

Kerkhof, Anna; Münster, Johannes

Working Paper

Detecting Coverage Bias in User-Generated Content

CESifo Working Paper, No. 8844

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Kerkhof, Anna; Münster, Johannes (2021) : Detecting Coverage Bias in User-Generated Content, CESifo Working Paper, No. 8844, Center for Economic Studies and Ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/232441>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Detecting Coverage Bias in User-Generated Content

Anna Kerkhof, Johannes Münster

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Detecting Coverage Bias in User-Generated Content

Abstract

The importance of user-generated content is growing as media consumption is moving online; yet, investigations of media bias on user-generated content platforms are rare. We develop a novel procedure to detect coverage bias – i.e., bias in the *amount of coverage* certain topics or issues receive – on user-generated content platforms. We proceed in two steps. First, we focus on a sample of homogeneous observations and control for observable differences. Second, we compare the coverage of our observations between different language versions of the same platform in a difference-in-differences framework, which allows us to disentangle coverage bias from unobserved heterogeneity between observations. We apply our procedure to Wikipedia and examine whether it has a coverage bias in its biographies of German (and French) Members of Parliament (MPs). Our analysis reveals a small to medium size coverage bias against MPs from the center-left parties in Germany and in France. A plausible explanation are partisan contributions to the Wikipedia biographies, as we show by analyzing patterns of authorship and Wikipedia’s talk pages for the German case. Practical implications of our results include raising users’ awareness of coverage bias when searching for and processing information obtained on user-generated content platforms.

JEL-Codes: L820, L860, P160.

Keywords: coverage bias, media bias, media economics, social media, user-generated content.

*Anna Kerkhof**
*ifo Institute – Leibniz Institute for Economic
Research at the University of Munich / Germany*
kerkhof@ifo.de

Johannes Münster
University of Cologne / Germany
johannes.muenster@uni-koeln.de

*corresponding author

December 26, 2020

We thank our editor Jan Becker, two anonymous referees, Felix Bierbrauer, Maria Cubel, Marc Egger, Matthew Ellman, Marc Fischer, Oliver Gürtler, Matthias Heinz, David Jaeger, Andre Kaiser, Roman Liesenfeld, Jonas Loebbing, Sebastian Mertesacker, Gregory Martin, Florian Morath, Susanne Prantl, Riccardo Puglisi, Santiago Sanchez-Pages, Wendelin Schnedler, Detlef Schoder, Olga Slivko, David Strömberg, Marius Vogel, Van Anh Vuong, Dainis Zegners, participants of the SMYE 2016, the Ammersee Workshop 2016, the 14th Annual Media Economics Workshop, the 2nd Media Bias Workshop, the 3rd Workshop on Industrial Organization in the Digital Economy, the 2018 Political Economy Workshop in Rotterdam, and seminar participants at Barcelona, DICE, Paderborn, and St. Gallen for helpful comments and suggestions. We also thank Annika Bohnhorst, Julian Freitag, Leja Riedl, Sarah Recht, and Youpeng Zhang for excellent research assistance. Anna Kerkhof acknowledges financial support from the Joachim-Herz-Stiftung. Johannes Münster acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy -EXC 2126/1-390838866. All errors are our own. Earlier versions of the paper circulated as “Coverage bias on Wikipedia? Evidence from biographies of German members of Parliament.”

1. Introduction

User-generated content is becoming more and more important for modern media markets. As media consumption is moving online, consumers turn to Yelp to find a restaurant, to TripAdvisor to plan a vacation, and to Wikipedia to search for information (Luca, 2016). The reach of such user-generated content platforms is tremendous: as of June 2020, more than a billion hours of YouTube videos are watched every day¹, Facebook attracts 2.6 billion active users every month², and 1.5 billion users per month access the 53 million Wikipedia articles.³ Despite the significance of user-generated content for the online media landscape, examinations of media bias on user-generated content platforms are rare. In particular, while it is comparatively easy to disclose fake news, the detection of *coverage bias* – i.e., bias in the *amount of coverage* certain topics or issues receive – is challenging.

Research on coverage bias has a long history in media economics and dates back to early studies on the agenda-setting power of the mass media (McCombs and Shaw, 1972). More recently, Puglisi and Snyder (2016) and Gentzkow et al. (2016) characterize coverage bias as insidious and more prevalent than media bias through the distortion of information (p.651 and p.626, respectively), calling for close attention of researchers and policy-makers.⁴ Considering coverage bias in the context of user-generated content platforms adds two additional dimensions to the existing literature. First, being composed of millions of individual contributions, coverage bias may arise without intention of the platform itself, but as a mere result of its users’ activities, which alters the practical implications regarding whether and how to be counteracted. Second, the enormous reach of many user-generated content platforms implies that any detrimental effect of coverage bias is likely to be huge in absolute terms. For instance, if sensitive topics such as politics are subject to biased coverage and distort users’ perception of political issues, a coverage bias on a user-generated content platform may affect election outcomes and, as a consequence, public policy.⁵ Similarly, a coverage bias in consumer reviews could have a large impact on product sales⁶, and coverage biases on information sharing websites such as Reddit and YouTube may affect many users’ well-being.⁷

Detecting coverage bias is a challenge. In particular, it is difficult to determine whether the amount of coverage certain topics receive corresponds to their newsworthiness, because the universe of facts that the media could have reported is unknown – a challenge that

¹See <https://www.youtube.com/intl/de/about/press/> (June 2020).

²See <https://investor.fb.com/home/default.aspx> (June 2020).

³See <https://en.wikipedia.org/wiki/Wikipedia:About> (June 2020).

⁴Puglisi and Snyder (2016) and Gentzkow et al. (2016) use the terms “implicit media bias” and “selection bias.”

⁵See, e.g., Strömberg (2004), Eisensee and Strömberg (2007), and Durante and Zhuravskaya (2018) for empirical contributions on the effects of media coverage on public policy.

⁶See, e.g., Zhu and Zhang (2010) and Hu et al. (2008) for studies on the effect of online reviews on product sales.

⁷E.g., YouTube videos on the health risks of smoking are outnumbered by videos that portray smoking in a positive way (Freeman and Chapman, 2007). Similarly, Briones et al. (2012) find that the majority of YouTube videos on the novel human papillomavirus vaccination distribute disapproving messages on the vaccine, including accusations of conspiracy theory and the infringement of civil liberties.

Groeling (2013) calls the “Problem of the Unobserved Population” (p. 137). Existing approaches to tackle this challenge include comparing the media coverage of homogeneous cases, controlling for observable differences between observations, comparing the coverage of the same topic in different media outlets, studying an observable population, and creating an observable population (Groeling, 2013).⁸ However, unobserved heterogeneity between topics or media outlets may confound the first three approaches, and observing or creating an appropriate population is not always feasible.

This paper presents a novel approach to detect coverage bias in user-generated content. We proceed in two steps. First, similar to existing procedures, we focus on a sample of homogeneous observations and control for observable differences. Second, we compare the coverage of our observations between *different language versions of the same platform* in a difference-in-differences framework. This allows us to disentangle coverage bias from unobserved heterogeneity: Under the identifying assumption that unobserved heterogeneity between observations affects the amount of coverage in either language version equivalently, the difference-in-differences estimates are less prone to omitted variable bias than the estimates from a comparison of coverage in one language alone. Moreover, comparing two language versions of the same platform prevents omitted variable bias stemming from unobserved heterogeneity between media outlets.

As an application of our procedure, we study if Wikipedia, the world’s largest online encyclopedia, has a coverage bias in its biographies of German (and French) Members of Parliament (MPs), where we define coverage bias as the unequal coverage of otherwise comparable MPs from different political parties. Thus, a coverage bias on Wikipedia corresponds to the more extensive coverage of MPs from a particular party that cannot be explained by differences in the newsworthiness between MPs, but only by party affiliation itself.

We start the examination by considering Wikipedia coverage in terms of biography length. Biography length is easy to measure objectively; moreover, it is a meaningful measure in itself. According to the agenda-setting theory (e.g., Weaver et al., 2004), the amount of coverage of an MP (here: the length of a biography) may influence opinion formation if it distorts voters’ awareness of one candidate relative to another. Similarly, voters could interpret the amount of coverage as an indication of an MP’s quality as a politician (Burden, 2002).⁹

As argued, any investigation of coverage bias requires the researcher to disentangle the effect of the newsworthiness of a subject from the bias itself. In our application, we must differentiate between the effect of an MP’s individual characteristics on her biography length from the effect of party affiliation. In other words, it is possible that MPs about whom there is more to write accumulate within particular parties; the resulting differences

⁸See Section 2 for an extensive overview.

⁹To further check the plausibility of this claim, we conducted a classroom and a representative survey, finding that a longer MP biography may signal valence characteristics such as knowledge, strength as a public servant, and the ability to inspire people to younger audiences, but not to older readers (see Appendix A).

in Wikipedia coverage between MPs would, according to our definition, not be classified as bias. If, in contrast, partisan contributors manipulate the biographies of MPs from a specific party, we would observe a coverage bias as defined above. Since many MP characteristics are unobserved (e.g., ability, wittiness, or looks), we cannot easily distinguish their impact on biography length from the impact of the MPs' party affiliation.

Following the outline from above, we address this issue in two steps. First, we study a sample of relatively homogeneous observations. We start by considering the 18th German Bundestag (2013 to 2017) and focus on MPs from Germany's two biggest political parties, the center-right CDU/CSU and the center-left SPD, who jointly comprised more than three quarters of all MPs and formed a coalition government. This is particularly convenient for our identification strategy: when we compare the biographies of MPs from the CDU/CSU and the SPD, differences in length cannot originate from differences in government versus opposition parties, centrist versus more extreme parties, or big versus fringe parties.¹⁰ To further increase comparability, we exclude MPs in distinguished offices such as Chancellor Angela Merkel, ministers, and party heads from the analysis and we control for observable characteristics such as gender, political experience, and constituency demographics.

