Journal Title
XX(X):1–26

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Polarization of Opinions on COVID-19 Measures: Integrating Twitter and Survey Data

Markus Reiter-Haas[1*], Beate Klösch[2*], Markus Hadler[2], and Elisabeth Lex[1]

Abstract

Polarization of public opinion is a major issue for societies, as high levels can promote adverse effects such as hostility. The present paper focuses on the polarization of opinions regarding COVID-19 prevention measures in survey data and on Twitter in the German-speaking regions of Germany, Austria, and Switzerland. The level of polarization is measured by dispersion and bimodality in the opinions based on the sentiment in Twitter data and the agreement in the survey data. Our paper, however, goes beyond existing research as we consider data from both sources separately and comparatively. For this purpose, we matched individuals' survey responses and tweets for those respondents who shared their Twitter account information. The analyses show that vaccination is more polarizing compared to mask wearing and contact tracing in both sources, that polarization of opinions is more pronounced in the survey data compared to the Twitter data, but also that individuals' opinions about the COVID-19 measures are consistent in both sources. We believe our findings will provide valuable insights for integrating survey data and Twitter data to investigate opinion polarization.

Keywords

Interdisciplinary Research, Opinion Polarization, Surveys, Twitter, Social Media, Integrating Data Sources, COVID-19 Measures

Opinion polarization is a major issue for a society as it leads to adverse effects such as the spread of misinformation (Bessi et al., 2015; Del Vicario et al., 2016). For instance, the opinions on COVID-19

[1]Graz University of Technology
[2]University of Graz
[*]Equal contribution

Corresponding author:
Markus Reiter-Haas, Graz University of Technology, Austria
Email: reiter-haas@tugraz.at

are polarized, as people disagree whether the virus is of natural origin or was created artificially (Reiter-Haas et al., 2020). Opinion polarization is characterized by extreme positions (Stroud, 2010) and can be defined as a state in terms of dispersion and modality of opinions (DiMaggio et al., 1996). Neither is a high dispersion of opinions negative (e.g. personal preferences like opinions on taste or weather) nor is a bimodality in itself harmful (e.g., whether people prefer cats or dogs). Even high polarization on both dispersion and bimodality might not be harmful, e.g., the perception of whether a dress is black and blue or white and gold. Nevertheless, research has shown that polarization in terms of sentiment and emotion, i.e., affective polarization, can lead to hostility in societies, e.g., alongside partisanship (Tucker et al., 2018). As a consequence, research on polarization (e.g., Baldassarri & Bearman, 2007; Borge-Holthoefer et al., 2015; Conover et al., 2011; Fiorina & Abrams, 2008; Garimella & Weber, 2017; J. Jiang et al., 2020) and related issues, such as group polarization (Sunstein, 1999), selective exposure (Knobloch-Westerwick & Meng, 2009), and echo chambers (Garrett, 2009), has been a longstanding research focus.

From a methodological perspective, polarization of public opinion over controversial topics has typically been analyzed via surveys (e.g., Bramson et al., 2017; Hetherington, 2001). In surveys, data about opinions and attitudes is primarily collected from a representative group of respondents to gain insights into the drivers of polarization. In addition, users increasingly exchange opinions and share their attitudes and beliefs via online social media platforms, making them an alternative source for public opinion. Thus, extensive research has been conducted on polarization in various online platforms using user-generated content and digital behavioral data (e.g., Adamic & Glance, 2005; An et al., 2013; Bakshy et al., 2015; Bessi et al., 2014; Conover et al., 2011; Darwish, 2019; Garcia et al., 2012).

While research on polarization at the intersection of surveys and online social media is still scarce, recent work has recognized the potential of linking social media and survey data to measure public opinion (Stier et al., 2020). Nevertheless, it is unclear if similar or different opinion dynamics can be observed in both sources. Moreover, Al Baghal et al. (2021) outline the asymmetry between survey and Twitter data, such as the differences in the quantity and information content, as well as its variability. Generally speaking, Twitter data is more abundant and provides longitudinal insights, whereas it typically lacks socio-demographic information and does not directly probe for the opinions of people, which is in turn provided by survey data. Hence, these two data sources are complementary to each other and taken together provide valuable insights into the opinions of people towards certain topics.

In this paper, we aim to study the relations of opinion polarization between survey responses and social media content with respect to three COVID-19 prevention measures, i.e., vaccination, mask wearing, and contact tracing. Our study analyzes the polarization in the German-speaking DACH region (D-Germany, A-Austria, and CH-Switzerland) at the beginning of August 2020, when the first wave of COVID-19 was over, and Austria, Germany, and Switzerland were almost entirely open. Yet, in this period, the number of COVID-19 cases started to rise again due to holiday traffic, which kept the public engaged in discussions of the COVID-19 prevention measures analyzed in the present work. We focus on COVID-19 as this topic is highly polarized and an emerging societal issue (Allcott et al., 2020; Bruine de Bruin et al., 2020; Dohle et al., 2020; Hart et al., 2020). Its societal relevance is exemplified, for instance, in the rise of dark web marketplaces for medical products, e.g., personal protective equipment and hydroxychloroquine, that were in short supply (Bracci et al., 2021). We deem the study of polarization on COVID-19-related topics as crucial since a high level of polarization can lead to biased reasoning in humans, which in turn may hinder public pandemic mitigation strategies (Van Bavel et al., 2020).

In our approach, we analyze opinion polarization in three sources, (i) Twitter data using an open dataset of tweet IDs (Chen et al., 2020), (ii) survey responses collected from a representative online survey, and (iii) an integrated dataset containing the survey responses and tweets of those survey respondents who shared their Twitter handle with us. Similar to previous work (Alamsyah & Adityawarman, 2017), we use sentiment analysis - also referred to as opinion mining (e.g., by Liu, 2010) - as a proxy to estimate opinions on Twitter. To quantify opinion polarization regarding COVID-19 prevention measures, we compare the extracted sentiment to the expressed agreement in survey responses using the bimodality coefficient (Ellison, 1987), which considers the skewness and kurtosis in the opinion distribution.

Since a direct comparison alone is imprecise due to the different nature of the data, e.g., multiple tweets per account vs. a single response in the survey, we analyze polarization from six different perspectives comprising of the three data sources (i.e., survey, Twitter, and integrated data) each on two levels of granularity (i.e., full and subset). Moreover, to avoid an ecological fallacy (Robinson, 2009), which states that correlations in aggregate data do not necessarily transfer to correlations in data of individuals, we investigate how the individual opinions expressed in the social media data align with the survey answers by using a sub-sample of respondents who agreed to share their Twitter handle. There, human annotators assign an agreement score to each Twitter account based on their tweets to evaluate the congruence of the expressed opinions on Twitter with the agreement in the survey answers.

We see the innovation of our research in bridging two lines of research i.e., *survey research* and *social media research* that discuss the same phenomenon, i.e., polarization, but have traditionally employed different data sources and measures for the task at hand. Specifically, we aim to investigate the congruence of polarization dispersion in our three data sources, i.e., survey, Twitter, and integrated data. Each of the data sources provides a state-of-the-art perspective for their respective line of research. In the Twitter data, we use a commonly referenced sample in the literature on COVID-19 (i.e, Chen et al., 2020); the survey is a representative quota sample of the population; with the integrated data, we consider consenting survey participants that are Twitter users, thus providing an intersection between the two other perspectives.

Our research outlines several similarities in the data sources, e.g., we show that vaccination is a more polarizing measure compared to mask wearing and contact tracing in both Twitter and survey data. Moreover, we observe that the expressed Twitter opinions, in general, agree with the survey answers in the integrated data. Hence, we find that the polarization is congruent between Twitter and survey data in the measured variables (i.e., sentiment for polarization on Twitter and agreement for polarization in the survey), but is more prominently displayed in the survey data. Nevertheless, the shared Twitter accounts predominantly express positive sentiment and agreement on the COVID-19 measures. As such, it might be subject to selection and observation biases.

Our study suggests that the analysis of polarization of opinions using social media content can complement survey research and act as a proxy for public opinions, but does not account for the characteristics of the people sharing their account information and their online engagement. We suspect that people with less extreme opinions are more willing to share their social media data, which we will investigate in future work. Additionally, we highlight the importance of combining social media data with survey data to obtain more comprehensive conclusions.

With this work, we contribute by providing a more holistic view on polarization by considering two complementary data sources and their integration to investigate their similarities in polarization effects. To the best of our knowledge, this is the first work that considers both polarization in surveys and social media, as well as integrates these two complementary data sources. Hence, we advance the state-of-the-art on polarization research by showing the general congruence between the different perspectives, while also paving the way for future research on specific differences between the individual data sources and their effects on the measurement of human behavior.

Related Work

There are many forms of polarization such as social polarization, political polarization, interactional polarization, positional polarization, affective polarization, and opinion polarization. Our work considers opinion polarization, which deals with polarization in terms of spread and formation of opinions (Matakos et al., 2017). Presently, we identify three lines of research that are related to our work: (i) investigating polarization using online data, (ii) studying polarization using survey data, and (iii) integrating survey data with digital behavioral data.

