicity (Kim et al., 2021). Therefore the toxicity of seed statements was considered (SI Fig. 6) and added as control to the adjusted models. Furthermore, in everyday life, the perception of toxicity depends substantially on individual characteristics (Kenski et al., 2020) and contexts (Sydnor, 2018). The results of this study are based on an operationalization of toxicity that focuses on disrespectful tone and word choice. However, others argue that related concepts, such as intolerance and antidemocratic attitudes can also be expressed without the use of profane language (Rossini, 2019).

The finding that reply toxicity was higher among people with high self-reported online activity closely aligns with research on selection effects into toxic commenting (Bor & Petersen, 2022; Kim et al., 2021) which gives reason for concern about the role of social media discussions for the discourse in democratic publics such as Germany and the United States. Those who select into active engagement appear to spread more toxicity than the majority of usually passive users (Oswald et al., 2022), skewing the public perception of salient political issues and discourse norms (Bor & Petersen, 2022).

Another interesting contribution to the literature on social media design changes is the finding that the intervention purely inducing friction, slowing down the interaction, was ineffective in reducing reply toxicity. However, the 'friction-nudge' had similar effects as the empathy and perspective taking prompts on comment length. In future iterations of the project, the design of the study could be extended, for example, by using ingroup salience

markers to improve the effectiveness of the empathy and perspective taking interventions through mechanisms of social categorization.

### Conclusion

Overall, despite the weak effects of most experimental treatments, especially the prompt-based ‘nudging’ interventions, this study does not allow concluding judgements on the role of empathy for democracy. Survey experiments, even when targeting the population of interest, suffer from various threats to external and internal validity such as an artificial survey settings, artificial materials, observer effects, sequence effects from previous questions, survey fatigue etc.

The results for the boosting condition give reason for hope in democratic education that emphasizes the role of empathy and perspective taking for public discourse. While this study only provided a very short and abstract text, more complex interventions conveying an understanding of empathy and practicing perspective taking in political discussions that involve conflicting arguments—which are essentially the core of democratic deliberation, should be explored by future research. Furthermore, going beyond the targeting of citizens’ motivation to engage in empathy but developing and training their cognitive competences is an interesting next step for follow-up projects.

This study took a first important step in the research landscape on online toxicity as it focused on the population of interest, highly active social media users that frequently write comments online. I hope other researchers are encouraged to explore the potentials of recruitment via social media advertisement instead of relying on frequent survey takers in social science panels, to research the actual roots of the issue. Another novel step this study took is the development of preemptive interventions that aim to reduce toxicity before it becomes visible to other users where it can spiral towards more toxicity in subsequent engagement.

While toxicity is a concept that is hard to universalize, both in terms of measurement approaches and normative implications (Masullo Chen et al., 2019; Mutz, 2016), I regard the development of interventions to reduce toxicity in online discussions as crucial endeavor. Toxicity is not randomly distributed online but silences important voices in the discourse

and predominantly forces already marginalized groups offline.

**Data Availability Statement.** Fully anonymized data and R-code will be made available for future analyses and replication in an open data repository under <https://osf.io/379hz/>.

**Supplementary Information.** Supplementary information, including experimental materials and supporting tables and figures can be found under <https://osf.io/c86a3>.

### References

- Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The “nasty effect:” Online incivility and risk perceptions of emerging technologies [Publisher: Oxford University Press Oxford, UK]. *Journal of computer-mediated communication*, *19*(3), 373–387.
- Arendt, H. (1968). *Between Past and Future*. Penguin Classics.
- Auguie, B., & Antonov, A. (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. Retrieved December 22, 2021, from <https://CRAN.R-project.org/package=gridExtra>
- Bächtiger, A., Dryzek, J. S., Mansbridge, J., & Warren, M. E. (2018). *The Oxford handbook of deliberative democracy*. Oxford University Press.
- Baek, Y. M., Wojcieszak, M., & Delli Carpini, M. X. (2012). Online versus face-to-face deliberation: Who? Why? What? With what effects? [Publisher: Sage Publications Sage UK: London, England]. *New media & society*, *14*(3), 363–383.
- Bail, C. A. (2021). *Breaking the Social Media Prism*. Princeton University Press.
- Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An Investigation of Adults with Asperger Syndrome or High Functioning Autism, and Normal Sex Differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. <https://doi.org/10.1023/B:JADD.0000022607.19833.00>
- Batson, C. D. (2010). Empathy-induced altruistic motivation.
- Binder, M. (2020). Twitter tests asking if you actually want to read an article before retweeting it [Section: Tech]. Retrieved November 22, 2022, from <https://mashable.com/article/twitter-asks-if-you-want-to-read-before-retweet>