As a second step, we compare the length of German and English Wikipedia biographies in a difference-in-differences framework. Partisan contributors are less likely to amplify the English biographies, since German voters are unlikely to read them. Assuming that unobserved MP characteristics affect the German and the English biography length equivalently, a difference-in-differences estimation using language as a first, and party affiliation as a second difference, yields estimates of the effects of party affiliation on biography length that are less prone to omitted variable bias. Since English biographies are only available for about a quarter of our observations, we also take potential selection effects into account.

We find that biographies of MPs from the SPD are, on average, about 1,200 characters shorter than biographies of MPs from the CDU/CSU, which roughly corresponds to half a DIN-A4 page.¹¹ These differences remain after controlling for gender, political experience, outside earnings, education, and MPs' constituency demographics. To put the effect size into perspective, note that the average biography length is about 5,900 characters (2.33 pages), the median length is about 4,500 characters (1.66 pages), and the standard deviation in biography length is about 4,900 characters (nearly two pages). The difference-in-differences estimate for coverage bias, i.e., the effect of party affiliation on biography length, is about twice as large as our basic estimate, confirming that the unequal coverage between MPs from CDU/CSU and SPD is not driven by unobserved MP characteristics.

We investigate several further dimensions of coverage, too, and find that biographies of MPs from the SPD exhibit fewer images, fewer adjectives, a lower adjective to word ratio, and a smaller number of links to external websites under the control of the MP or her party than biographies on MPs from the CDU/CSU. Images and adjectives brighten texts and

¹⁰The literature on the news coverage of politicians has found a clear incumbency effect on coverage (Vos, 2014).

¹¹2,500 characters roughly correspond to one DIN-A4 page of the "Printable version" of a Wikipedia biography.

contribute to a more positive coverage, and a high number of weblinks under party control indicates that Wikipedia is used more extensively for election campaigns. Biographies of MPs from the SPD are also assigned to a lower number of Wikipedia categories, which makes them harder to find.

We stress that our approach is widely applicable. To demonstrate this point, we study if there is a coverage bias in the biographies of French Members of the National Assembly in a second application. Using the same procedure as above, we find that biographies of MPs from the center-right LR are around half a page longer than biographies of the center-left PS and its allies; this difference corresponds to around 35% of a standard deviation in biography length. The finding is supported by the difference-in-differences estimates, whose magnitude and effect size are similar to the baseline results.

We also provide a brief theoretical framework on the emergence of coverage bias in user-generated content, where we highlight the role of a platform’s users. In particular, we argue that differences in the users’ preferences, characteristics, and ulterior marketing aspirations can result in unbalanced coverage.

Regarding Wikipedia, we show that differences in the users’ partisan contributions are a likely driver of the coverage bias between center-left and center-right parties. Focusing on German MPs, we check the plausibility of this explanation in three ways. First, we identify all anonymous edits conducted from the Bundestag network by tracking the users’ IP addresses. Consistent with our explanation, we find that biographies of MPs from the CDU/CSU are edited nearly 50% more often from the Bundestag building than biographies of MPs from the SPD. Second, we document that there are fewer authors who repetitively contribute to SPD biographies only. Finally, if partisan contributions drive the differences in coverage, they should generate debates within the Wikipedia community. A testable implication is that Wikipedia’s talk pages for MPs from the SPD should be shorter. Indeed, there are fewer talk pages for MPs from the SPD, and these pages are also shorter on average.

Our paper makes three contributions to research in media economics. First, we present a novel empirical strategy to detect coverage biases in user-generated content. While existing approaches use variation in institutional features across or within online platforms (e.g., Anderson and Magruder, 2012; Mayzlin et al., 2014), we exploit variation between different language versions of a platform. The procedure is applicable beyond the context of our paper and may be used as basis for further research on coverage bias in the online media landscape, e.g., to study coverage bias in consumer reviews on Amazon, Yelp, or TripAdvisor, or to study the coverage of companies, celebrities, and certain (types of) products on information sharing websites like YouTube or Tumblr.¹²

Second, while media biases in the offline media landscape are well understood (see Puglisi and Snyder, 2016, for a survey), our results add to the scarce knowledge on media bias in user-generated content. As far as we know, we are the first to unveil a partisan coverage bias on Wikipedia, whereby we complement the findings by Greenstein and Zhu

¹²See Section 7 for further discussion.

(2012), who measure political slant in Wikipedia articles on political issues and conclude that the English language version of Wikipedia is slanted towards the Democratic party in the USA. In contrast to that, our measure is based on the amount of coverage of German (and French) MPs from different political parties.

Third, we provide evidence of a partisan coverage bias in the world’s largest online encyclopedia, which – given that Wikipedia serves as information source for billions of users – is a relevant result in itself and utters a general word of caution regarding the use of Wikipedia. If users are unaware of this type of bias during their online searches, they may be especially susceptible to it.

The paper proceeds as follows. Section 2 discusses the related literature. Section 3 describes our dataset. The empirical strategy and the results with respect to German MPs are presented in Section 4. Section 5 demonstrates the applicability of our approach beyond our main application. In Section 6, we propose a brief theoretical framework and show that partisan contributors are a likely driver of coverage bias. Section 7 discusses the practical implications, the external validity, and limitations of our analysis.

2. Related literature

Our paper contributes to three overlapping strands of literature, summarized below and in Tables 2, 3, and 4.

2.1. Media bias

First, we contribute to the large literature on media bias. Our paper is especially close to studies that develop empirical measures for media bias; see Table 2 for an overview of the most relevant papers to our context and D’Alessio and Allen (2000), Groeling (2013), and Puglisi and Snyder (2016) for extensive reviews.¹³

Manifestations of media bias typically fall into one out of two categories: the systematic selection of news items (*Which* news items are covered by the media?), and the systematic distortion of news items (*How* do the media report about news items?). Following D’Alessio and Allen (2000), we classify the former manifestation as *coverage bias* and the latter as *statement bias* (Table 2, Column 3).¹⁴ As our paper presents a novel approach to detect coverage bias in user-generated content, we focus on reviewing papers that present alternative measures for coverage bias, but we also include the most important papers detecting statement bias at the bottom of Table 2 for the sake of completeness.

As argued in Section 1, detecting coverage bias is a challenge, because it is difficult to determine whether the amount of coverage certain topics receive corresponds to their newsworthiness. Column 4 in Table 2 summarizes the existing approaches to disentangle

¹³The theoretical literature on media bias is surveyed by Gentzkow et al. (2016).

¹⁴Puglisi and Snyder (2016) refer to “implicit” and “explicit bias”; Gentzkow et al. (2016) speak of “filtering” and “distortion bias.”

coverage bias from (unobserved) variables affecting the newsworthiness of a topic.¹⁵

A straight-forward procedure is to control for observable differences between observations as in Schiffer (2006), whereby any remaining differences in coverage can be ascribed to coverage bias. However, many determinants of a topic’s newsworthiness are unobserved (e.g., the ability of a politician or the quality of a product), whereby the estimates from this approach are prone to omitted variable bias.

Procedures that either compare the coverage of homogeneous observations within the same media outlet or the coverage of the same observations across different media outlets are better able to take unobserved heterogeneity between observations into account. Niven (2003), Puglisi (2011), Brandenburg (2005), and Morris and Francia (2010), for instance, pursue the former approach, arguing that differences in the coverage of comparable observations are unlikely to result from differences in the observations’ newsworthiness.¹⁶ Conceptually similar, Barrett and Peake (2007), Groeling and Baum (2008), Larcinese et al. (2011), Puglisi and Snyder (2011), Lott and Hassett (2014), and Galvis et al. (2016) pursue the latter approach, where comparing the coverage of the same observations across media outlets rules out that differences in coverage are driven by unobserved heterogeneity between observations.¹⁷ Yet, estimates from comparing the coverage of homogeneous observations within the same media outlet may still be affected by unobserved differences in newsworthiness between observations, and estimates from comparing the same observations across different media outlets may be confounded by heterogeneity between news outlets.

A final approach to study coverage bias is to focus on observations with an observable population, either because it is observable per se as in Groeling (2008), Aday (2010), Soroka (2012), Garz (2014), and Heinz and Swinnen (2015), or because the authors could create it as in Butler and Schofield (2010) and Dertwinkel-Kalt et al. (2019). This procedure, however, is often not feasible for the particular research question at hand. Our application, for instance, would require an observable population of all MP characteristics that determine MPs’ newsworthiness.

Our paper presents a novel approach to detect coverage bias in user-generated content that improves upon the shortcomings of the existing procedures. As outlined in Section 1, we proceed in two steps. First, similar to Schiffer (2006), Niven (2003), Puglisi (2011), Brandenburg (2005), and Morris and Francia (2010), we focus on a sample of homogeneous observations and control for observable differences. Second, we compare the coverage of our observations between *different language versions of the same platform* in

¹⁵Our classifications of approaches follows Groeling (2013).

¹⁶Puglisi (2011), for instance, analyses the New York Times’ coverage of several topics from public policy and finds that during a presidential campaign, The New York Times gives more emphasis to topics on which the Democratic party is perceived as more competent (civil rights, health care, labor and social welfare) when the incumbent president is a Republican.

¹⁷E.g., Larcinese et al. (2011) consider coverage on unemployment numbers and show that newspapers with pro-Democratic endorsement patterns systematically give more coverage to high unemployment when the incumbent president is a Republican than when the president is Democratic, compared to newspapers with pro-Republican endorsement patterns.

a difference-in-differences framework, whereby we can disentangle coverage bias from unobserved heterogeneity between observations. In particular, if unobserved heterogeneity between observations affects the amount of coverage in either language version equivalently, the difference-in-differences estimates are less prone to omitted variable bias than the estimates from a comparison of coverage in one language alone. Moreover, comparing two language versions of the same platform prevents omitted variable bias stemming from unobserved heterogeneity between media outlets. We perceive this novel procedure to detect coverage bias as the major contribution of our paper.