Investigating Polarization in Online Social Media

Related work on polarization in online social media predominantly considers how opinions form, spread, and relate between users. Such network-based approaches have been researched extensively in the past, primarily in terms of user interactions (e.g., using the network topology) and political affiliations. Conover et al. (2011) used hashtags as a proxy for political affiliation to analyze polarization in terms of network topology (i.e., interactional polarization) on Twitter and found high segregation in the retweet network, but less so in the mention network. An et al. (2013) explored the effects of selective exposure on partisan differences on political news consumption on Facebook and found evidence for users predominantly sharing like-minded articles. Bakshy et al. (2015) investigated the media exposure on Facebook considering the friends' network and found that homophily is the most important factor for limiting the mix of content encountered. In this regard, the research of Zhang and Ho (2020) provides evidence that homophily-evoked interactions and fragmentation exists among actors of data journalism on Twitter and the crucial role that organizations hold within the network. Adamic and Glance (2005) studied the linking patterns of political blogs and concluded that both liberals and conservatives primarily link within their communities. In a similar vein, Hagen et al. (2020) investigated the influence of social bots on Twitter, which among other factors amplify messages of fringe actors and smaller communities. However, they show that such amplification when done along ideological lines, can increase fragmentation and polarization in a network. More broadly, the thesis of Garimella (2018) deals with multiple aspects of polarization in networks, e.g., quantifying polarization using a random walk algorithm (Garimella et al., 2018). Moreover, Cota et al. (2019) studied information diffusion on Twitter and found that users are more likely to receive information from others with similar political positions regarding the impeachment of former Brazilian President Dilma Rousseff. However, Esteve Del Valle et al. (2021) analyzed the Twitter mention network of Dutch members of parliament, which only shows a low level of homophily, thus refuting the existence of echo chambers in the analyzed network. Nevertheless, the authors note that the communication patterns in the mention network have dialogical properties. Whereas, the follower and retweet networks, which show support instead, were not analyzed.

In comparison to network-based studies, our research considers how polarization differs between social media and surveys. We achieve this by performing our analysis not only from a macroscopic but also from a microscopic view. This approach has similarities with the information diffusion models from network-based analyses, as it considers whether the views from people expressed in surveys also transfer to social media and vice versa.

Other studies focus more on polarization towards given events, which often contains a temporal dimension as the subject of analysis. Several recent works consider the effects of online conversations on polarization towards given events. Demszky et al. (2019) found that the reactions on Twitter to mass shootings are highly polarized and driven by partisan differences in their messages. Yarchi et al. (2020) conducted an over-time analysis of interactional, positional, and affective polarization on Facebook, Twitter, and WhatsApp on the killing of a Palestinian assailant by an Israeli soldier. They concluded that polarization cannot be seen as a unified phenomenon in social media, as the three platforms showed significant differences. J. Jiang et al. (2020) studied the political polarization of conversations on the COVID-19 pandemic on Twitter using Hashtags and found that partisanship correlates with government prevention measures. In a similar vein, our research focuses on affective polarization in tweets regarding the COVID-19 pandemic. We perform our study in the German-speaking Twitter data on three specific prevention measures, i.e., vaccination, mask wearing, and contact tracing.

Finally, several approaches deal with the differences in content found online and often consider emotions or similar aspects as proxies to quantify opinions. They often consider *affective polarization*, i.e., the emotional reaction of users. Garcia et al. (2012) quantify affective polarization in YouTube videos using likes and dislikes and performed sentiment analysis on comments. Pellert et al. (2020) modeled temporal dynamics of emotions on Facebook using emotional valence, i.e., the positivity of emotions, and arousal, i.e., the energy of the emotion. They find that both valence and arousal relax exponentially towards a baseline level after stimulation, which is relevant to estimate the actual impact of affect. Alternatively, sentiment can be used to determine affective polarization. Alamsyah and Adityawarman (2017) use the sentiment to label nodes in a network as positive, negative, or neutral for structural analysis in an Indonesian case study on Twitter regarding the reclamation of land through filling ocean waters and found that sentiment reliably captures the polarization process as far fewer conversations happen between the pro and counter reclamation nodes. Affect, i.e., emotions and sentiment, can be used to estimate the opinions when considering opinion polarization. Moreover, sentiment analysis is even used interchangeably with the term opinion mining (Liu, 2010).

Similarly, we perform sentiment analysis as a proxy for opinions in the analysis of the polarization on Twitter. Additionally, we quantify the results using statistical measures such as the bimodality coefficient, which allows a comparison of those results with the survey responses.

Unlike many other studies (Adamic & Glance, 2005; Conover et al., 2011; Garcia et al., 2012; J. Jiang et al., 2020; Yarchi et al., 2020), we go beyond considering political polarization since we analyze the polarization in all Twitter users and tweets that express their opinions on the prevention measures regardless of their political affiliation. Moreover, instead of relying only on social media data, we concurrently conducted an online survey in the DACH region to contrast the results. Additionally, we combine a subset of the survey participants with their shared Twitter accounts to directly compare and discuss the differences between survey answers and their views expressed in social media. Thus, our approach of analyzing polarization in both surveys and social media also mitigates concerns of Sloan (2017), who showed that the demographic of Twitter is not representative of the population as a whole, and D. Lee et al. (2015), who showed that there is a discrepancy between the opinions expressed offline and online.

Studying Polarization in Survey Data

The polarization of the public has been considered extensively in the United States, with an emphasis on the divide between the two main parties and individuals that identify with them. Some researchers concluded that the polarization of the political elites contributes to the polarization of the mass, at least to ideological polarization of the identifiers with political parties (Hetherington, 2001). In the political context, a general distinction can be made between *affective political polarization* and *ideological political polarization*. Affective polarization describes the extent to which supporters of one political party oppose other parties, whereas ideological polarization refers to the range of ideological positions and policies of different political parties (Tucker et al., 2018). Further research on political polarization often focuses on the influence of news, online information, and social media on the differentiation of opinions among different constituencies (Abril, 2018; Bail et al., 2018; F. L. Lee, 2016). Besides the influence of information and social media, educational inequality is also a relevant determinant of political polarization. Moreover, when education is taken into account, the impact of income on the differentiation of opinions fades (Bosancianu, 2017).

Other researchers questioned an ongoing and overall ideological or moral polarization of the public and rather perceive short periods of polarization for specific topics and thus support the thesis that attitudes of the public remain rather stable over time (Baldassarri & Bearman, 2007; Evans, 2003; Fiorina & Abrams, 2008). Recent social debates, such as the political discussions and events during the Trump administration from 2017 to 2021 or the ongoing controversies concerning the handling of the COVID-19 pandemic, however, point towards much stronger polarization processes, which also is in line with the observation of a global trend of increasing polarization that entails radical and populist tendencies, especially in political contexts (Deitelhoff et al., 2020). Researchers have used different ways of assessing the polarization of public opinion, which can be a reason for the different findings and conclusions. In an overview, Bramson et al. (2017) identified nine different concepts of assessing polarization. Some concepts are based on the spread and range of answers as well as the distance between extreme positions across an entire population, other concepts are based on the overall shape of a distribution and the dispersion of the data and consider indicators such as mean values, differences, standard deviations, and other related statistical measures. Furthermore, polarization can be understood as little diversity of opinion (*narrow bands of opinion space*) or, in contrast, ideally distinctive groups or diversity of opinion within groups. Other conceptions rather focus on the temporal changes of groups or the group size as such (*size parity*). Our analyses of the public opinion data consider the distribution of answers, mean values, and other statistical measures within the entire sample at a given time.

In addition to the existing focus on political polarization, current research turns towards polarization regarding the COVID-19 debate. Bruine de Bruin et al. (2020) examined US citizens' attitudes towards COVID-19 policies, risk perception, and protective behavior depending on political orientation. They found that Democrats perceived the virus to be more risky in terms of health and economics than Republicans. Likewise, they were more supportive of COVID-19 policies and more likely to fear their early repeal. Allcott et al. (2020) addressed the relationship between political party differences and social distancing during the pandemic in the U.S. population. In addition to the analysis of GPS data, they conducted an online survey, according to which respondents reduced their social contacts by 70% on average (self-reported behavior). This study again showed that Democrats take the pandemic more serious, as they reduced their social contacts more and considered social distancing to be more effective in prevention than Republicans. In addition, Democrats estimate future infection rates higher than Republicans.

In this article, we also examine opinion polarization regarding preventive COVID-19 measures, but without the focus on political orientation. Rather, we seek to present a general overview of the polarization regarding different COVID-19 measures in the DACH region in summer 2020.

Integrating Survey Data with Digital Trace Data

The combination and integration of survey and digital trace data is an emerging field. Recent work by Pasek, McClain, et al. (2020) compares presidential approval with sentiment among population subgroups and found that sentiment is infeasible as a proxy from a microscopic viewpoint while being similar from a macroscopic viewpoint. Thus, their research outlines that a macroscopic comparison is not enough to draw valid conclusions. In our work, we, therefore, also consider the microscopic perspective to mitigate spurious correlations in the data.

In another study, Pasek, Singh, et al. (2020) compared the attention towards various campaign events in the 2016 presidential election between tweets and open-ended survey responses. They found that Twitter and survey data, in general, provide a similar picture on attention but differ in certain details, e.g., in event peak days. Similarly, we compare polarization between Twitter and closed-ended survey responses on a macroscopic level and discuss their similarities and differences. Moreover, we also address a limitation in their work, as we account for more comparable subsets such as Twitter users in the survey data.