- Bor, A., & Petersen, M. B. (2022). The Psychology of Online Political Hostility: A Comprehensive, Cross-National Test of the Mismatch Hypothesis [Publisher: Cambridge University Press]. *American Political Science Review*, *116*(1), 1–18. <https://doi.org/10.1017/S0003055421000885>
- Borah, P. (2014). Does it matter where you read the news story? Interaction of incivility and news frames in the political blogosphere [Publisher: Sage Publications Sage CA: Los Angeles, CA]. *Communication Research*, *41*(6), 809–827.
- Bright, J., Bermudez, S., Pilet, J.-B., & Soubiran, T. (2019). Power users in online democracy: Their origins and impact. *Information, communication & society*, 1–16.
- Chambers, S. (2003). Deliberative democratic theory [Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA]. *Annual review of political science*, *6*(1), 307–326.
- Chen, G. M. (2017). *Online Incivility and Public Debate*. Springer. Retrieved November 22, 2022, from <https://link.springer.com/book/10.1007/978-3-319-56273-5>
- Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. [Publisher: McGraw-Hill].
- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658–679.
- Dahl, D. B. (2016). *Xtable: Export Tables to LaTeX or HTML*.
- Dryzek, J. S. (2002). *Deliberative democracy and beyond: Liberals, critics, contestations*. Oxford University Press on Demand.
- Epley, N., & Caruso, E. M. (2009). Perspective taking: Misstepping into others' shoes. *Handbook of imagination and mental simulation* (pp. 295–309). Psychology Press.
- Firke, S. (2018). *Janitor: Simple Tools for Examining and Cleaning Dirty Data*.
- Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in medicine*, *27*(15), 2865–2873.
- Google. (2022). Perspective API. Retrieved December 22, 2022, from <https://perspectiveapi.com/>

- Grimmelmann, J. (2015). The virtues of moderation [Publisher: HeinOnline]. *Yale JL & Tech.*, 17, 42.
- Grönlund, K., Herne, K., & Setälä, M. (2017). Empathy in a citizen deliberation experiment. *Scandinavian Political Studies*, 40(4), 457–480.
- Hangartner, D., Gennaro, G., Alasiri, S., Bahrich, N., Bornhoft, A., Boucher, J., Demirci, B. B., Derksen, L., Hall, A., Jochum, M., Munoz, M. M., Richter, M., Vogel, F., Wittwer, S., Wüthrich, F., Gilardi, F., & Donnay, K. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences*, 118(50). <https://doi.org/10.1073/pnas.2116310118>
- Hartman, R., Blakey, W., Womick, J., Bail, C., Finkel, E. J., Han, H., Sarrouf, J., Schroeder, J., Sheeran, P., Van Bavel, J. J., Willer, R., & Gray, K. (2022). Interventions to reduce partisan animosity [Number: 9 Publisher: Nature Publishing Group]. *Nature Human Behaviour*, 6(9), 1194–1205. <https://doi.org/10.1038/s41562-022-01442-3>
- Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, 12(6), 973–986.
- Huling, J. (2019). Jcolors: Colors Palettes for R and 'ggplot2', Additional Themes for 'ggplot2'. Retrieved December 22, 2021, from <https://CRAN.R-project.org/package=jcolors>
- Katsaros, M., Yang, K., & Fratamico, L. (2022). Reconsidering Tweets: Intervening during Tweet Creation Decreases Offensive Content. *Proceedings of the International AAAI Conference on Web and Social Media*, 16, 477–487. <https://doi.org/10.1609/icwsm.v16i1.19308>
- Kenski, K., Coe, K., & Rains, S. A. (2020). Perceptions of Uncivil Discourse Online: An Examination of Types and Predictors [Publisher: SAGE Publications Inc]. *Communication Research*, 47(6), 795–814. <https://doi.org/10.1177/0093650217699933>
- Kim, J. W., Guess, A., Nyhan, B., & Reifler, J. (2021). The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity [Publisher: Oxford University Press]. *Journal of Communication*, 71(6), 922–946.