Table 2 also demonstrates that most of the research on media bias has focused on traditional media – such as newspapers and television – while comparatively little is known about biases in user-generated content. A notable exception is Greenstein and Zhu (2012) who apply automated text analysis to measure the slant in Wikipedia articles on political issues. In their paper, slant is an intrinsic property of an article, measuring whether its language is more typical for the Republican or Democratic party in the USA. In contrast to that, our measure of coverage bias is based on the overall coverage of German and French MPs from different political parties. In addition, research on media bias has mainly focused on media in the USA, while research on biases of media in other languages is rare. The political systems and the media systems of the USA, Germany and France differ along many dimensions (Persson and Tabellini, 2005; Hallin and Mancini, 2004), and it is a priori unclear whether results obtained in the US will hold in the European cases. Indeed, a naive extrapolation of the result that the English Wikipedia has a pro-liberal bias (Greenstein and Zhu, 2012) might suggest that, when comparing the two biggest German parties, the German Wikipedia is biased in favor of the center-left SPD relative to the center-right CDU/CSU. Our results, however, show the exact opposite.

2.2. User-generated content

Our paper also adds to the growing research on user-generated content and social media (see Luca, 2016, for a survey), where it is particularly close to the literature on (promotional) consumer reviews, which we summarize in Table 3. These papers typically develop procedures to detect promotional or fake reviews as in Mayzlin et al. (2014) and Luca and Zervas (2016), or they assess the effect of promotional reviews on prices (Melnik and Alm, 2002), sales (Chevalier and Mayzlin, 2006; Anderson and Magruder, 2012; Lu et al., 2013), or product growth (Clemons et al., 2006).¹⁸

Promotional reviews are conceptually similar to ulterior marketing aspirations that lead to coverage bias on user-generated content platforms (see also Section 6.1 on this), whereby the former strand of research is particularly close to our paper. Luca and Zervas (2016) use the results of Yelp’s filtering algorithm as a proxy for fake reviews and find that a restaurant is more likely to commit review fraud on Yelp when its reputation is weak. Mayzlin et al. (2014) study promotional content in hotel reviews and show that hotels

¹⁸Theoretical contributions are provided by Mayzlin (2006) and Dellarocas (2006).

with a high incentive to fake have more positive reviews on TripAdvisor – where it is less costly to leave a review – relative to Expedia; the neighbors of hotels with a high incentive to fake, in contrast, have more negative reviews on TripAdvisor than on Expedia. In contrast to that, our approach does not compare equivalent observations across platforms, but within two language versions of the same platform. Our empirical strategy hinges on the assumption that the difference in the coverage of two subjects in two different language versions of the same platform was identical in the absence of bias, which might be more plausible than the assumption that the difference in the coverage of two subjects between *different platforms* was identical. Moreover, our data requirements are more modest: our procedure does not require an observation to be covered by two independent platforms, but in two language versions of the same platform instead. Since many relevant user-generated content platforms such as Wikipedia, YouTube, TripAdvisor, Twitter, Yelp, and Facebook are available in several languages, we consider the latter condition as easier to meet. In contrast to Luca and Zervas (2016), our approach does not rely on the availability of pre-specified promotional content.

A further group of papers on user-generated content examines the individual motives for users to contribute to platforms in general and to Wikipedia in particular; the results from these papers link to the theoretical framework from Section 6.1. Wang (2010) shows that a positive reputation of a user increases her productivity, while Chen et al. (2010) demonstrate that users become more productive when they realize that others are more productive than they are. In the context of Wikipedia, Zhang and Zhu (2011) show that users’ activity decreases in community size, i.e., users contribute more if the number of potential readers grows.

Several analyses, surveyed by Mesgari et al. (2015), are linked to our paper, because they examine contributions to Wikipedia or Wikipedia itself. For instance, it is often claimed that Wikipedia suffers from systemic coverage biases induced by its users’ demographics – a majority is English-speaking, white, male, and Internet-affine (Halavais and Lackaff, 2008) – leading, e.g., to the under-representation of women (Reagle and Rhue, 2011; Hinno Saar, 2019) as well as to the over-representation of Western culture (Callahan and Herring, 2011) and other “male” topics such as naval sciences, military, science fiction, and fantasy (Rosenzweig, 2006). Similarly, while several studies confirm that Wikipedia exhibits only few factual errors (Mesgari et al., 2015), Brown (2011) unveils frequent errors of omission, which motivates our focus on coverage.

2.3. Social media and political outcomes

Third, our analysis is linked to research on social media and political outcomes (see Goldfarb and Tucker, 2019, for a survey), in particular to papers that examine how broadband diffusion (e.g., Falck et al., 2014; Campante et al., 2018; Gavazza et al., 2019) and social media (e.g., Bond et al., 2012; Enikolopov et al., 2020) affect voting behavior and political participation (Zhuravskaya et al., 2020, provide a survey). Since coverage bias on

Wikipedia that stems from partisan contributions is a type of political advertising, the paper by Liberini et al. (2020) is especially close to our study. The authors show that micro-targeted political advertising through social media has a significant effect on voters during the 2016 US presidential elections: exposure to these ads made individuals less likely to change their initial voting intentions, particularly among those who had expressed an intention to vote for Donald Trump.

3. Data

Kürschner (2015) provides a list of all members of the 18th German Bundestag, including information on their education, political experience, and party affiliation. Data on the MPs’ offices during the 18th Bundestag stems from bundestag.de. Information on the MPs’ ancillary incomes is collected from abgeordnetenwatch.de: German MPs are obliged to declare their ancillary income by means of ten different categories; following the literature, we use these categories’ mean values in our analysis (e.g., Becker et al., 2009).¹⁹ Data on constituency demographics stems from the electoral management body.²⁰ Data on the French MPs in the 14th legislature stems from the National Assembly’s homepage assemblee-nationale.fr.

From the initial list of German MPs, we exclude Chancellor Angela Merkel, 35 MPs in distinguished offices (party heads and ministers from the 18th or a preceding Bundestag), and nine MPs who had already left the 18th Bundestag before we started our data collection.²¹ Of 598 remaining MPs in our dataset, 294 MPs are from the CDU/CSU, 184 from the SPD, 62 from the Left, and 58 from the Greens; no MP switched parties during the observation period. Moreover, 266 of the 598 MPs were directly elected to the Bundestag.²²

German Wikipedia biographies exist for all MPs in our dataset, English biographies are available for 138 German MPs. Similarly, French Wikipedia biographies exist for all MPs in our dataset, English biographies are available for 296 French MPs. The numbers of characters, words, adjectives, images, and categories per biography are obtained via Wikipedia’s API. In addition, each biography links to a background site that provides a list of unique authors and the number of edits. Information on the number of weblinks under party control, translation indicators, MPs’ English homepages, and criticism is hand coded. Table 1 provides summary statistics of all the variables used in our analysis.

¹⁹See www.abgeordnetenwatch.de/blog/nebeneinkuenfte2016 (June 2015).

²⁰See www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_13/strukturdaten/ (Dec 2017).

²¹See Appendix B for robustness checks that include these observations.

²²In Germany, each voter casts two votes in elections to the Bundestag. The first vote decides which local candidate from each of Germany’s 299 constituencies will be sent to the Bundestag. The second vote is cast for a party list and determines the parties’ relative strength in the Bundestag. The Bundestag has a minimum number of 598 seats. In its 18th election term, it was amplified by four additional “overhang” seats, since the CDU won more constituency seats than it would have been entitled to based on the second-vote share. A further 29 “balance seats” sustain the parties’ relative strengths, leading to a total number of 631 seats.

4. Evidence from the German Bundestag

Our goal is to answer if there is a coverage bias on Wikipedia, i.e., if comparable MPs from different political parties are covered differently in terms of their biography length. We proceed in two steps. First, we focus on the biographies of a relatively homogeneous group of backbenchers from CDU/CSU and SPD. Any differences in biography length can therefore not originate from differences in government versus opposition parties, centrist versus more extreme parties, or big versus fringe parties. Moreover, the personalities of single prominent MPs cannot affect our results. Second, to disentangle the effect of party affiliation from unobserved MP characteristics, we compare the MPs’ German and English biography length in a difference-in-differences framework.

4.1. Basic estimation

Figure 1 shows the average biography length per party in characters. Since 2,500 characters roughly correspond to one DIN-A4 page of a biography’s PDF print version, biographies of MPs from the CDU/CSU and the Greens are around two and a half pages, biographies of MPs from the SPD are around two pages, and biographies of MPs from the Left nearly three pages long.²³ To put these numbers into perspective, note that the average biography length across all MPs is around 2.33 pages (5901.6 characters) and the median around 1.66 pages (4487.5 characters), with a standard deviation of nearly two pages (4890.87 characters). Hence, the difference in average biography length between the CDU/CSU and the SPD corresponds to roughly 25% of a standard deviation.

To control for observable MP characteristics, we further estimate the regression equation

$$length_i^G = \beta_0^G + \beta_1^G P_i + \beta_2^G X_i + u_i^G, \quad (1)$$

by OLS, where $length_i^G$ corresponds to the German biography length of MP i , P_i is a vector of party dummies with the SPD as omitted category, and X_i is a vector of control variables including MP i ’s gender, political experience, education, ancillary income, and – if MP i is directly elected – constituency demographics.²⁴ The parameter of interest is β_1^G , as it measures the effect of party affiliation on biography length relative to the omitted category SPD. In other words, β_1^G corresponds to the average coverage bias relative to the

²³There are three potential explanations for the significantly longer Wikipedia biographies of Left MPs. First, from the parties whom we consider, the Left are positioned furthest away from the center of the political spectrum and may therefore have worse access to alternative media channels – such as newspaper or television interviews – to communicate with their voters. Second, Forschungsgruppe Wahlen (2014) has shown that voters of the Left are more Internet-affine than, for instance, voters of CDU/CSU and SPD. Finally, MPs from the Left may consider an extensive web presence as more important than MPs from other parties. Freitag et al. (2020), for instance, find that 74% of the MPs from the Left have used Twitter in the six months following the constitution of the 19th Bundestag in 2017, while only 55.6% of the MPs from the SPD, 36% of the MPs from the CDU and 30.4% of the MPs from the CSU have done so.