Bach et al. (2019) investigate whether voting behavior can be predicted using digital trace data in Germany and find that online behavior is not a good predictor for voting choices, but achieved different results depending on the party, with voting predictions for the right-wing populist and progressive environmentalist party performing slightly better. Their research outlines that even the microscopic data in social media is not enough to accurately predict user choices offline. Hence, we link the microscopic data to ensure that the online behavior of users corresponds to their survey opinions.

Regarding polarization, Joseph et al. (2019) studied the manifestation of polarization between survey and Twitter data by considering the support of tweets from Donald Trump depending on its content, e.g., tweet sentiment. They found that, while Republicans show higher support in general, tweets of Trump that contain positive language, e.g., express positive sentiment, have higher relative support across partisan lines than tweets with negative language. Their findings are also consistent between survey and Twitter data, which is congruent with our findings on opinion polarization in the COVID-19 prevention measures. Unlike their study, we directly relate the levels of polarization in both survey responses and Twitter content using statistical measures. Also, we do not restrict our analyses to political parties.

The research of Al Baghal et al. (2019) discusses the problems of linking individual survey data. They found that the consent rates are very low, especially on web surveys, which may introduce bias in the data. Our research might be subject to low consent rates and possible biases in the integrated data. For this reason, we also compare the data on a macroscopic level that does not require consent and use our small sample with linked data to further strengthen and verify our findings.

Integrating survey data with digital trace data is challenging in several aspects. Stier et al. (2020) describe three key issues that emerge when integrating survey data with digital trace data, i.e., (i) consent when linking individual data, (ii) methodological and ethical issues of the analysis, and (iii) dealing with the multi-dimensionality of such data. All three issues apply to our research. Hence, to address issue (i), we collected individual data only from survey respondents who gave their informed consent.

Table 1. Dataset description of initial survey data collection. We list the collected data separately for each of the three German-speaking countries. For the integrated data set, we use the Twitter handles for which the users provided their consent.

Survey	Austria	Germany	Switzerland
Start	30.07.2020	30.07.2020	30.07.2020
End	07.08.2020	10.08.2020	08.08.2020
Participants	565	1721	274
Twitter Handles	25	77	17

We informed the respondents about the nature of our research and explained that we will analyze their social media posts in case they share their handles with us. That procedure, plus the anonymization of all identifying personal information in any publication, reduces the ethical concerns (ii). The main emphasis of our paper is on the methodological challenges as expressed in (ii) and (iii). We tackle the problem of multi-modality of the data sources by comparing similar statistics across the different data types, i.e., we compare agreement and sentiment using mean, variance, skewness, and kurtosis, as well as derived statistics such as the bimodality coefficient. Yet, we are aware of the different nature of our data and strive to avoid fallacies on inference between survey and social media data compared to just considering aggregate data. Regarding the linking types introduced by Stier et al. (2020), we use both aggregate-level and individual-level ex post linking, i.e., both of which use historical data. To the best of our knowledge, no other study exists that considers the linking of data from both perspectives. On the aggregate-level ex post linking, we combine the data on all three dimensions, i.e., temporally as our macro perspective considers the same time period, topically as our data is on the very narrow subject of COVID-19 prevention measures, and geographically since our data is linked via the German language mainly spoken in the DACH region, where the survey was performed. On the individual-level ex post linking, we use the Twitter API to collect data from the handles provided by the survey respondents.

Data and Methods

We study opinion polarization on COVID-19 prevention measures in German-speaking countries from multiple different perspectives using three data sources. Firstly, we study polarization on Twitter using a multilingual COVID-19 Twitter dataset provided by Chen et al. (2020), whose collection started at the end of January 2020. We considered German tweets posted until August 10TH, 2020. Secondly, we conducted an online survey in the DACH region. In this survey, we collected individual opinions on COVID-19 prevention measures in the form of a survey, which also considers the study participants' socio-demographics and their social media behavior. The survey started on July 30TH 2020 and ended at a different end date depending on the location to meet the country-specific requirements for the quota sample, i.e., August 7TH 2020 for Austria, August 8TH for Switzerland, and August 10TH 2020 for Germany. Table 1 contains details about the study sample. Thirdly, we also asked for the study participants' social media accounts and integrated them with their historical tweets about the COVID-19 prevention measures.

We further focus our data sources to increase comparability between the three perspectives while preserving a decent amount of data for each individual perspective. Specifically, all three perspectives consider the same narrow topic, i.e., COVID-19 prevention measures, in the same language, i.e., German.

There is also considerable similarity regarding geographical information (since German is mostly spoken in the DACH region) and temporal information due to the overlap time period of the data sources. We also consider a subset of the Twitter and survey data, which makes them more comparable. For the Twitter data subset, we focus on the tweets with a direct overlap of the temporal dimension. For the survey data subset, we focus on the answers to respondents that use Twitter (according to their answers). Moreover, there is also a direct overlap between the integrated data, as the integrated survey data is a subset of the overall survey data, while certain tweets of the integrated data also appear in the Chen Twitter data.

Our analyses are driven by comparable statistics derived from the agreement expressed in the survey and sentiment extracted from the Twitter data. Whereas, for the integrated data, we first annotate the tweets with agreement ratings.

COVID-19 situation. Given that the responses and Tweets are influenced by the actual state of the pandemic, we now offer a brief overview of the situation during our data collection period. After the first peak in spring 2020, the pandemic situation in the German-speaking countries was rather calm during the summer. However, due to holiday traffic, infection rates started to rise again and containing measures were discussed anew. At the time of the survey, the stringency index (Ritchie et al., 2020), which records the strictness of active COVID-19 policies (from 0 to 100, 100 = strictest), was between 55.09 and 56.94 in Germany, between 39.35 and 43.06 in Switzerland, and stable at 37.96 in Austria. In all three countries, face masks were required in some public spaces over the whole period, comprehensive contact tracing of all cases was conducted and vaccination was not yet available. Furthermore, there were no stay-at-home requirements during this time span in all three countries, workplace closures and public event cancellations were required for some regions in Germany and Switzerland, in Austria it was recommended (Ritchie et al., 2020).

Twitter Data

We analyze the *Twitter data* using the publicly available dataset from Chen et al. (2020). This dataset contains the 1% sample of tweet IDs from the Twitter streaming API[1] on a predefined set of COVID-19 related accounts and keywords, e.g., *COVID-19* and *Coronavirus*. Using these tweet IDs, we gathered tweets in the period of the survey, i.e., we use the maximum length from January 29TH 2020 to August 10TH 2020, as shown in Table 1, retroactively, which is called hydration. As suggested by the authors, we hydrate the tweets using twarc[2], which uses Twitter's lookup API. As a consequence of the hydration, some tweets might no longer be available, i.e., the tweets were deleted[3]. We filtered the tweets to only include German tweets, resulting in $3,336,562$ tweets for our analyses. Additionally, we also focused on the subset of tweets within the same time period of the survey data, i.e., July 30TH 2020 to August 10TH 2020, which resulted in $547,579$ tweets. When referring to this subset, we explicitly state it, otherwise, we refer to the full Twitter Data.

We further filter the tweets according to three predefined word stems that resemble the three prevention measures to be considered. The rationale for using stems as identification of tweets is due to its simplicity, and thus interpretability while capturing virtually all target tweets[4]. Specifically, we use *impf* for vaccination, *mask* for mask wearing, and *trac* for contact tracing. These three stems capture virtually all tweets related to these measures. The stem *impf* captures the German noun *Impfung* and verb *impfen* for vaccination, as well as other words related to it such as the vaccine itself (i.e., *Impfstoff* in German). The stem *mask* captures the German noun *Maske* and related words such as mask mandate (i.e., *Maskenpflicht* in German). The stem *trac* captures both contact tracing and contact tracer, which have been Germanized and used by the public for COVID-19. For the full dataset, this results in $63,676$ for *impf*, $136,198$ for *mask*, and $13,151$ for *trac* tweets respectively. Considering the subset number of tweets gets reduced to $12,260$ for *impf*, $31,856$ for *mask*, and $1,385$ for *trac*.

We conduct the sentiment analysis using the TextBlob library with the German language extension[5], which includes a sentiment polarity lexicon that we use for sentiment extraction. After extracting the sentiment, we remove tweets that express no sentiment to exclude purely objective statements, e.g., in scientific discussions or very short statements in tweets, which would otherwise dominate the resulting distribution. The final full dataset for comparison consists of $25,769$ tweets expressing sentiment for vaccination, $60,218$ for mask-wearing, and $4,819$ for contact tracing. For the subset, the numbers of tweets with sentiment are $5,420$ for vaccination, $15,425$ for mask-wearing, and 634 for contact tweets. The extracted sentiments are on a numerical scale from -1 for negative to $+1$ for positive sentiment.

Please note that with this procedure, we exclude 57.37% of the tweets since they do not contain any words contained in the sentiment polarity lexicon. Here, potentially valuable tweets might be excluded, e.g., from users, who choose to write their tweets using words not associated with sentiment. Furthermore, TextBlob only uses simple rules in combination with the lexicon, e.g., to detect negations. Thus, more nuanced types of statements, such as sarcasm, are unlikely to be detected and might be associated with the wrong polarity. Finally, the method only considers the text itself, but not contextual features such as conversation threads and user attributes. As a result, the sentiment analysis, while being well-established, can only act as a proxy since true opinions are unavailable in Twitter data.