- Klimecki, O. M. (2019). The Role of Empathy and Compassion in Conflict Resolution [Publisher: SAGE Publications]. *Emotion Review*, 11(4), 310–325. <https://doi.org/10.1177/1754073919838609>
- Konrath, S., Meier, B. P., & Bushman, B. J. (2018). Development and validation of the single item trait empathy scale (SITES). *Journal of Research in Personality*, 73, 111–122. <https://doi.org/10.1016/j.jrp.2017.11.009>
- Kosmidis, S., & Theocharis, Y. (2020). Can Social Media Incivility Induce Enthusiasm?: Evidence from Survey Experiments. *Public Opinion Quarterly*, 84(S1), 284–308. <https://doi.org/10.1093/poq/nfaa014>
- Kozyreva, A., Wineburg, S., Lewandowsky, S., & Hertwig, R. (2022). Critical ignoring as a core competence for digital citizens [Publisher: SAGE Publications Sage CA: Los Angeles, CA]. *Current Directions in Psychological Science*, 09637214221121570.
- Krause, S. R. (2008). *Civil passions*. Princeton University Press.
- Kriplean, T., Morgan, J., Freelon, D., Borning, A., & Bennett, L. (2012). Supporting reflective public thought with considerit. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, 265–274. <https://doi.org/10.1145/2145204.2145249>
- Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American economic review*, 111(3), 831–70.
- Lewandowsky, S., Smillie, L., Garcia, D., Hertwig, R., Weatherall, J., Egidy, S., Robertson, R. E., O'Connor, C., Kozyreva, A., Lorenz-Spreen, P., et al. (2020). Technology and Democracy: Understanding the influence of online technologies on political behaviour and decision-making.
- Lim, D., Condon, P., & DeSteno, D. (2015). Mindfulness and compassion: An examination of mechanism and scalability [Publisher: Public Library of Science San Francisco, CA USA]. *PloS one*, 10(2), e0118221.
- Long, J. A. (2021). Panelr: Regression Models and Utilities for Repeated Measures and Panel Data. Retrieved December 22, 2021, from <https://CRAN.R-project.org/package=panelr>



- Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2022). A systematic review of worldwide causal and correlational evidence on digital media and democracy | Nature Human Behaviour. *Nature Human Behaviour*. <https://doi.org/https://doi.org/10.1038/s41562-022-01460-1>
- Maier, M., Bartoš, F., Stanley, T. D., Shanks, D. R., Harris, A. J. L., & Wagenmakers, E.-J. (2022). No evidence for nudging after adjusting for publication bias [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *119*(31), e2200300119. <https://doi.org/10.1073/pnas.2200300119>
- Masullo Chen, G., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We Should Not Get Rid of Incivility Online [Publisher: SAGE Publications Ltd]. *Social Media + Society*, *5*(3), 2056305119862641. <https://doi.org/10.1177/2056305119862641>
- Matias, J. N. (2019). Preventing harassment and increasing group participation through social norms in 2,190 online science discussions [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *116*(20), 9785–9789. <https://doi.org/10.1073/pnas.1813486116>
- Monti, C., Aiello, L. M., De Francisci Morales, G., & Bonchi, F. (2022). The language of opinion change on social media under the lens of communicative action [Number: 1 Publisher: Nature Publishing Group]. *Scientific Reports*, *12*(1), 1–11. <https://doi.org/10.1038/s41598-022-21720-4>
- Moretto, M., Ortellado, P., Kessler, G., Vommaro, G., Rodriguez-Raga, J. C., Luna, J. P., Heinen, E., Cely, L. F., & Toro, S. (2022). People are more engaged on Facebook as they get older, especially in politics: Evidence from users in 46 countries. *Journal of Quantitative Description: Digital Media*, *2*. <https://doi.org/10.51685/jqd.2022.018>
- Morrell, M. E. (2007). Empathy and democratic education. *Public Affairs Quarterly*, *21*(4), 381–403.
- Morrell, M. E. (2010). *Empathy and democracy: Feeling, thinking, and deliberation*. Penn State Press.
- Munger, K. (2017). Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment. *Political Behavior*, *39*(3), 629–649. <https://doi.org/10.1007/s11109-016-9373-5>