²⁴Several large cities such as Cologne and Berlin are divided into several electoral districts. The data on those cities are aggregated and therefore not independent. We account for that by clustering the standard errors at the city level, and obtain 249 clusters.

SPD.

Table 5 shows the results. Column 1 replicates the party averages from Figure 1 relative to the omitted category SPD, i.e., the CDU/CSU coefficient corresponds to around half a DIN-A4 page. When we control for gender (column 2), political experience (column 3), and doctoral degrees as well as ancillary incomes (column 4), the size of the CDU/CSU coefficient decreases to a third of a DIN-A4 page (which corresponds to 15% of a standard deviation in the dependent variable), but remains statistically significant.²⁵

In column 5, we control for an MP’s constituency’s demographics. Since only about 50% of the MPs are directly elected in a constituency and the other 50% enter the Parliament via state lists and thus lack a clear constituency affiliation, we consider only 266 observations here. Plausibly, population density has a highly significant effect on biography length: broadband connections in urban areas are usually better than in rural areas, which facilitates the use of Wikipedia and increases the biographies’ likelihood of being read. In contrast to that, education or a constituency’s share of voters aged 18 to 35 – those who are most prone to use the Internet as a source of political information – have no statistically significant effect. The CDU/CSU coefficient is statistically significant despite the decreased sample size and corresponds to around 30% of a standard deviation in the dependent variable.²⁶

In addition, we perform several two-sided *t*-tests to check whether differences in coverage between other parties are statistically significant. In columns (1) and (2), no other differences in biography length between any two parties are significant, while in columns (3) and (4), the difference between CDU/CSU and Left is significant at the 5%-level. Moreover, Section 6.4 demonstrates that negative coverage plays only a minor role and does not affect our results.

4.2. Difference-in-differences estimation

If party affiliation is correlated to unobservable MP characteristics (e.g., if more salient politicians accumulate within particular parties) an OLS estimation of equation (1) might suffer from an omitted variable bias. To disentangle the effect of party affiliation on biography length from the effect that unobserved MP characteristics could have, we exploit variation between the MPs’ biographies in the German and in the English Wikipedia. Under the identifying assumption that unobserved MP characteristics affect the German and the English biography length equivalently, a difference-in-differences estimation using language as a first, and party affiliation as a second difference, yields estimates of the effects of party affiliation on biography length that are less prone to omitted variable bias.

English Wikipedia biographies exist for 138 MPs in our sample; Figure 2 shows their average length. English biographies of MPs both from the CDU/CSU and the SPD are

²⁵We obtain similar estimates when we consider more prominent political figures such as ministers and party heads, too (see Appendix B).

²⁶Note that we do not consider observations from the Greens and the Left here, since only three Left MPs and one Green MP are directly elected.

nearly two pages, English biographies of MPs from the Greens are around three pages, and English biographies of MPs from the Left are around one and a third page long. To put these numbers into perspective, note that the average English biography length across all MPs is around two pages (4909.5 characters) and the median around one and a half pages (3817 characters), with a standard deviation of 1.66 pages (4119.6 characters).

Suppose that the MPs' English biography length is determined analogously to equation (1). Then,

$$length_i^E = \beta_0^E + \beta_1^E P_i + \beta_2^E X_i + u_i^E. \quad (2)$$

Differencing equations (1) and (2) yields

$$length_i^G - length_i^E = \beta_0 + \beta_1 P_i + \beta_2 X_i + u_i, \quad (3)$$

where $\beta_k = \beta_k^G - \beta_k^E$ and $u_i = u_i^G - u_i^E$. The parameter of interest in equation (3) is β_1 , whose interpretation hinges on further assumptions on β_1^E . As argued, biography length may serve as a positive signal about MPs' valence characteristics (see Appendix A), so partisan contributors have an incentive to amplify the Wikipedia biographies of MPs from one specific party. While partisan contributions may lead to a coverage bias in the German Wikipedia, this is unlikely to occur in the English Wikipedia version, since German voters are unlikely to read their MPs' English biographies. Thus, we assume that $\beta_1^E = 0$; see Appendix C for further discussion. If party affiliation has no effect on the MPs' English biography length, β_1 corresponds to β_1^G in equation (1). Hence, under the identifying assumption that unobserved MP characteristics affect the German and the English biography length equivalently and thereby cancel out, equation (3) yields an unbiased estimate for β_1^G , the average coverage bias relative to the SPD.

Naively regressing equation (3) on the subsample of MPs where we observe English biographies may lead to sample selection bias, though. Thus, we also consider a selection model consisting of equation (3) and the selection equation

$$d_i = 1[\delta Z_i + \epsilon_i > 0], \quad (4)$$

where the dependent variable d_i indicates if there exists an English biography for MP i (see also Greene, 2003). For the selection model to work well, Z_i has to include additional variables that determine whether or not there exists an English Wikipedia biography for MP i , but do not affect the dependent variable in equation (3). We use $Z_i = (P_i, X_i, I_i)$, where I_i is a vector of variables that determine the international political relevance of MP i : the number of election terms in the European Parliament and the number of international offices during the 18th Bundestag (Commission of Foreign Affairs, Commission of European Affairs, and being head of an international parliamentary group). While international offices plausibly increase the probability that an English biography will be set up, they are not very salient, and hardly anything is written on them. We estimate the

selection model by Heckman two-step and by maximum likelihood.²⁷

Table 6 shows the results. Column 1 shows the results of an OLS estimation of equation (3). Although the size of the CDU/CSU coefficient is considerable – it corresponds to nearly a DIN-A4 page, which is nearly 60% of a standard deviation in the dependent variable and around twice as much as in Table 5 – it is not statistically significant. Columns 2 and 4 show the results from a two-step and a maximum likelihood estimation of equations (3) and (4), respectively. Since the magnitude of the coefficients is not directly comparable to the OLS estimates in column 1 of Table 6, we also display their marginal effects at the mean (MEM) in columns 3 and 5, which are similar to the OLS estimates in column 1.²⁸ In contrast to the OLS regression, the CDU/CSU coefficient is statistically significant at the 10%-level in the two-step, and at the 1%-level in the maximum likelihood specification.²⁹ Hence, the difference-in-differences estimation provides even stronger evidence of a coverage bias against MPs from the SPD than our results from section 4.1.

4.3. Further dimensions of coverage

This section shows that our results on coverage bias from Section 4.1 hold for alternative dependent variables in equation (1): adjectives, images, categories, and weblinks under party control. First, we study the occurrence of adjectives in the biographies. Adjectives can make a text more lively and colorful. Moreover, the literature on sentiment analysis shows a correlation between the presence of adjectives and the subjectivity of a sentence, i.e., the degree to which opinions are expressed (e.g., Bruce and Wiebe, 1999; Wiebe et al., 2004; Pang and Lee, 2008). We find that biographies on MPs from the CDU/CSU have around ten more adjectives and a higher adjective-to-word ratio than biographies on MPs from the SPD; the former difference corresponds to around 18% of a standard deviation in the dependent variable and is statistically significant at the 5%-level (Table 7, columns 1 and 2).

Similarly, images make biographies more lively and attractive. We find that biographies of MPs from the CDU/CSU contain on average 0.33 images (30% of a standard deviation) more than biographies of MPs from the SPD; the difference is statistically significant at the 1%-level (Table 7, column 3).

Next, Wikipedia articles are usually assigned to categories, that “help readers to find, and navigate around, a subject area, to see pages sorted by title, and to thus find article relationships.”³⁰ Thus, assigning a biography to many different categories – examples

²⁷We do not believe that the probability of having an English Wikipedia biography is affected by MP preferences: only five of the 138 MPs with an English Wikipedia biography even provide their personal homepage in English.

²⁸MEMs measure the effect of a change in one of the regressors on the conditional mean of the difference in biography length, evaluated at the mean values of all other covariates, and given that this difference is observed. MEMs for dummy variables show how the biography length changes as the dummy changes from 0 to 1, holding all other covariates at their mean values.

²⁹By using the likelihood function rather than the method of moments, the latter estimation procedure is more efficient and yields smaller standard errors; see Greene (2003) for details.

³⁰See <https://en.wikipedia.org/wiki/Help:Category> (Dec 2018.)

include “Member of Parliament”, “Politician”, or “German” – enhances its chances to be found by readers. Biographies on MPs from the CDU/CSU are assigned to 0.65 more categories than biographies on MPs from the SPD; this difference corresponds to 25% of a standard deviation in the dependent variable and is statistically significant at the 1%-level (Table 7, column 4).

Finally, many Wikipedia biographies provide weblinks to external websites. We classify a link as “under party control” if it directs to a website that is under the obvious and substantial influence of an MP’s party (e.g., MPs’ personal or party homepages). Weblinks under party control facilitate the use of Wikipedia as a platform for political campaigns. We find that biographies on MPs from the CDU/CSU contain about half a weblink more than biographies on MPs from the SPD (60% of a standard deviation); the difference is statistically significant at the 1%-level (Table 7, column 5).³¹

5. Evidence from the French National Assembly

We argue that our approach to detect coverage bias in user-generated content is widely applicable. Considering Wikipedia, it can be applied to all settings where a sufficient number of Wikipedia articles in the native and at least one foreign language exist. To demonstrate the point, this section studies coverage bias in the Wikipedia biographies of French MPs as a second application of our approach. Moreover, as we find a coverage bias in favor of the center-right *Les Républicains*, we also show that the results from above are not a country specific phenomenon.