Survey Data

The *survey data* was collected through an online survey in July and August 2020 in Germany, Austria, and Switzerland as a representative quota sample and comprises 2560 respondents. The targeted quotas were based on the official distribution of gender, age, and federal state/canton in the population of the three DACH countries. As these quotas were met, it can be assumed that the sample is representative of the overall population in these countries regarding those aspects. However, only people with Internet access were considered for the sample, as the survey was conducted online. The questionnaire consists of opinions on polarizing topics (including the COVID-19 prevention measures of vaccination, mask wearing, and contact tracing), social media use (including private Twitter handles), and socio-demographics[6].

The items concerning attitudes towards polarizing topics were taken from the questionnaire of the project ”The measurement of CO2 relevant environmental behaviors and other environmental attitudes through surveys”, funded by the Austrian National Bank (OeNB) and carried out by the Institute of Sociology (University of Graz) in 2019, and adapted to the desired requirements, i.e., COVID-19. The items regarding social media use were taken from the questionnaire of the project ”Future of Life”, conducted by the Institute of Sociology (University of Graz) in 2018/2019.

In our sample, 67% of the respondents are from Germany, another 22% live in Austria, and the remaining 11% are from Switzerland, whereas Austrians are oversampled and the Swiss sample is limited to the German-speaking area. Gender is equally distributed, and the average age is 44 years. The sample shows a high level of education as it contains an above-average number of respondents with a university degree, with almost 30% (the rate of people with university degrees varies in the DACH region between 16% and 21% (Bundesamt für Statistik Schweiz, 2021; Statistik Austria, 2018; Statistisches Bundesamt Deutschland Destatis, 2020).

We analyze polarization in terms of agreement with the COVID-19 prevention measure on an ordinal scale from 1 for strong disagreement to 5 for strong agreement.

In the survey, we also asked participants about their Twitter use and consent to use their private data for our analyses. First, we assured them of the confidential treatment of their Twitter data, and asked them for their consent to link this data to the survey data as well. Respondents first had to give their consent to provide us with their personal Twitter username and access to their data before being asked about their actual Twitter handle in a follow-up question (where we provided an example, i.e., @jane.doe). Of the 2560 respondents in our population survey, 705 people (27.5%) use Twitter between ”several times a day” and ”less than once a week”. In this respect, our data reflect the findings of social media use statistics that Twitter is far less widespread in German-speaking countries compared to other social media platforms, such as Facebook or Instagram (Newman et al., 2021). As in the overall sample, 67% of Twitter users are from Germany, around 21% are from Austria, and 13% live in Switzerland. According to gender, more men (60%) than women use Twitter in our sample, and the average age is slightly lower at around 41 years. The rate of individuals with a university degree is even higher among Twitter users, with almost 38%, compared to the overall sample. In addition to the full sample, we consider the polarization of those 705 respondents separately to have a more comparable subset of survey users for the Twitter platform. Again, we refer to this subset explicitly.

Integrated Data

We use the survey also to generate our dataset of *integrated data*. 119 respondents (29.5%) granted us access to their public Twitter information. At this point, several challenges in linking the two data sources become visible, such as the low number of Twitter users in German-speaking countries or the reluctance to share one’s private social media account. Furthermore, some respondents have provided a false name or a protected account, therefore we can only match a total of 79 survey respondents and Twitter accounts. The distribution by country in this integrated dataset is almost identical to the overall survey sample. The gender ratio is 67% male; the average age (39 years) and the educational qualification of this integrated data (31% university degree) is similar to the overall survey sample.

Comparing our integrated data to all Twitter users in our population survey indicates that our sample is similar in terms of residency (DACH) and educational qualifications. Men and younger respondents, however, are slightly overrepresented. The sample thus is useful to study the similarities and differences concerning their survey statements on COVID-19 prevention measures, but not to draw inferences to the entire Twitter platform.

Using this sample of Twitter accounts, we collected tweets that referred to the COVID-19 pandemic by using the Twitter timeline API for manual annotation. This collection resulted in 221 tweets for 20 accounts - referred to as subset - with original, i.e., non-retweet, tweets in German that contain the term *Corona* or *Covid*. In this step, the sample of Twitter accounts was further reduced, as only 20 of the 79 people who granted us access to their Twitter handles posted about COVID-19 in their tweets. Out of these 221 tweets, 28 are also found in the subset of Twitter data of the survey time period. We combine the tweets with the survey answers to perform analyses from an individual perspective by integrating the Twitter accounts with the survey respondents. We acknowledge that the amount of data is small, which is why we analyze this dataset from a qualitative social science perspective as well. Thus, these individual cases can be used to provide a basis to describe and understand the relationship between the opinions directly addressed to us researchers in the survey and the public opinions posted on Twitter.

Quantifying Polarization

For all three datasets, i.e., Twitter, survey, and integrated data, we first analyze polarization separately. In our analyses, we use the variance to gauge the dispersion and the kurtosis to estimate the modality. A higher variance and a lower kurtosis (especially a negative one) suggest a high level of polarization. Moreover, we measure the bimodality coefficient for a finite sample (SAS, 2012), which indicates bimodality on a scale between 0 and 1 with greater numbers favoring bimodality. It is given by Equation 1, where γ represents the skewness, κ represents the kurtosis, and n represents the sample size. The sample size is used as a normalization factor becomes negligible as the sample size grows large enough, i.e., converges to 1.

$$\beta = \frac{\gamma^2 + 1}{\kappa + 3\frac{(n-1)^2}{(n-2)(n-3)}} \tag{1}$$

The bimodality coefficient has some caveats regarding its use for identifying true bimodal distributions (Pfister et al., 2013). However, it captures the basic intuition for quantifying polarization, i.e., both high skewness and low kurtosis are associated with a higher amount of polarization. Consequently, and in line with intuition, it also assigns a high value in the case of an unimodal but highly skewed distribution.

Evaluating Congruence

In the integrated data, we first investigate the tweet content to get a better understanding of the specific topics that Twitter users are discussing. Following qualitative content analysis according to Mayring (2015), we inductively categorize a subset of COVID-19 related tweets of our integrated data survey users. This way, we want to identify the specific topics that survey users talk about on Twitter when using keywords regarding COVID-19 or hashtags such as *#COVID-19* and *#Corona*.

To evaluate the congruence of the survey and Twitter data, we manually annotated the subset of 20 users with a total of 221 tweets on the COVID-19 prevention measures by two annotators and compared these with the survey data. For the annotations, we chose the same labels as in the survey, i.e., an ordinal rating scale of agreement. We calculate the binary inter-annotator agreement, which only considers perfect matches, between the survey answers and the Twitter annotations. Evaluating the congruence is an important aspect for ensuring the comparability between the survey data and the Twitter data.

Inductive Category Formation. To analyze whether the provided content fits the case for the congruence evaluation, we perform a qualitative content analysis to inductively categorize the content. This approach provides insights into the topics discussed by the integrated users.

The content analysis includes 221 tweets from 20 survey users and discovers a huge variety of categories. Political topics (both local and global politics, over 70 times in total) were most frequently addressed in connection with COVID-19. Here, politicians' handling of the pandemic was frequently discussed and criticized. There was also frequent debate about how dangerous COVID-19 was (almost 50 times). Comparisons were often made with influenza, or personal experiences with COVID-19 were reported. Furthermore, different prevention measures were mentioned about 25 times, and individual problems, as well as societal challenges due to the pandemic, were reported (about 20 times). In addition, private and professional changes in everyday life were reported a few times (more than 10 times). Financial support from the government and how relief funds should be distributed was mentioned

Table 2. Descriptive statistics of the COVID-19 prevention measures, i.e., Vaccination (*Vacc.*), *Mask* Wearing, and Contact Tracing (*CT*), of the three different perspectives, i.e., Twitter, Survey, and Integrated Data. Survey and Twitter results are reported on two levels of granularity, i.e., full and a more comparable subset. The Twitter subset has a direct temporal overlap with the survey; the survey subset focuses on Twitter users; the integrated subset considers the users that post about COVID-19. Note that Twitter results report sentiment, whereas, Survey and Integrated results report the agreement.

Dataset		Statistics	Mean μ	Std σ	Variance σ^2	Median $Q2$	Skew γ	Kurtosis κ	BC β	Sample n
Twitter	all	Vacc.	0.18	0.57	0.32	0.25	-0.29	-0.81	0.49	25,769
		Mask	0.05	0.50	0.25	0.17	-0.29	-0.56	0.44	60,218
		CT	0.15	0.51	0.26	0.23	-0.39	-0.40	0.44	4,819
	subset	Vacc.	0.18	0.60	0.36	0.28	-0.27	-1.02	0.54	5,420
		Mask	0.02	0.49	0.24	-0.05	-0.06	-0.40	0.39	15,425
		CT	0.20	0.46	0.21	0.24	-0.20	-0.39	0.40	634
Survey	all	Vacc.	3.19	1.52	2.31	4	-0.25	-1.42	0.67	2497
		Mask	2.99	1.51	2.27	3	0.05	-1.47	0.65	2523
		CT	3.10	1.39	1.94	3	-0.22	-1.23	0.59	2502
	subset	Vacc.	3.24	1.45	2.11	4	-0.29	-1.30	0.63	690
		Mask	3.09	1.47	2.15	3	-0.05	-1.41	0.63	699
		CT	3.20	1.36	1.84	3	-0.29	-1.12	0.57	691
Integrated	all	Vacc.	3.24	1.37	1.88	4	-0.33	-1.14	0.56	78
		Mask	3.38	1.33	1.78	4	-0.40	-1.04	0.56	79
		CT	3.56	1.26	1.58	4	-0.85	-0.18	0.59	79
	subset	Vacc.	3.53	1.26	1.60	4	-0.44	-0.94	0.45	19
		Mask	3.60	1.19	1.41	4	-0.58	-0.44	0.43	20
		CT	3.75	1.07	1.15	4	-1.74	3.21	0.60	20

in similar frequency. Tweets about scientific research results and data were shared around 10 times, and tweets about fake news and conspiracy theories similarly often. Topics that were less frequently mentioned were polarizing role attributions (e.g., COVID-19 deniers), Corona apps, demonstrations, maintaining occupations, future scenarios, or toilet paper.