- Muradova, L. (2020). Seeing the other Side? Perspective-taking and reflective political judgments in interpersonal deliberation. *Political Studies*, 0032321720916605.
- Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press.
- Mutz, D. C. (2016). *In-your-face politics: The consequences of uncivil media*. Princeton University Press. Retrieved November 22, 2022, from <https://press.princeton.edu/books/hardcover/9780691165110/in-your-face-politics>
- Neundorf, A., & Öztürk, A. (2021a). *Recruiting Research Participants through Facebook Advertisements: A Handbook* (tech. rep.). OSF Preprints. Retrieved April 13, 2022, from <https://osf.io/87rg3/>
- Neundorf, A., & Öztürk, A. (2021b). Using Facebook Advertisements to Recruit Hard-to-Reach Populations: A Study on the Advertisement Content. Retrieved June 29, 2022, from <https://osf.io/https://osf.io/dax9e>
- Ooms, J. (2018). *Writexl: Export Data Frames to Excel 'xlsx' Format*. <https://CRAN.R-project.org/package=writexl>
- Oswald, L. (2022). Automating the Analysis of Online Deliberation? A comparison of manual and computational measures applied to climate change discussions. <https://doi.org/10.31235/osf.io/qmzwx>
- Oswald, L., Munzert, S., Barbera, P., Guess, A., & Yang, J. (2022). Beyond the tip of the iceberg? Exploring Characteristics of the Online Public with Digital Trace Data. <https://doi.org/10.31235/osf.io/yfmzh>
- R Core Team. (2018). R: A language and environment for statistical computing.
- Robinson, D. (2015). *Broom: Convert Statistical Analysis Objects into Tidy Data Frames*.
- Rossini, P. (2019). Disentangling uncivil and intolerant discourse in online political talk [Num Pages: 238]. In R. G. Boatright, T. J. Shaffer, S. Sobieraj, & D. G. Young (Eds.). Routledge. Retrieved November 22, 2022, from <https://eprints.gla.ac.uk/272713/>
- Santos, L. A., Voelkel, J. G., Willer, R., & Zaki, J. (2022). Belief in the Utility of Cross-Partisan Empathy Reduces Partisan Animosity and Facilitates Political Persuasion

- [Publisher: SAGE Publications Inc]. *Psychological Science*, 33(9), 1557–1573. <https://doi.org/10.1177/09567976221098594>
- Seering, J., Fang, T., Damasco, L., Chen, M. ', Sun, L., & Kaufman, G. (2019). Designing User Interface Elements to Improve the Quality and Civility of Discourse in Online Commenting Behaviors. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3290605.3300836>
- Stietz, J., Jauk, E., Krach, S., & Kanske, P. (2019). Dissociating Empathy From Perspective-Taking: Evidence From Intra- and Inter-Individual Differences Research. *Frontiers in Psychiatry*, 10, 126. <https://doi.org/10.3389/fpsyt.2019.00126>
- Sunstein, C. R. (2002). The law of group polarization. *Journal of political philosophy*, 10(2), 175–195.
- Sydnor, E. (2018). Platforms for incivility: Examining perceptions across different media formats [Publisher: Taylor & Francis]. *Political Communication*, 35(1), 97–116.
- Taylor, S. H., DiFranzo, D., Choi, Y. H., Sannon, S., & Bazarova, N. N. (2019). Accountability and empathy by design: Encouraging bystander intervention to cyberbullying on social media. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–26.
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving Decisions About Health, Wealth, and Happiness* (rev. a. exp. ed. edition). Penguin Books.
- Thøgersen, J. (1999). Spillover processes in the development of a sustainable consumption pattern. *Journal of Economic Psychology*, 20(1), 53–81. [https://econpapers.repec.org/article/eeejoepsy/v\\_3a20\\_3ay\\_3a1999\\_3ai\\_3a1\\_3ap\\_3a53-81.htm](https://econpapers.repec.org/article/eeejoepsy/v_3a20_3ay_3a1999_3ai_3a1_3ap_3a53-81.htm)
- Todd, A. R., & Galinsky, A. D. (2014). Perspective-Taking as a Strategy for Improving Intergroup Relations: Evidence, Mechanisms, and Qualifications [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/spc3.12116>]. *Social and Personality Psychology Compass*, 8(7), 374–387. <https://doi.org/10.1111/spc3.12116>
- Voelkel, J. G., Stagnaro, M. N., Chu, J., & Pink, S. (2022). Megastudy identifying successful interventions to strengthen Americans' democratic attitudes. Retrieved August 19, 2022, from <https://www.strengtheningdemocracychallenge.org/paper>