We proceed analogously to our main analysis. First, we focus on a relatively homogeneous group of backbenchers. To this end, we limit our attention to the National Assembly of the 14th legislature of the French Fifth Republic (2012 to 2017), where the majority of French MPs was affiliated to the center-right *Les Républicains* (LR) or the center-left *Parti Socialiste* and its allies (SER).

We retrieve a list of all MPs in the 14th legislature from the National Assembly’s homepage `assemblee-nationale.fr`. Next, we use the Wikipedia API to collect data on all MPs’ French and English biography length in characters, the number of images, external weblinks, and categories. To facilitate a comparison with the results from the German MPs, we retrieved this information as of Oct 12th 2015, just as we did for the main sample. Data on international offices – MPs’ membership in the Committee on Foreign Relations and the Committee on European Affairs – stems from `assemblee-nationale.fr`, too. Excluding all MPs who left the French parliament early and excluding the 10% MPs with the longest biographies to focus on the backbenchers, we are left with 504 MPs of whom 172 are affiliated to LR, 261 to SER, 21 to the *Union des Démocrates et Indépendants*

³¹The finding is supported by one particular incident. According to the Wikipedia talk pages, it is difficult to incorporate weblinks into articles, partly because of the German umlauts (ä, ö, ü). There exists, however, a user called “Cducsu” who has written a program to facilitate the procedure and has used it to install weblinks underneath biographies of MPs affiliated to the CDU/CSU that redirect to the homepage of the CDU/CSU parliamentary group.

(UDI), 15 to the *Groupe de la Gauche Démocrate et Républicaine* (GDR), 16 to the *Radical, Républicain, Démocrate et Progressiste* (RRDP), and further 17 *députés non-inscrits* (NI) who do not belong to any parliamentary group in the National Assembly. English Wikipedia biographies exist for 296 (or 58.7%) of these MPs.

Figures 3 and 4 illustrate our data. Figure 3 shows that French biographies of MPs from the SER are about 1,500 characters shorter than biographies of MPs from the LR, which is slightly more than half a DIN-A4 page. To put the numbers into perspective, note that the average biography for a French MP is around three and a half pages (8850.3 characters) and the median biography is around three pages (7687 characters) long, with a standard deviation of 4310.3 characters. The sign of the effect reverses in Figure 4: English biographies of French MPs from the SER are on average around 200 characters longer than English biographies of MPs from the LR.

Table 9 shows the results from running regression equations (1), (3), and (4) on the sample of French MPs, including only party dummies, and using the SER as baseline. In Column 1, we estimate equation (1) and use $length_i^F$, i.e., the length of the French biography of MP i , as dependent variable. In line with Figure 3, the estimate for LR is positive and highly statistically significant; the difference in coverage between MPs from the LR and the SER corresponds to around 35% of a standard deviation in the dependent variable. Thus, just as in the German case, French conservatives receive more coverage than MPs from the center-left. The pattern is confirmed by further dimensions of coverage. In columns 2 to 4, we use the number of images, external weblinks, and categories as dependent variables. The estimates in columns 3 and 4 are positive and significant at the 1% level; the effect sizes correspond to 24% and 44% of a standard deviation in the dependent variable, respectively. The estimate in column 2 is negative, but not statistically significant.

Since the difference in coverage could also be driven by unobserved heterogeneity, we use our difference-in-differences approach from equation (3) in Columns 5 and 6, where $length_i^F - length_i^E$ is the dependent variable. In Column 5, we do not take potential selection into having an English Wikipedia biography into account. The estimate for LR is around 50% smaller than in Column 1 and only weakly statistically significant. Column 6 shows the ML results from estimating the selection model consisting of equations (3) and (4), using indicators for membership in the Committee on Foreign Relations and the Committee on European Affairs for identification. The estimate for LR is positive, too, and highly statistically significant. To compare its magnitude with the OLS estimates, Column 7 provides the MEMs for the coefficients. The MEM for LR is very similar to the OLS estimate from Column 1 and confirms the existence of a coverage bias against MPs from the SER. We conclude that the application of our approach to French MPs yields comparable results to our main analysis, although the coverage bias is smaller in absolute terms for French MPs.

6. Partisan contributions

How does coverage bias on user-generated content platforms emerge? This section develops a brief theoretical framework that highlights the role of a platform’s users. In particular, we argue that differences in the users’ preferences, characteristics, and ulterior marketing aspirations may result in the unbalanced coverage of certain topics. Coming back to the German Wikipedia, we demonstrate that differences in users’ partisan contributions are a likely driver of the unbalanced coverage of MPs from CDU/CSU and SPD.

6.1. Theoretical considerations

In contrast to traditional media outlets, coverage bias on user-generated content platforms accrues from the contributions of individual users and not necessarily from the intentions of the platform itself. For instance, if two topics are equivalently newsworthy, but a larger number of users contributes to one of them, coverage bias may arise. Similarly, if the users’ contributions to a specific topic are relatively large, the overall coverage may become unbalanced.

Figure 5 illustrates our theoretical considerations. A user-generated content platform covers two topics, *A* and *B*, which are equivalently newsworthy. Yet, the amount of coverage on topic *B* exceeds the amount of coverage on topic *A*; in other words, there is a coverage bias against *A*. There are two potential drivers of this bias: First, a larger number of users could contribute to topic *B*; second, the individual contributions to *B* could be relatively larger than the contributions to *A*. These drivers are not mutually exclusive. Figure 5 also shows that the emerging coverage bias may, in turn, affect political outcomes, sales, and users’ well-being (see also Section 1 on the potential effects of coverage bias).

There are many potential reasons for differences in user contributions to the platform. On the one hand, individual characteristics, preferences, and interests determine how much and to which topics the users contribute. E.g., as discussed in Section 2.2, Wikipedia’s users are predominantly white, male, English-speaking, and Internet-affine (Halavais and Lackaff, 2008), which has induced some structural biases on the platform. Reagle and Rhu (2011) and Hinnosaar (2019), for instance, show that women are underrepresented on Wikipedia. Halavais and Lackaff (2008) document a coverage bias towards naval sciences and military, whereas Rosenzweig (2006) speaks of “geek priorities” (p.127) that manifest, e.g., in a large number of articles on obscure characters from science fiction and fantasy.

On the other hand, it is possible that users’ contributions are driven by ulterior marketing aspirations, i.e., intentional causation of coverage bias (in contrast to the unintentional causation of bias through the channels outlined above). Users could, for instance, raise the amount of coverage of a certain politician or of a certain product on purpose to increase the vote share of the politician or the sales of the product. E.g., Mayzlin et al. (2014) unveil the existence of promotional reviews on TripAdvisor; Lu et al. (2013) provide similar evidence from online restaurant reviews.³² Likewise, the coverage of politicians may be

³²See also Mayzlin (2006) for a game theoretic model on promotional chat on the Internet.

influenced by hidden marketing aspirations – i.e., partisan contributions – of the users.

Theoretically, coverage bias on Wikipedia may thus be driven by structural biases of the platform (e.g., if there are more users who are interested in MPs from the center-right parties) or by partisan contributions (e.g., if supporters of the center-right parties manipulate the biographies of associated MPs). In the former case, differential coverage is likely to accrue in all language versions; in the latter case, it is more likely to occur in the language version that matters to the MPs’ potential voters. Since differential coverage that occurs both on the German (French) *and* the English biographies cancels out in a difference-in-differences estimation, the results from Sections 4 and 5 support the role of ulterior marketing aspirations. To further endorse this claim, the remainder of this section shows that partisan contributions are a likely driver of our main results.³³

6.2. Authorship patterns

To support the idea that partisan contributions drive the coverage bias on Wikipedia, this section shows that there are more Wikipedia authors who repetitively contribute to the biographies of MPs from the CDU/CSU than to biographies of MPs from the SPD. Moreover, we show that biographies of MPs from the CDU/CSU are edited more often from the Bundestag building.

First, we check if there are authors who repetitively amplify the biographies of MPs from one specific party. For each article, Wikipedia displays either the authors’ user name or, in case of anonymous contributions, their IP address. We identify all authors who contribute to at least 10% of the biographies of MPs from the CDU/CSU or from the SPD and classify them as “repetitive contributors.” Next, we check which of these repetitive contributors amplify the biographies of just one specific party and classify them as “party-specific repetitive contributors.” We find that there exist 37 repetitive and three party-specific repetitive contributors for the SPD. Moreover, there exist 42 repetitive and five party-specific repetitive contributors for the CDU/CSU.

Next, we track all contributions of anonymous users whose IP addresses are displayed. Building on a study by Bayerischer Rundfunk (2017), we consider all IP addresses that can be linked to the Bundestag building.³⁴ We find that 50.3% of the biographies of MPs from the CDU/CSU, and 52.2% of the biographies of MPs from the SPD were edited at least once from the Bundestag. Moreover, biographies of MPs from the CDU/CSU were edited on average 5.2 times, while biographies of MPs from the SPD were edited 3.8 times on average. There exist also differences in terms of the number of characters added or deleted: while on average 627.02 characters are added and 482.20 characters deleted from biographies of MPs from the CDU/CSU by Bundestag contributors, only 232.49 characters are added and 157.65 characters deleted from biographies of MPs from the SPD. Note that the difference in the net number of characters added between the CDU/CSU and the SPD

³³For the sake of conciseness, we focus on the German case.

³⁴According to `bundesedit.de` (June 2018), the digits “193.17.” at the beginning of an IP address indicate the Bundestag network.

– around thirty – is much smaller than the effect we find in Section 4. Hence, edits by anonymous users from the Bundestag alone cannot explain the coverage bias documented above.

6.3. Talk pages

Partisan contributions are likely to entail discussions if additional content should be included or not. The purpose of a Wikipedia talk page “is to provide space for editors to discuss changes to its associated article or project page.”³⁵ Hence, existence and length of a talk page can indicate the occurrence of partisan contributions.³⁶

We find that talk pages exist for 132 MPs from the CDU/CSU (44.9%) and for 79 MPs from the SPD (42.9%). Moreover, talk pages on MPs from the SPD are on average half a page shorter than talk pages on MPs from the CDU/CSU (3589.52 versus 5029.30 characters), which corresponds to 10% of a standard deviation in talk page length.