Alongside such content-related topics, jokes (sarcasm, irony) about the current situation as well as emotions were frequently found in the COVID-19-related tweets (around 30 times). Here, mainly negative emotions such as annoyance, disappointment, or nervous breakdowns were reported. However, there were also positive emotions mentioned, such as good wishes or hope.

Apart from this, there are some tweets whose content is not directly related to the pandemic (climate crisis, soccer, racism, nature, advertising) or which do not concern German-speaking countries (e.g., the U.S. election). It should also be noted that altogether, only a few survey users (20 out of 79) have posted tweets related to COVID-19. Furthermore, within these 20 individuals, some posted only one or two tweets within the surveyed period while others shared over 40 tweets, which leads to distortions in the frequencies of the topics mentioned.

Overall, we conclude that the majority of data is suitable for the task of a congruence analysis with the survey data, as the tweets mostly reflect individual opinions on the pandemic and the related measures, which enables a content-based comparison with the opinions shared in the survey.

Results

We present the results of our polarization analyses on the Twitter, survey, and integrated data separately before comparing the results. Afterward, we consider the integrated data to determine the congruence between the survey and Twitter data. We summarized the statistics in Table 2. The analyses of the results are predominantly performed in terms of the bimodality coefficient (BC) denoted as β, which is an indicator of polarization.

Polarization in Twitter Data

We analyzed polarization regarding the prevention measures in the COVID-19 German dataset and found that all three measures are polarized as shown in Figure 1. We observe that vaccination has the highest dispersion, i.e., a variance of 0.32 compared to 0.25 and 0.26, which is already an indicator for polarization. Investigating the kurtosis further strengthens this observation, which is far lower than the other two measures, i.e., -0.81 compared to -0.56 and -0.4. Considering the skewness, we observe similar results, but vaccination has the highest mean (0.18) and median (0.25). This shows that the approval of the measures is higher than the rejection. Moreover, all three prevention measures are leaning more towards the positive, i.e., approval side with the mean and median being positive. Computing the bimodality coefficient reaffirms the observation that vaccination is the most polarizing with $\beta = 0.49$.

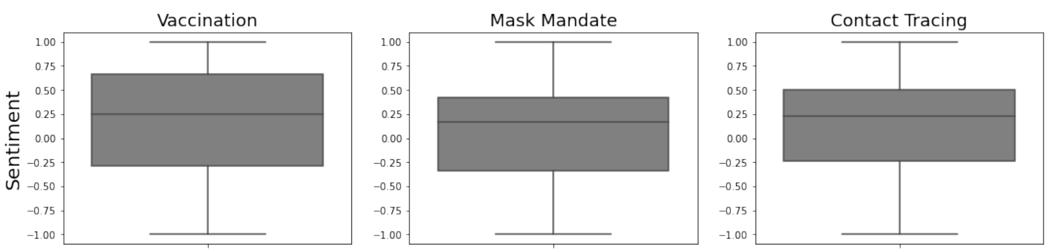

Figure 1. Polarization in Twitter data (all: $n = 90,806$ tweets) in terms of sentiment in the three prevention measures, i.e., vaccination, mask wearing, and contact tracing. The sentiments are measured per tweet on a range from -1 for the maximum negative sentiment to $+1$ for the maximum positive sentiment. Tweets with neutral sentiment are excluded. Vaccination shows high variance which indicates a high level of polarization, but also the highest median suggesting a more positive leaning towards the measure.

The results are very similar for the temporal subset of Twitter data as it has similar medians and dispersion of the data (due to its marginal differences to Figure 1 we omitted showing the boxplot). However, we observe a noticeable change in the bimodality coefficient. This results in an increase for the bimodality coefficient in vaccination with $\beta = 0.54$ and a decrease for the other two prevention measures with a β of 0.39 and 0.4. Overall, we conclude that the prevention measures are polarizing in terms of sentiment, and find that there are differences in the opinions depending on the prevention measures, as vaccination is substantially more polarizing compared to the other two prevention measures.

Polarization in Survey Data

The polarization of public opinion is particularly evident with regard to socio-political measures addressing the COVID-19 pandemic. We investigated the agreement to the introduction of compulsory vaccination, voluntary wearing of face masks, and contact tracing. In the entire sample, both supporters and opponents of all three prevention measures can be found to a similar extent. The highest level of support can be found for the introduction of compulsory vaccination (51% agree absolutely or rather agree), the strongest opposition can be observed against the voluntary wearing of face masks (45% disagree or rather disagree).

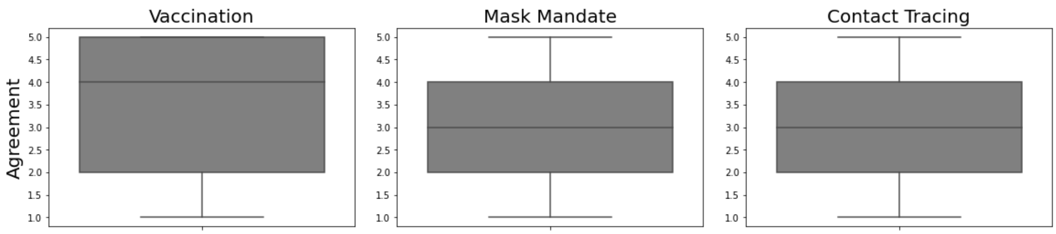

Figure 2. Polarization in Survey data (all: $n = 2560$ respondents) in terms of agreement to the three prevention measures, i.e., vaccination, mask wearing, and contact tracing. The agreement is measured per respondent on a range from 1 for strong disagreement to 5 for strong agreement. Vaccination shows high variance which indicates a high level of polarization, but also the highest median suggesting a more positive leaning towards the measure.

The distribution of the variables regarding the different COVID-19 prevention measures for the overall sample can be seen in Figure 2. Compulsory vaccination receives the highest level of agreement, with a mean of 4. The variables considered here are ordinal, but we nevertheless consider certain statistical indicators of dispersion for the sake of comparability with the Twitter analysis. The first quartile for all three prevention measures lies at 2, which means that 25% of respondents are below this level and do not agree with the prevention measures. The 75% quartile is highest for compulsory vaccination, which again indicates the highest level of agreement with this prevention measure. An additional comparison by country shows that respondents from Germany express the strongest support for all three prevention measures. Meanwhile, respondents from Austria show the highest level of rejection of the prevention measures, especially of contact tracing and compulsory vaccination. It can be noted that in the overall DACH region there tends to be a higher level of support for those three COVID-19 prevention measures than the rejection of the same.

Considering the bimodality coefficient, we observe that all three prevention measures are polarizing with vaccination being the most polarizing by having a bimodality coefficient of 0.67 compared to 0.65 for mask wearing, and 0.59 for contact tracing. Considering the subset of Twitter users shows similar results (the boxplot is almost identical to Figure 2 and thus omitted), but leads to a noticeable drop in the bimodality coefficient. This observation suggests that Twitter users in our sample are less polarized compared to the overall population.

Polarization in Integrated Data

Here, we analyzed the 79 respondents, whose Twitter handles could be successfully matched between the opinions expressed in the survey and the tweets posted online. This group turned out to be more likely in favor of the prevention measures compared to all respondents who use Twitter - especially contact tracing (63% versus 48%) and wearing of face masks (55% versus 44%), whereas compulsory vaccination is seen similar (51% versus 51%).

Figure 3 shows the distributions of agreement on the three COVID-19 prevention measures of the respondents analyzed in this section. The median of all three prevention measures is located at the upper end of the boxes and, in case of contact tracing and face masks, higher than in the overall sample as well as in the subsample of Twitter users.

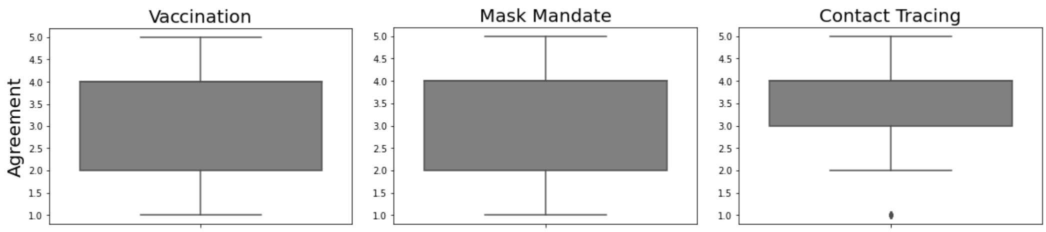

Figure 3. Polarization in the Integrated data (all: $n = 79$ respondents) in terms of agreement among the three prevention measures, i.e., vaccination, mask wearing, and contact tracing. The agreement is measured per respondent on a range from 1 for strong disagreement to 5 for strong agreement. Both, vaccination and mask wearing, show a high variance which indicates a high amount of polarization. All three measures have a median of 4, suggesting a leaning towards approval of the measures.