- Votta, F. (2019). peRerspective: A wrapper for the Perspective API. *R*. <https://github.com/favstats>
- Warnes, G. R., Bolker, B., Gorjanc, G., Grothendieck, G., Korosec, A., Lumley, T., MacQueen, D., Magnusson, A., Rogers, J., et al. (2015). *Gdata: Various R Programming Tools for Data Manipulation*. <http://CRAN.R-project.org/package=gdata>
- Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686.
- Wickham, H., & Bryan, J. (2018). *Readxl: Read Excel Files*. <https://CRAN.R-project.org/package=readxl>
- Wickham, H., & François, R. (2015). Dplyr: A Grammar of Data Manipulation.
- Wickham, H., Hester, J., & François, R. (2017). *Readr: Read Rectangular Text Data*.
- Wickham, H., & Miller, E. (2018). *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*.
- Wikipedia. (2022). Troll (slang) [Page Version ID: 1123122900]. Retrieved November 24, 2022, from [https://en.wikipedia.org/w/index.php?title=Troll\\_\(slang\)&oldid=1123122900](https://en.wikipedia.org/w/index.php?title=Troll_(slang)&oldid=1123122900)
- Windisch, S., Wiedlitzka, S., & Olaghere, A. (2022). Online Interventions for Reducing Hate Speech and Cyberhate: A Systematic Review. *Campbell Systematic Reviews*, 17, 1–17. <https://doi.org/10.1002/cl2.1133>
- Wu, Y. W., & Bailey, B. P. (2021). Better Feedback from Nicer People: Narrative Empathy and Ingroup Framing Improve Feedback Exchange. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3), 236:1–236:20. <https://doi.org/10.1145/3432935>
- Xie, Y. (2018). Knitr: A comprehensive tool for reproducible research in R. *Implementing reproducible research* (pp. 3–31). Chapman; Hall/CRC.
- Zhu, H. (2019). KableExtra: Construct complex table with 'kable' and pipe syntax. *R package version*, 1(0).

Zumbach, D., & Bauer, P. C. (2021). Deeplr: Interface to the 'DeepL' Translation API. Retrieved January 28, 2023, from <https://CRAN.R-project.org/package=deeplr>

## Appendix A

### Regression Diagnostics

The main results of this paper were presented in the form of linear models which imply various diagnostic checks. The distribution of errors was checked for all models. A check for heteroscedasticity suggested the use of robust standard errors which were used throughout for consistency (HC2 robust standard errors). Furthermore, the presence of high leverage observations was checked. For most models, at least some were found when using Cook's distance to perform this check. According to this definition, observations are labeled as high leverage when they have a distance of greater than  $\frac{4}{Num.Obs}$ . Further models without these high leverage observations were estimated (see A1) but the results were substantially the same.

## Appendix B

### List of included publications

1. Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., and Hertwig, R. (2023). A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nature Human Behaviour*, 7(1), Article 1.  
<https://doi.org/10.1038/s41562-022-01460-1>
2. Oswald, L. (2022). More than News! Mapping the Deliberative Potential of Political Online Ecosystems with Digital Trace Data.  
<https://osf.io/preprints/socarxiv/68udg/>
3. Oswald, L., Munzert, S., Barberá, P., Guess, A., and Yang, J. (2022). Beyond the tip of the iceberg? Exploring Characteristics of the Online Public with Digital Trace Data. <https://osf.io/preprints/socarxiv/yfmzh/>
4. Oswald, L. (2022). Automating the Analysis of Online Deliberation? A Comparison of Manual and Computational Measures Applied to Climate Change Discussions. <https://osf.io/preprints/socarxiv/qmzwx/>
5. Oswald, L. (2023). Effects of Preemptive Empathy Interventions on Reply Toxicity among Highly Active Social Media Users.  
<https://osf.io/preprints/socarxiv/evdxy/>