6.4. Negative coverage

Finally, we show that negative coverage does not drive our results on coverage bias. Our analysis assumes that Wikipedia coverage is beneficial for MPs. Media coverage is generally beneficial for politicians, e.g., because it increases name recognition (Burden 2002); this applies in particular to less well-known politicians such as the MPs in our dataset. In addition, we have conducted a classroom survey that shows that a longer biography signals knowledge, strength as a public servant, and the ability to inspire people (see Appendix A). Wikipedia coverage could hurt MPs, however, if the biographies contained large amounts of criticism. To resolve such concerns, this section shows that negative coverage is a minor concern.

To explore the extent of negative coverage, we systematically identify negative sentences in the biographies. In a first step, we search each biography for sentences that contain the word stems of “Kritik” (“criticism”), “Diskussion” (“discussion”), “Rück- / Austritt” (“resignation”), “Skandal” (“scandal”), and “Affaire” (“affair”). Next, we determine if these sentences actually criticize the MP. We find that negative coverage is a minor issue: only 7% of the biographies contain more than one sentence of negative coverage, and 90% do not contain any negative coverage at all.

To confirm that the results from Section 4 are not driven by different amounts of negative coverage, we estimate equations (1), (3) and (4) on the subsample of MPs whose biographies do not contain any criticism at all. Table 8 shows the results. Although the

³⁵See https://en.wikipedia.org/wiki/Wikipedia:Talk_page_guidelines (Dec 2018).

³⁶To support this presumption, we let a Research Assistant code whether the talk pages contain “criticism on the biography’s length”, “criticism on the biography’s content”, or neither of it. She found that found that eighteen talk pages for MPs from the SPD exhibit criticism on length (22.8% of the existing talk pages), and four talk pages exhibit criticism on content (5% of the existing talk pages). Moreover, 27 talk pages for MPs from the CDU/CSU exhibit criticism on length (20.5% of the existing talk pages), and four talk pages exhibit some criticism on content (3% of the existing talk pages).

estimates are smaller and not as statistically significant as in Tables 5 and 6, they are qualitatively similar.³⁷

7. Conclusion

This paper presents a novel approach to detect coverage bias in user-generated content. The procedure involves two steps: First, we focus on a sample of homogeneous observations and control for observable differences; second, we compare the coverage of our observations between different language versions of the same user-generated content platform in a difference-in-differences framework. As opposed to existing approaches to detect coverage bias in the media – such as the comparison of similar cases within the same media outlet or the comparison of equivalent cases across media outlets – the estimates from our approach are less prone to omitted variable bias.

An application of our procedure to Wikipedia unveils a coverage bias against MPs from the center-left relative to MPs from the center-right in Germany and in France. We also present a brief theoretical framework on the driving forces behind our results and, focusing on the German case, provide empirical evidence that supports the role of partisan contributions as a potential driver of the emergence of coverage bias on Wikipedia.

Our study is relevant for media researchers, policy makers, and practitioners for two broad reasons. First, our approach to detect coverage bias in user-generated content is widely applicable and could be used to unveil further biases on Wikipedia or on alternative user-generated content platforms. One could, for instance, use the procedure to study coverage bias in consumer reviews on Amazon, Yelp, or TripAdvisor. Similarly, one could examine the coverage of companies, celebrities, and certain (types of) products on information sharing websites like YouTube or Tumblr. Extending our analysis to further Members of Parliament, Acts of Parliament, or political initiatives such as Fridays For Future or Occupy Wall Street would be feasible as well. Practitioners could employ our approach to monitor the coverage of their own business relative to their competitors.

Second, the results of our application provide the basis for a general debate on how coverage bias on user-generated content platforms could be counteracted. As argued in Section 6.1 and illustrated in Figure 5, coverage bias on user-generated content platforms stems from the individual contributions of its users and not necessarily from the intentions of the platform itself. Hence, any media policy that aims to diminish coverage biases must address the users' actions. While a full-fledged discussion is beyond the scope of this paper, we suggest two general approaches. On the one hand, one could raise the platform's and the platform's users' awareness of coverage bias. E.g., while warning templates for poor quality articles on Wikipedia already exist, a similar type of template could be introduced, stating that not only the information itself, but also the amount of information

³⁷We also perceive pure vandalism as a minor issue: false statements are quickly detected by Wikipedia's control mechanisms and are thereupon erased. Moreover, if MPs are involved in scandals such as plagiarism or the consumption of illegal drugs, they usually resign, and these observations are excluded from our analysis.

matters and should be processed accordingly.³⁸ On the other hand, user-generated content platforms could encourage users from underrepresented groups to make more and larger contributions. Wikipedia could, for instance, embolden women and non-English speakers to set up new articles or to extend existing ones. Similarly, user review websites like Yelp or TripAdvisor could use vouchers or discounts to increase the incentive to contribute of users from underrepresented groups.

Our empirical approach to detect coverage bias on user-generated content platforms is limited in two ways. First, while we show that marketing aspirations are one likely driver of unbalanced coverage, we cannot fully rule out other potential causes. That is, we cannot determine precisely whether and to which extent the coverage bias is driven by user preferences, characteristics or ulterior marketing aspirations. Regarding the latter, we are also unable to distinguish between demand and supply side as driving force behind this channel. Thus, in terms of our application, we cannot say with certainty *why* there exists a coverage bias against MPs from the center-left in Germany and France.³⁹ The SPD in Germany has fewer potential voters than the CDU/CSU; hence, there may be relatively less demand for Wikipedia biographies of MPs from the SPD, such that partisan contributions have a comparatively smaller payoff. Although voters of CDU/CSU and SPD are equally Internet affine (Forschungsgruppe Wahlen, 2014), potential voters of the SPD may also perceive the new media as a less relevant information source. On the other hand, it is possible that the differences in partisan activity reflect the parties' perceptions of how important an extensive Internet presence is and to which extent they are aware of the potentials of user-generated content. For instance, Peter Tauber, who was secretary general during our observation period, provides a social media compendium that also points to the importance of Wikipedia (Tauber, 2013, p.12), while nothing comparable exists for the SPD – in other words, the CDU/CSU may be more successful in their ulterior marketing activities.

Second, the approach is not applicable to unilingual platforms. We consider this to be a minor disadvantage, though. Many relevant user-generated content platforms such as Wikipedia, YouTube, TripAdvisor, Twitter, Yelp, and Facebook are available in several languages; notable exceptions are the English-only Reddit⁴⁰ and the Chinese-only Sina Weibo.⁴¹ Similarly, different language versions must not be mere translations of each other, and one language version must be expected to be less biased.

One further limitation of our paper is specific to our application to Wikipedia. To disentangle the effect of party affiliation on coverage from the effect that MP characteristics

³⁸Although Wikipedia's author guidelines stress that any content must be written from a "neutral point of view" (see https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view (June 2020)), which includes assigning proper weights to topics, the average user is unlikely to be aware of the issue and its pitfalls.

³⁹We contacted the German parties' press offices to inquire whether there are coordinated party activities in Wikipedia. According to all replies, there exist no official guidelines for the handling of Wikipedia; every MP is responsible herself for her Wikipedia biography.

⁴⁰See [reddit.com](https://www.reddit.com) (Sep 2020).

⁴¹See [weibo.com](https://www.weibo.com) (Sep 2020).

may have, we focus on MPs from just one election period, which raises concerns about the generalizability of our results. The external validity of our study is, however, supported by similar patterns of Wikipedia coverage of members of the 16 German State Parliaments and of the European Parliament.⁴² Biographies of CDU/CSU affiliates in a State Parliament are on average about a quarter page (or about 0.15 standard deviations) longer than biographies of SPD affiliates. Similarly, biographies of CDU/CSU affiliates in the European Parliament are on average about half a page (or about 0.25 standard deviations) longer than biographies of SPD affiliates. These numbers suggest that our main results reflect a general pattern, rather than being specific to the sample at hand. Interestingly, when we compare the Wikipedia coverage of judges in the German Constitutional Court – who are usually nominated by a particular party – we do not find differences in coverage. These judges are, however, elected for a lifetime and not by the public, and thus have no incentive to amplify their Wikipedia biographies. These observations therefore also fit our hypothesis about partisan contributions as the main driver of our results.

The implications and limitations of our analysis highlight several avenues to future research. First, it would be interesting to study the effects of different remedies against coverage bias. It is, for instance, unclear by which means underrepresented user groups can best be encouraged to contribute. Moreover, since the literature on fact-checking measures documents that factual information has ambiguous effects on individuals' misperceptions (Jerit and Zhao, 2020), warning templates for articles that cover sensitive topics need to be well-designed to be beneficial rather than harmful. Second, a comprehensive analysis of the channels that ultimately lead to unbalanced coverage on user-generated content platforms promises valuable insights. If these channels were better understood and distinguishable from each other, more concerted remedies against coverage bias would be feasible.

⁴²The data was collected on March 29, 2017.

A. Survey evidence

To investigate whether voters perceive an extensive Wikipedia biography as a positive signal, we conducted a classroom survey at the University of Cologne and a representative online survey. The goal was to study if participants rate unknown MPs with a longer Wikipedia biography better in terms of their valence characteristics, i.e., qualities of a politician on which all voters agree (Stokes, 1963). We used nine valence characteristics that are frequently discussed in the political science literature (e.g., Kinder et al., 1980; Funk, 1999; Stone and Simas, 2010).