A decrease of polarization is also reflected in the bimodality coefficient of 0.56 for vaccination and mask wearing, and 0.59 for contact tracing. Interestingly, in this dataset vaccination has a lower bimodality coefficient than contact tracing, whereas vaccination was consistently the highest in terms of the bimodality coefficient for all other datasets.

Considering the subset, we observe an additional drop for vaccination to 0.45 and mask wearing to 0.43. Whereas contact tracing rises to 0.6 as a result of a very negative skewness, that is only partially counteracted by a high kurtosis. This is mainly due to the size of the subsample ($n = 20$), since the smaller the data set, the greater the impact of outliers. Also, agreement on contact tracing is far more unevenly distributed than among the other two preventive measures (65% of respondents expressed an agreement value of 4 on a scale from 1 to 5).

Comparison between Polarization Results

To discuss the COVID-19 measures holistically, we compare the distributions of the provided boxplots, i.e., Figure 1 for Twitter, Figure 2 for the survey, and Figure 3 for the integrated data.

We observe that the sentiments in the Twitter data are less dispersed compared to the survey data and also have a lower bimodality coefficient. Note that Twitter data is collected at the level of tweets and measured in terms of sentiment, whereas survey data is based on a single response per item and respondent and measured in terms of agreement. Nevertheless, we find that the overall characteristics are rather similar in all distributions. The prevention measures of contact tracing and mask wearing are less polarized and do not display a clear tendency towards either side, whereas prevention measures on vaccinations are highly polarized and skewed towards agreement/positive sentiment.

Our observation indicates that the opinions of survey participants directly relate to the opinions of Twitter users. To test this assumption, we compare the tweets in the integrated data, i.e., which were provided by the survey participants, with their respective survey answers.

Congruence of Opinions in Integrated Data. Multiple tweets in our Twitter data can belong to one account, whereas, for the survey data, we have a single answer per respondent. This fact limits the comparability between the two data sources. To mitigate this issue, we also consider the association between the opinions expressed in the survey and through their Twitter accounts within the integrated data. In this regard, we enable a direct comparison of the two different data types, i.e., sentiment and agreement, by manually annotating the tweets.

The binary inter-annotator agreement for the assessment of tweets is $\alpha = 0.7$ and includes missing values, i.e., where the stance towards the prevention measure could not be derived. In comparison, random annotations would only agree $1/6$ of the time (scale 1-5 and missing). The rating scale is similar to the survey scale, based on agreement on a prevention measure of 1 to 5 (1 for strong disagreement and 5 for strong agreement). Both rating scales also allow for missing values, but the meaning differs slightly. In the case of the survey data, missing values mean that participants either have no opinion or that the participants do not want to specify their opinion. In the case of the Twitter data, the missing value means that the opinion could not be derived from the tweets' content.

In summary, a relatively high level of consistency between survey responses and tweet content regarding their opinions toward COVID-19 prevention measures can be observed among the 20 people considered. Only one person shows a discrepancy between their opinion in the survey and their tweets.

Nevertheless, it should be noted that the classification of the analyzed tweets was quite challenging. On the one hand, not all survey users directly addressed COVID-19 prevention measures in their tweets. In this case, the assessment was made based on other related statements or was ambiguous. On the other hand, some survey users did not comment at all on COVID-19 prevention measures on Twitter, which is why no assessment was possible for them.

Discussion

We portray polarization on three COVID-19 prevention measures - vaccination, mask wearing, and contact tracing - from multiple perspectives. Specifically, we use three data sources to investigate whether similar mechanisms exist. Indeed, we find that opinions expressed in our survey and on Twitter show similar polarization across the prevention measures. Generally, vaccination seems to be the most polarizing of the three investigated measures. Moreover, we evaluate congruence in the integrated dataset and find that there is a high congruence between the tweets and survey answers. To improve the comparability, we also consider a subset from both data sources. While the subset is more comparable, this leads to a decrease in the amount of data available for analysis. Hence, our approach considers multiple perspectives to provide a holistic view on the topic of COVID-19 prevention measures.

Our multi-perspective view, however, also faces some trade-offs. We detail three of those trade-offs in our study and discuss how the multi-perspective view mitigates those.

Firstly, the Twitter and survey data consists of *different data types* which are distinct in specific ways. In the Twitter data, we measure a collection of tweets from user accounts. In this scenario, multiple tweets can correspond to the same account. Thus, there is the possibility that a single account posts diverging opinions on Twitter, even within a short time span. In comparison, for the survey data, each respondent expresses a single predefined answer to each question within the given survey. Nevertheless, it is also possible that survey respondents answer differently across multiple surveys and also across multiple items within a survey. Aggregating tweets per account would allow assigning a singular value per account, but would only conceal the underlying problem instead of solving it. For instance, averaging the opinions of diverging tweets would result in a neutral value, even if not a single value expresses a neutral stance on a topic. Considering the value spectrum, in the survey data we use ordinal values, whereas, in the Twitter data we use numerical values. We address this issue with the perspective of the integrated data that combines the two different data types and by mapping tweet content to survey agreement.

Secondly, there is an issue regarding the *representativeness of tweets*, as very active accounts are over-represented. Applying an inverse weighting function (e.g., by simply weighting each tweet with the inverse number of tweets for a given account) could alleviate this bias in the data and achieve balance on an account basis. On the other side, the public perception of the opinions on Twitter is more likely related to tweet visibility, which means that tweets from popular accounts get a lot more attention. For tweet visibility, a weighting function according to tweet engagement, i.e., the number of interactions with a given tweet, might be more suitable. However, tweet engagement is a function of time that tends to increase over time, i.e., the total number of interactions on older tweets is typically higher than for new tweets, while the increase of interactions is higher for newer tweets as they get more attention. We opted for a naive approach and omitted weighting tweets due to a lack of knowledge on which weighting function best captures the relevance of each tweet to its corresponding account. Moreover, treating tweets uniformly lies between the account-level weighting, i.e., treating each account as equally important, and visibility-level weighting, i.e, according to the public perception. As such, it provides a balance between those extreme weightings, while providing a natural way of representing the importance of social media content. Our approach mitigates some of the issues of representativeness, as we consider the polarization at different levels of granularity, including the very fine-grained level of our integrated data. In the integrated data, individual tweets are aggregated, and an overall assessment is derived, thus, alleviating the issue.

Thirdly, we analyze the polarization in the Twitter data *using sentiment exclusively*, but not in terms of positions or emotions. Considering positions would be non-trivial due to a lack of well-defined dimensions such as political ideology. Regarding the measuring of affective polarization using emotions, we performed a prestudy in terms of emotions, which did not lead to noteworthy results. In particular, the results were comparable to sentiment in terms of emotional valence but less distinct. Thus, we focus our analyses on sentiment for conciseness reasons. This single view on the Twitter data becomes less prevalent as we also report the perspective of the agreement in the survey data.

Although we find that polarization is similar between the perspectives, there are still differences between each of the data sources. Comparing the Twitter data with the survey data, we observe that the

Twitter data is less polarized considering the bimodality coefficient. However, we cannot conclude that Twitter as a platform acts in a depolarizing manner. Although we observe that the subsample of survey respondents that use Twitter are less polarized, two other observations indicate that other effects could be the cause for this phenomenon. Firstly, a temporal focus on the Twitter data within the survey time period results in a change of the bimodality coefficient. The subset of Twitter data shows higher polarization for vaccination but decreases for mask wearing and contact tracing. This outlines the importance of considering temporal factors in the analysis. Secondly, we also observe that there is a lower level of polarization in agreement in the integrated data compared to the complete and Twitter subset of the survey data. Still, the polarization is substantially higher compared to the level of polarization in the Twitter data. We attribute this difference to the different kinds of data that are measured respectively. In the Twitter data, the sentiments of tweets are measured, and multiple tweets can belong to the same account, whereas agreement of individuals in the survey data is measured at the particular time of the fieldwork. These observations show that a direct comparison would be infeasible and is the reason we also evaluated the congruence in the integrated data.

Overall, we show that both survey data and social media data have their merits when studying opinion polarization; however, both provide an incomplete picture. Twitter data is more abundant, whereas, survey data provides representativeness. Additionally, considering the integrated data combines the advantage of both perspectives, but comes at the cost of difficulties in obtaining the data. As we found in our experiments, only a limited amount of data can be collected with such an approach. A possible remedy to increase the sample size could be to move the recruitment of survey participants to social media platforms, e.g., similar to the approach described in Pötzschke and Weiß (2021), or to target specific user interests and user demographics.

Considering the perspectives for our topic, i.e., COVID-19 prevention measures in the German-speaking DACH region, we find that there is a congruence of the different perspectives, but with variations in how prononced the observed polarization is. Thus, each individual perspective would result in a similar conclusion, but the polarization is more noticeable in the survey data. Still, our research illustrates the importance of considering multiple perspectives, as there are noticeable differences between the perspectives. Whether our findings also apply to other topics than COVID-19 prevention measures remains a subject for further research, as our study design needed to be restricted to a specific topic to improve comparability amoung data sources.