We randomized participants into two groups. Participants in group 1 received the instruction: “Consider a politician from your preferred party. The Wikipedia biography of this politician (politician A) is three pages long. Consider another politician from the same party. The Wikipedia biography of this politician is one page long. Please answer the following questions.” Participants in group 2 received the same instructions, only that the biography of politician A was one, and of politician B was three pages long. We did not show actual (fake) biographies to the participants, because adding text – that inevitably provides further information on the politician – would prevent us from disentangling the effect of biography length as such from the effect of providing a particular piece of additional information.

Next, we asked the participants which politician would probably score better with respect to each of the nine valence characteristics. They could either reply “Politician A”, “Politician B”, or “Don’t know.” We considered participants’ replies if they answered all nine questions.

A.1. Classroom survey

The classroom survey was conducted among sixty undergraduate students of economics at the University of Cologne.⁴³ Table 10 shows the results. Columns 1 to 3 show the shares of students who opted for the politician with the three-page biography (s_3), the politician with the one-page biography (s_1), and “Don’t know”, respectively. Column 4 displays the difference between s_3 and s_1 , which is positive for all valence characteristics except for intelligence and honesty. We test the statistical significance of this difference against the null hypothesis that $s_3 = s_1$ by means of a likelihood ratio test as in Giordan and Diana (2008) and find that the difference between s_3 and s_1 is statistically significant for knowledge, strength as a public servant, and inspiring ($p < 0.01$); and weakly statistically significant for empathy ($p < 0.1$)

A.2. Representative survey

To explore whether the results from the classroom survey are specific to the particular age group under consideration, we repeated the survey among a representative sample of

⁴³The survey was carried out via classEx, a free software for interactive classroom experiments (Giamattei and Lambsdorff, 2019).

participants ($N = 500$).⁴⁴ Columns 5 to 20 of Table 10 show the results. To compare the results to the classroom survey while still maintaining a reasonable sample size, we partitioned the sample into terciles based on the participants' age in Columns 5 to 16; Columns 17 to 20 show the results for the entire sample.

In line with the results from Section A.1, we find that the difference between s_3 and s_1 for respondents younger than 37 years is positive and statistically significant for knowledge, strength as a public servant, inspiring, and empathy ($p < 0.05$, Column 8). This does not hold for respondents older than 37, though. In contrast to Columns 4 and 8, the difference between s_3 and s_1 is negative for many valence characteristics in Columns 12 and 16, which means that older people do not perceive a relatively longer biography as a positive signal.

To sum up, we find that younger subjects perceive a longer Wikipedia biography as a positive signal about the politician's valence characteristics, whereas it makes older subjects more critical. In his widely cited paper, Converse (1969) argues that voters' party identification becomes stronger over time; Falter (2010) provides empirical evidence from Germany (p.14, Figure 15). Thus, young individuals are more likely to be swing voters and thereby comprise the relevant target group for (partisan) coverage bias on Wikipedia. Moreover, young individuals are far more likely to use the Internet in general and in particular to obtain political information.⁴⁵

B. Robustness checks

In our main analysis, we excluded 35 MPs in distinguished positions (ministers and party heads) and nine further MPs who had already left the Bundestag. Table 11 shows that including these observations does not affect our results. Moreover, the estimates that were added – dummies for early resignations and distinguished offices – are very large and statistically significant, legitimizing the presumption that they are not comparable to other MPs in the sample.

In addition to that, we prove the robustness of the results in Table 6 by taking into account that English texts are about a fourth to a fifth shorter than German texts.⁴⁶ To this end, we scale the MPs' German biography length in equation (3) with the factors 0.6, 0.75, 0.8, and 0.9, respectively. Moreover, we use the difference in logs of the biography lengths as a dependent variable. Table 11 shows that our results are qualitatively unaffected.

⁴⁴The survey was conducted online by respondi in December 2020. The sample of participants was representative for the German population between 18 and 74 years.

⁴⁵See <https://de.statista.com/infografik/20068/informationsquellen-fuer-politische-nachrichten-bei-jungen-und-alten/> (Dec 2020), and <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/IT-Nutzung/Tabellen/durchschnittl-nutzung-alter-ikt.html> (Dec 2020).

⁴⁶See, e.g., www.orbis-uebersetzungen.de (Feb 2016).

C. Party affiliation and English biography length

In Section 4, we argue that it is plausible to assume that there are no effects of party affiliation on the English biography length, because partisan contributors have no incentive to contribute to the English Wikipedia. In this section, we perform four additional plausibility checks. First, only thirteen of 598 MPs in our dataset provide more than a short CV in English on their personal homepage, suggesting that they do not consider an English web presence as important. Second, only eight of 138 English biographies (5.8%) were edited from the Bundestag building; with one exception, only small changes were undertaken. Third, the lion’s share of the English biographies is not translated from their German counterparts. Translated articles have to be marked by a translation template on the article’s talk page and by a link to the source article; only ten out of 138 English biographies are marked like this, and no biography is translated from a foreign language into German. In addition, Wikipedia advises against one-to-one translation.⁴⁷ Finally, while the assumption of no party effects may fail for foreign languages that are spoken in countries adjacent to Germany or by large minorities, Germany has no direct border with any English speaking country, and a low number of immigrants whose native language is English (Statistisches Bundesamt, 2017).

⁴⁷See <https://en.wikipedia.org/wiki/Wikipedia:Translation> (Dec 2018).

References

- Aday, Sean (2010) “Leading the charge: Media, elites, and the use of emotion in stimulating rally effects in wartime,” *Journal of Communication*, **60** (3), 440–465.
- Anderson, Michael and Jeremy Magruder (2012) “Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database,” *The Economic Journal*, **122** (563), 957–989.
- Barrett, Andrew W and Jeffrey S Peake (2007) “When the president comes to town: Examining local newspaper coverage of domestic presidential travel,” *American Politics Research*, **35** (1), 3–31.
- Bayerischer Rundfunk, (2017) “Wikipedia-Artikel über Abgeordnete vom Bundestag aus geschönt,” *URL: <https://www.br.de/bundestagswahl/wikipedia-artikel-abgeordnete-bundestag-100.html>*.
- Becker, Johannes, Andreas Peichl, and Johannes Rincke (2009) “Politicians’ outside earnings and electoral competition,” *Public Choice*, **140** (3-4), 379.
- Bond, Robert, Christopher Fariss, Jason J. Jones, Adam I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler (2012) “A 61-million-person experiment in social influence and political mobilization,” *Nature*, **489** (7415), 295–298.
- Brandenburg, Heinz (2005) “Political bias in the Irish media: A quantitative study of campaign coverage during the 2002 general election,” *Irish Political Studies*, **20** (3), 297–322.
- Briones, Rowena, Xiaoli Nan, Kelly Madden, and Leah Waks (2012) “When vaccines go viral: an analysis of HPV vaccine coverage on YouTube,” *Health Communication*, **27** (5), 478–485.
- Brown, Adam R (2011) “Wikipedia as a data source for political scientists: Accuracy and completeness of coverage,” *PS: Political Science & Politics*, **44** (2), 339–343.
- Bruce, Rebecca F and Janyce M Wiebe (1999) “Recognizing subjectivity: a case study in manual tagging,” *Natural Language Engineering*, **5** (2), 187–205.
- Burden, Barry C (2002) “When bad press is good news: The surprising benefits of negative campaign coverage,” *Harvard international journal of press/politics*, **7** (3), 76–89.
- Butler, Daniel M and Emily Schofield (2010) “Were newspapers more interested in pro-Obama letters to the editor in 2008? Evidence from a field experiment,” *American Politics Research*, **38** (2), 356–371.
- Callahan, Ewa S and Susan C Herring (2011) “Cultural bias in Wikipedia content on famous persons,” *Journal of the American society for information science and technology*, **62** (10), 1899–1915.

- Campante, Filipe, Ruben Durante, and Francesco Sobbrío (2018) “Politics 2.0: The multifaceted effect of broadband internet on political participation,” *Journal of the European Economic Association*, **16** (4), 1094–1136.
- Chen, Yan, F Maxwell Harper, Joseph Konstan, and Sherry Xin Li (2010) “Social comparisons and contributions to online communities: A field experiment on movielens,” *American Economic Review*, **100** (4), 1358–98.
- Chevalier, Judith and Dina Mayzlin (2006) “The effect of word of mouth on sales: Online book reviews,” *Journal of Marketing Research*, **43** (3), 345–354.
- Clemons, Eric K, Guodong Gordon Gao, and Lorin M Hitt (2006) “When online reviews meet hyperdifferentiation: A study of the craft beer industry,” *Journal of Management Information Systems*, **23** (2), 149–171.
- Converse, Philip E (1969) “Of time and partisan stability,” *Comparative Political Studies*, **2** (2), 139–171.
- D’Alessio, Dave and Mike Allen (2000) “Media bias in presidential elections: A meta-analysis,” *Journal of Communication*, **50** (4), 133–156.
- Dellarocas, Chrysanthos (2006) “Strategic manipulation of internet opinion forums: Implications for consumers and firms,” *Management Science*, **52** (10), 1577–1593.
- Dertwinkel-Kalt, Markus, Anna Kerkhof, and Johannes Münster (2019) “Incumbency Dominance in Letters to the Editor: Field Experimental Evidence,” *Political Communication*, **36** (3), 337–356.
- Durante, Ruben and Ekaterina Zhuravskaya (2018) “Attack when the world is not watching? US news and the Israeli-Palestinian conflict,” *Journal of Political Economy*, **126** (3), 1085–1133.
- Eisensee, Thomas and David Strömberg (2007) “News droughts, news floods, and US disaster relief,” *The Quarterly Journal of Economics*, **122** (2), 693–728.
- Enikolopov, Ruben, Alexey Makarin, and Maria Petrova (2020) “Social media and protest participation: Evidence from Russia,” *Econometrica*, **88** (4), 1479–1514.
- Falck, Oliver, Robert Gold, and Stephan Heblich (2014) “E-lelections: Voting Behavior and the Internet,” *American Economic Review*, **104** (7), 2238–65.
- Falter, Jürgen W (2010) “Demographischer Wandel und Wahlverhalten,” Working Paper. Forschungsgruppe Wahlen, (2014) “Internet Strukturdaten 2014,” *URL: <http://www.forschungsgruppe.de/Umfragen/Internet-Strukturdaten/>*.