Alongside these methodological insights, our approach can be of value in supporting policymakers to gauge polarization on controversial topics, such as COVID-19 prevention measures. Here, we observe that compulsory vaccination is a very polarizing prevention measure in the DACH region and needs special consideration when discussed in the public sphere. This observation agrees with previous studies that suggest that vaccinations are indeed polarizing (X. Jiang et al., 2021; Schmidt et al., 2018).

Limitations

While considering multiple perspectives provides a holistic view on polarization effects, we identify three limitations of our work.

Firstly, we focus on *polarization as a state* instead of also considering the definition of polarization as a process by DiMaggio et al. (1996). However, temporal effects could play a major role in the understanding of how a topic gets polarized in the first place. Thus, considering polarization as a state only could greatly influence the interpretation of the results. While temporal information was available for Twitter data, the survey data is available only for the time of fieldwork. Moreover, using the short time span of the experiment would likely not reveal interesting dynamics in the process.

Secondly, there might be *potential biases* in the data, especially in terms of the respondents who share their accounts. While we briefly discussed the differences between survey respondents in general, survey respondents using Twitter, and survey respondents who shared their Twitter accounts, we did not perform an in-depth analysis of the characteristics of the respondents who shared their Twitter accounts. This might introduce biases, i.e., selection and observation biases, into the analysis of tweets. We suspect that certain characteristics favoring account sharing could also explain the less polarizing nature of our Twitter sample. For instance, we presume that users with extreme positions might be reluctant to share their account information. Also, Twitter users are not necessarily users who post on Twitter but might be using the platform passively.

Thirdly, we again emphasize the challenging issues of *comparing survey data with Twitter data*, which are different by their very nature. Their integration lets us combine the advantages of both data types, but results in a small number of users and tweets for analysis. Since in our approach, we perform sentiment analysis on the tweets and measure agreement in the survey data to quantify polarization, our study is subject to the limitations of these techniques, as we discussed in the methods section.

Future Work

As for future work, we plan to reproduce this experiment in a follow-up survey on a larger sample size to further validate our results. To increase the amount of data in the integrated data, we will conduct the recruitment on the social media platforms to acquire more active users alongside the representative sample. Additionally, we aim to repeat the survey multiple times with the same set of users and questions. In these questionnaires, we will ask users to state their reason for sharing or not sharing their accounts, which allows the analysis of biases in the integration of data. Overall, this longitudinal study should provide in-depth insights into the process of how polarization changes over time.

Furthermore, we will also incorporate advanced models for opinion formation and spread in the social media analyses. For instance, we want to investigate how the multiple expressed opinions in tweets relate to the single innate opinion of a social media account user. Using these models, we will try to further improve the understanding of how online content relates to the survey answers.

Acknowledgements

We thank the rectorates of Graz University of Technology and University of Graz for their support of this research. We thank the anonymous reviewers for their helpful comments on this paper.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplemental material

Supplemental material for this article is available online. The archive of the survey information is available at: https://data.aussda.at/dataset.xhtml?persistentId=doi:10.11587/OVHKTR. The repository with the code is available at: https://github.com/socialcomplab/sscr-opinion-polarization.

Notes

1. https://developer.twitter.com/en/docs/tutorials/consuming-streaming-data Note that, while the sample is supposedly random, Pfeffer et al. (2018) showed that it should not be regard as such due limitations in the sampling algorithm.
2. https://github.com/DocNow/twarc
3. This is due to the Twitter policy that researchers are only allowed to share tweet IDs instead of the complete tweets: https://developer.twitter.com/en/developer-terms/agreement-and-policy
4. We also experimented with topic models such as Latent Dirichlet Allocation (LDA), but perceived the interpretation of the resulting topics and their unsatisfactory quality as a hindrance in our analysis.
5. https://textblob-de.readthedocs.io
6. The questionnaire items are provided in the supplemental materials.

References

Abril, E. P. (2018). Subduing attitude polarization?: How partisan news may not affect attitude polarization for online publics. *Politics and the Life Sciences*, *37*(1), 68–77. https://doi.org/10.1017/pls.2017.1

Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 US election: Divided they blog. *Proceedings of the 3rd international workshop on Link discovery*, 36–43. https://doi.org/10.1145/1134271.1134277

Al Baghal, T., Sloan, L., Jessop, C., Williams, M. L., & Burnap, P. (2019). Linking Twitter and survey data: The impact of survey mode and demographics on consent rates across three uk studies. *Social Science Computer Review*, *38*(5), 517–532. https://doi.org/10.1177/0894439319828011

Al Baghal, T., Wenz, A., Sloan, L., & Jessop, C. (2021). Linking Twitter and survey data: Asymmetry in quantity and its impact. *EPJ Data Science*, *10*(1), 32. https://doi.org/10.1140/epjds/s13688-021-00286-7

Alamsyah, A., & Adityawarman, F. (2017). Hybrid sentiment and network analysis of social opinion polarization. *2017 5th International Conference on Information and Communication Technology (ICoIC7)*, 1–6. https://doi.org/10.1109/ICoICT.2017.8074650

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of Public Economics*, *191*, 1–11. https://doi.org/10.1016/j.jpubeco.2020.104254

An, J., Quercia, D., & Crowcroft, J. (2013). Fragmented social media: A look into selective exposure to political news. *Proceedings of the 22nd International Conference on World Wide Web*, 51–52. https://doi.org/10.1145/2487788.2487807

Bach, R. L., Kern, C., Amaya, A., Keusch, F., Kreuter, F., Hecht, J., & Heinemann, J. (2019). Predicting voting behavior using digital trace data. *Social Science Computer Review*. https://doi.org/10.1177/0894439319882896

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. https://doi.org/10.1073/pnas.1804840115

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Baldassarri, D., & Bearman, P. (2007). Dynamics of political polarization. *American Sociological Review*, *72*(5), 784–811. https://doi.org/10.1177/000312240707200507

Bessi, A., Caldarelli, G., Del Vicario, M., Scala, A., & Quattrociocchi, W. (2014). Social determinants of content selection in the age of (mis) information. *International Conference on Social Informatics*, 259–268. https://doi.org/10.1007/978-3-319-13734-6_18

Bessi, A., Petroni, F., Del Vicario, M., Zollo, F., Anagnostopoulos, A., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Viral misinformation: The role of homophily and polarization. *Proceedings of the 24th International Conference on World Wide Web*, 355–356. https://doi.org/10.1145/2740908.2745939

Borge-Holthoefer, J., Magdy, W., Darwish, K., & Weber, I. (2015). Content and network dynamics behind egyptian political polarization on Twitter. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 700–711. https://doi.org/10.1145/2675133.2675163

Bosancianu, C. M. (2017). A growing rift in values? Income and educational inequality and their impact on mass attitude polarization. *Social Science Quarterly*, *98*(5), 1587–1602. https://doi.org/10.1111/ssqu.12371

Bracci, A., Nadini, M., Aliapoulios, M., McCoy, D., Gray, I., Teytelboym, A., Gallo, A., & Baronchelli, A. (2021). Dark web marketplaces and COVID-19: Before the vaccine. *EPJ Data Science*, *10*(1), 6. https://doi.org/10.1140/epjds/s13688-021-00259-w

Bramson, A., Grim, P., Singer, D. J., Berger, W. J., Sack, G., Fisher, S., Flocken, C., & Holman, B. (2017). Understanding polarization: Meanings, measures, and model evaluation. *Philosophy of Science*, *84*(1), 115–159. https://doi.org/10.1086/688938

Bruine de Bruin, W., Saw, H.-W., & Goldman, D. P. (2020). Political polarization in US residents' COVID-19 risk perceptions, policy preferences, and protective behaviors. *Journal of Risk and Uncertainty*, *61*, 177–194. https://doi.org/10.1007/s11166-020-09336-3

Bundesamt für Statistik Schweiz. (2021). Höchste abgeschlossene Ausbildung in der Schweiz. https://www.bfs.admin.ch/bfs/de/home/statistiken/bildung-wissenschaft/bildungsstand.assetdetail.11627129.html

Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. *JMIR Public Health and Surveillance*, *6*(2), e19273. https://doi.org/10.2196/19273

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on Twitter. *Fifth International AAAI Conference on Weblogs and Social Media*.

Cota, W., Ferreira, S. C., Pastor-Satorras, R., & Starnini, M. (2019). Quantifying echo chamber effects in information spreading over political communication networks. *EPJ Data Science*, *8*(1), 35. https://doi.org/10.1140/epjds/s13688-019-0213-9

Darwish, K. (2019). Quantifying polarization on Twitter: The Kavanaugh nomination. *International Conference on Social Informatics*, 188–201.

Deitelhoff, N., Groh-Samberg, O., & Middell, M. (Eds.). (2020). *Gesellschaftlicher Zusammenhalt. Ein interdisziplinärer Dialog.* Campus Verlag.

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

Demszky, D., Garg, N., Voigt, R., Zou, J., Gentzkow, M., Shapiro, J., & Jurafsky, D. (2019). Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. *arXiv preprint arXiv:1904.01596*.

DiMaggio, P., Evans, J., & Bryson, B. (1996). Have American's social attitudes become more polarized? *American journal of Sociology*, *102*(3), 690–755. https://doi.org/10.1086/230995

Dohle, S., Wingen, T., & Schreiber, M. (2020). Acceptance and adoption of protective measures during the COVID-19 pandemic: The role of trust in politics and trust in science. *Social Psychological Bulletin*, *15*(4), 1–23. https://doi.org/10.32872/spb.4315

Ellison, A. M. (1987). Effect of seed dimorphism on the density-dependent dynamics of experimental populations of atriplex triangularis (chenopodiaceae). *American Journal of Botany*, *74*(8), 1280–1288. https://doi.org/10.1002/j.1537-2197.1987.tb08741.x

Esteve Del Valle, M., Broersma, M., & Ponsioen, A. (2021). Political interaction beyond party lines: Communication ties and party polarization in parliamentary Twitter networks. *Social Science Computer Review*. https://doi.org/10.1177/0894439320987569

Evans, J. H. (2003). Have Americans' attitudes become more polarized?—an update*. *Social Science Quarterly*, *84*(1), 71–90. https://doi.org/10.1111/1540-6237.8401005

Fiorina, M. P., & Abrams, S. J. (2008). Political polarization in the American public. *Annual Review of Political Science*, *11*(1), 563–588. https://doi.org/10.1146/annurev.polisci.11.053106.153836

Garcia, D., Mendez, F., Serdült, U., & Schweitzer, F. (2012). Political polarization and popularity in online participatory media: An integrated approach. *Proceedings of the first edition workshop on Politics, elections and data*, 3–10. https://doi.org/10.1145/2389661.2389665

Garimella, K. (2018). *Polarization on social media* (Doctoral dissertation). Aalto University. Aalto University.

Garimella, K., Morales, G. D. F., Gionis, A., & Mathioudakis, M. (2018). Quantifying controversy on social media. *ACM Transactions on Social Computing*, *1*(1), 1–27. https://doi.org/10.1145/3140565

Garimella, K., & Weber, I. (2017). A long-term analysis of polarization on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1).

Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among internet news users. *Journal of Computer-mediated Communication*, *14*(2), 265–285. https://doi.org/10.1111/j.1083-6101.2009.01440.x

Hagen, L., Neely, S., Keller, T. E., Scharf, R., & Vasquez, F. E. (2020). Rise of the machines? Examining the influence of social bots on a political discussion network. *Social Science Computer Review*. https://doi.org/10.1177/0894439320908190

Hart, P. S., Chinn, S., & Soroka, S. (2020). Politicization and polarization in COVID-19 news coverage. *Science Communication*, *42*(5), 679–697.

Hetherington, M. J. (2001). Resurgent mass partisanship: The role of elite polarization. *American Political Science Review*, *95*(3), 619–631. https://doi.org/10.1017/S0003055401003045

Jiang, J., Chen, E., Yan, S., Lerman, K., & Ferrara, E. (2020). Political polarization drives online conversations about COVID-19 in the United States. *Human Behavior and Emerging Technologies*, *2*(3), 200–211. https://doi.org/10.1002/hbe2.202

Jiang, X., Su, M.-H., Hwang, J., Lian, R., Brauer, M., Kim, S., & Shah, D. (2021). Polarization over vaccination: Ideological differences in Twitter expression about COVID-19 vaccine favorability and specific hesitancy concerns. *Social Media+ Society*, *7*(3). https://doi.org/10.1177/20563051211048413

Joseph, K., Swire-Thompson, B., Masuga, H., Baum, M. A., & Lazer, D. (2019). Polarized, together: Comparing partisan support for trump's tweets using survey and platform-based measures. *Proceedings of the International AAAI Conference on Web and Social Media*, *13*, 290–301.

Knobloch-Westerwick, S., & Meng, J. (2009). Looking the other way: Selective exposure to attitude-consistent and counterattitudinal political information. *Communication Research*, *36*(3), 426–448. https://doi.org/10.1177/0093650209333030

Lee, D., Hahn, K. S., Yook, S.-H., & Park, J. (2015). Quantifying discrepancies in opinion spectra from online and offline networks. *PloS one*, *10*(4), e0124722. https://doi.org/10.1371/journal.pone.0124722

Lee, F. L. (2016). Impact of social media on opinion polarization in varying times. *Communication and the Public*, *1*(1), 56–71. https://doi.org/10.1177/2057047315617763

Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing*, *2*(2010), 627–666.

Matakos, A., Terzi, E., & Tsaparas, P. (2017). Measuring and moderating opinion polarization in social networks. *Data Mining and Knowledge Discovery*, *31*(5), 1480–1505. https://doi.org/10.1007/s10618-017-0527-9

Mayring, P. (2015). *Qualitative Inhaltsanalyse* (12th ed.). Beltz Verlagsgruppe.

Newman, N., Fletcher, R., Schulz, A., Andı, S., Robertson, C. T., & Nielsen, R. K. (2021). Reuters institute digital news report 2021.

Pasek, J., McClain, C. A., Newport, F., & Marken, S. (2020). Who's tweeting about the president? What big survey data can tell us about digital traces? *Social Science Computer Review*, *38*(5), 633–650. https://doi.org/10.1177/0894439318822007

Pasek, J., Singh, L. O., Wei, Y., Soroka, S. N., Ladd, J. M., Traugott, M. W., Budak, C., Bode, L., & Newport, F. (2020). Attention to campaign events: Do Twitter and self-report metrics tell the same story? *Big Data Meets Survey Science: A Collection of Innovative Methods*, 193–216. https://doi.org/10.1002/9781118976357.ch6

Pellert, M., Schweighofer, S., & Garcia, D. (2020). The individual dynamics of affective expression on social media. *EPJ Data Science*, *9*(1), 1–14. https://doi.org/10.1140/epjds/s13688-019-0219-3

Pfeffer, J., Mayer, K., & Morstatter, F. (2018). Tampering with Twitter's sample API. *EPJ Data Science*, *7*(1), 50. https://doi.org/10.1140/epjds/s13688-018-0178-0

Pfister, R., Schwarz, K. A., Janczyk, M., Dale, R., & Freeman, J. (2013). Good things peak in pairs: A note on the bimodality coefficient. *Frontiers in Psychology*, *4*, 700. https://doi.org/10.3389/fpsyg.2013.00700

Pötzschke, S., & Weiß, B. (2021). Realizing a global survey of emigrants through Facebook and Instagram. https://doi.org/10.31219/osf.io/y36vr

Reiter-Haas, M., Kloesch, B., Hadler, M., & Lex, E. (2020). Bridging the gap of polarization in public opinion on misinformed topics. *Challenging Misinformation: Exploring Limits and Approaches, workshop co-located with Social Informatics'20*.

Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D., & Roser, M. (2020). Policy responses to the coronavirus pandemic. https://ourworldindata.org/policy-responses-covid

Robinson, W. S. (2009). Ecological correlations and the behavior of individuals. *International journal of epidemiology*, *38*(2), 337–341. https://doi.org/10.1093/ije/dyn357

SAS, S. (2012). Stat 12.1: User's guide cary. *NC: SAS Institute Inc*.

Schmidt, A. L., Zollo, F., Scala, A., Betsch, C., & Quattrociocchi, W. (2018). Polarization of the vaccination debate on Facebook. *Vaccine*, *36*(25), 3606–3612. https://doi.org/10.1016/j.vaccine.2018.05.040

Sloan, L. (2017). Who tweets in the United Kingdom? Profiling the Twitter population using the British social attitudes survey 2015. *Social Media+ Society*, *3*(1). https://doi.org/10.1177/2056305117698981

Statistik Austria. (2018). Bildungsstand der Bevölkerung. https://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bildung/bildungsstand_der_bevoelkerung/index.html

Statistisches Bundesamt Deutschland Destatis. (2020). Bildungsstand. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Bildungsstand/_inhalt.html

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. https://doi.org/10.1177/0894439319843669

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, *60*(3), 556–576. https://doi.org/10.1111/j.1460-2466.2010.01497.x

Sunstein, C. R. (1999). The law of group polarization. *University of Chicago Law School, John M. Olin Law & Economics Working Paper*, (91). https://doi.org/10.2139/ssrn.199668

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. https://doi.org/10.2139/ssrn.3144139

Van Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., et al. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, *4*(5), 460–471. https://doi.org/10.1038/s41562-020-0884-z

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2020). 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. https://doi.org/10.1080/10584609.2020.1785067

Zhang, X., & Ho, J. C. (2020). Exploring the fragmentation of the representation of data-driven journalism in the Twittersphere: A network analytics approach. *Social Science Computer Review*. https://doi.org/10.1177/0894439320905522

Author Biographies

Markus Reiter-Haas is a university assistant and PhD student at the Institute of Interactive Systems and Data Science at Graz University of Technology. His research focuses on applying computational models in social systems to analyze how user behavior leads to emergent phenomena such as polarization. Email: reiter-haas@tugraz.at

Beate Klösch is a PhD student and research associate at the University of Graz, department of Sociology. Her research focuses on quantitative social research, polarization in public opinion and environmental sociology. Email: beate.kloesch@uni-graz.at

Markus Hadler is a professor of Sociology, University of Graz, Austria, and an honorary professor, Department of Sociology, Macquarie University, Australia. His research interests lie in the areas of social inequality, political sociology, and environmental sociology. Email: markus.hadler@uni-graz.at

Elisabeth Lex is an associate professor at the Institute of Interactive Systems and Data Science at Graz University of Technology. Her main research interests include user modeling and recommender systems, web and social media mining, computational social science, and information retrieval. Email: elisabeth.lex@tugraz.at