- Freeman, Becky and Simon Chapman (2007) “Is “YouTube” telling or selling you something? Tobacco content on the YouTube video-sharing website,” *Tobacco Control*, **16** (3), 207–210.
- Freitag, Julian, Anna Kerkhof, and Johannes Münster (2020) “Selective Sharing of News Items and the Political Position of News Outlets,” Working Paper.
- Funk, Carolyn L (1999) “Bringing the candidate into models of candidate evaluation,” *The Journal of Politics*, **61** (3), 700–720.
- Galvis, Ángela Fonseca, James M Snyder Jr, and BK Song (2016) “Newspaper market structure and behavior: Partisan coverage of political scandals in the United States from 1870 to 1910,” *The Journal of Politics*, **78** (2), 368–381.
- Garz, Marcel (2014) “Good news and bad news: evidence of media bias in unemployment reports,” *Public Choice*, **161** (3-4), 499–515.
- Gavazza, Alessandro, Mattia Nardotto, and Tommaso Valletti (2019) “Internet and politics: Evidence from uk local elections and local government policies,” *The Review of Economic Studies*, **86** (5), 2092–2135.
- Gentzkow, Matthew, Jesse M Shapiro, and Daniel F Stone (2016) “Media bias in the marketplace: Theory,” in *Handbook of media economics*, 1: Elsevier, 623–645.
- Giamattei, Marcus and Johann Graf Lambsdorff (2019) “classEx—an online tool for lab-in-the-field experiments with smartphones,” *Journal of Behavioral and Experimental Finance*, **22**, 223–231.
- Giordan, Marco and Giancarlo Diana (2008) “Tests for the equality of multinomial parameters,” *Journal of Statistical Planning and Inference*, **138** (12), 3789–3796.
- Goldfarb, Avi and Catherine Tucker (2019) “Digital economics,” *Journal of Economic Literature*, **57** (1), 3–43.
- Greene, William H (2003) *Econometric analysis*: Pearson Education India.
- Greenstein, Shane and Feng Zhu (2012) “Is Wikipedia Biased?” *American Economic Review*, **102** (3), 343–48.
- Groeling, Tim (2008) “Who’s the fairest of them all? An empirical test for partisan bias on ABC, CBS, NBC, and Fox News,” *Presidential Studies Quarterly*, **38** (4), 631–657.
- (2013) “Media bias by the numbers: Challenges and opportunities in the empirical study of partisan news,” *Annual Review of Political Science*, **16**.
- Groeling, Tim and Matthew A Baum (2008) “Crossing the water’s edge: Elite rhetoric, media coverage, and the rally-round-the-flag phenomenon,” *The Journal of Politics*, **70** (4), 1065–1085.

- Halavais, Alexander and Derek Lackaff (2008) “An analysis of topical coverage of Wikipedia,” *Journal of computer-mediated communication*, **13** (2), 429–440.
- Hallin, Daniel C and Paolo Mancini (2004) *Comparing media systems: Three models of media and politics*: Cambridge university press.
- Heinz, Matthias and Johan Swinnen (2015) “Media slant in economic news: A factor 20,” *Economics Letters*, **132**, 18–20.
- Hinnosaar, Marit (2019) “Gender inequality in new media: Evidence from Wikipedia,” *Journal of Economic Behavior & Organization*, **163**, 262–276.
- Hu, Nan, Ling Liu, and Jie Jennifer Zhang (2008) “Do online reviews affect product sales? The role of reviewer characteristics and temporal effects,” *Information Technology and Management*, **9** (3), 201–214.
- Jerit, Jennifer and Yangzi Zhao (2020) “Political misinformation,” *Annual Review of Political Science*, **23**, 77–94.
- Kinder, Donald R, Mark D Peters, Robert P Abelson, and Susan T Fiske (1980) “Presidential prototypes,” *Political Behavior*, **2** (4), 315–337.
- Kürschner (2015) *Kürschners Volkshandbuch Deutscher Bundestag 18. Wahlperiode*: Neue Darmstädter Verlagsanstalt.
- Larcinese, Valentino, Riccardo Puglisi, and James M Snyder Jr (2011) “Partisan bias in economic news: Evidence on the agenda-setting behavior of US newspapers,” *Journal of Public Economics*, **95** (9-10), 1178–1189.
- Liberini, Federica, Antonio Russo, Ángel Cuevas, and Ruben Cuevas (2020) “Politics in the Facebook Era-Evidence from the 2016 US Presidential Elections.”
- Lott, John R and Kevin A Hassett (2014) “Is newspaper coverage of economic events politically biased?” *Public Choice*, **160** (1-2), 65–108.
- Lu, Xianghua, Sulin Ba, Lihua Huang, and Yue Feng (2013) “Promotional marketing or word-of-mouth? Evidence from online restaurant reviews,” *Information Systems Research*, **24** (3), 596–612.
- Luca, Michael (2016) “User-generated content and social media,” in *Handbook of Media Economics*, **1**: Elsevier, 563–592.
- Luca, Michael and Georgios Zervas (2016) “Fake it till you make it: Reputation, competition, and Yelp review fraud,” *Management Science*, **62** (12), 3412–3427.
- Mayzlin, Dina (2006) “Promotional chat on the Internet,” *Marketing Science*, **25** (2), 155–163.

- Mayzlin, Dina, Yaniv Dover, and Judith Chevalier (2014) “Promotional Reviews: An Empirical Investigation of Online Review Manipulation,” *American Economic Review*, **104** (8), 2421–2455.
- McCombs, Maxwell E and Donald L Shaw (1972) “The agenda-setting function of mass media,” *Public Opinion Quarterly*, **36** (2), 176–187.
- Melnik, Mikhail I and James Alm (2002) “Does a seller’s ecommerce reputation matter? Evidence from eBay auctions,” *The Journal of Industrial Economics*, **50** (3), 337–349.
- Mesgari, Mostafa, Chitu Okoli, Mohamad Mehdi, Finn Arup Nielsen, and Arto Lanamaki (2015) “The sum of all human knowledge: A systematic review of scholarly research on the content of Wikipedia,” *Journal of the Association for Information Science and Technology*, **66** (2), 219–245.
- Morris, Jonathan S and Peter L Francia (2010) “Cable news, public opinion, and the 2004 party conventions,” *Political Research Quarterly*, **63** (4), 834–849.
- Niven, David (2003) “Objective evidence on media bias: Newspaper coverage of congressional party switchers,” *Journalism & Mass Communication Quarterly*, **80** (2), 311–326.
- Pang, Bo and Lillian Lee (2008) “Opinion mining and sentiment analysis,” *Foundations and Trends® in Information Retrieval*, **2** (1–2), 1–135.
- Persson, Torsten and Guido Enrico Tabellini (2005) *The economic effects of constitutions*: MIT press.
- Puglisi, Riccardo (2011) “Being the New York Times: the political behaviour of a newspaper,” *The BE journal of economic analysis & policy*, **11** (1).
- Puglisi, Riccardo and James M Snyder (2011) “Newspaper coverage of political scandals,” *The Journal of Politics*, **73** (3), 931–950.
- (2016) “Empirical studies of media bias,” in *Handbook of media economics*, **1**: Elsevier, 647–667.
- Reagle, Joseph and Lauren Rhue (2011) “Gender bias in Wikipedia and Britannica,” *International Journal of Communication*, **5**, 21.
- Rosenzweig, Roy (2006) “Can history be open source? Wikipedia and the future of the past,” *The Journal of American History*, **93** (1), 117–146.
- Schiffer, Adam J (2006) “Assessing partisan bias in political news: The case (s) of local senate election coverage,” *Political Communication*, **23** (1), 23–39.
- Soroka, Stuart N (2012) “The gatekeeping function: Distributions of information in media and the real world,” *The Journal of Politics*, **74** (2), 514–528.

- Statistisches Bundesamt, (2017) “Ausländische Bevölkerung nach Geschlecht und ausgewählten Staatsangehörigkeiten,” URL:<http://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/MigrationIntegration/Au>
- Stokes, Donald E (1963) “Spatial models of party competition,” *American Political Science Review*, **57** (2), 368–377.
- Stone, Walter J and Elizabeth N Simas (2010) “Candidate valence and ideological positions in US House elections,” *American Journal of Political Science*, **54** (2), 371–388.
- Strömberg, David (2004) “Radio’s impact on public spending,” *The Quarterly Journal of Economics*, **119** (1), 189–221.
- Tauber, Peter (2013) “Social Media Leitfaden 2013,” URL:<http://www.petertauber.de/wp-content/uploads/2013/04/SocialMedia-Leitfaden-Peter-Tauber-2013.pdf>.
- Vos, Debby (2014) “Which politicians pass the news gates and why? Explaining inconsistencies in research on news coverage of individual politicians,” *International Journal of Communication*, **8**, 2438–2461.
- Wang, Zhongmin (2010) “Anonymity, social image, and the competition for volunteers: a case study of the online market for reviews,” *The BE Journal of Economic Analysis & Policy*, **10** (1), 1–35.
- Weaver, David, Maxwell McCombs, and Donald L Shaw (2004) “Agenda-setting research: Issues, attributes, and influences,” *Handbook of Political Communication Research*, **257**.
- Wiebe, Janyce, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin (2004) “Learning subjective language,” *Computational linguistics*, **30** (3), 277–308.
- Zhang, Xiaoquan (Michael) and Feng Zhu (2011) “Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia,” *American Economic Review*, **101** (4), 1601–1615.
- Zhu, Feng and Xiaoquan Zhang (2010) “Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics,” *Journal of Marketing*, **74** (2), 133–148.
- Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov (2020) “Political effects of the internet and social media,” *Annual Review of Economics*, **12**.

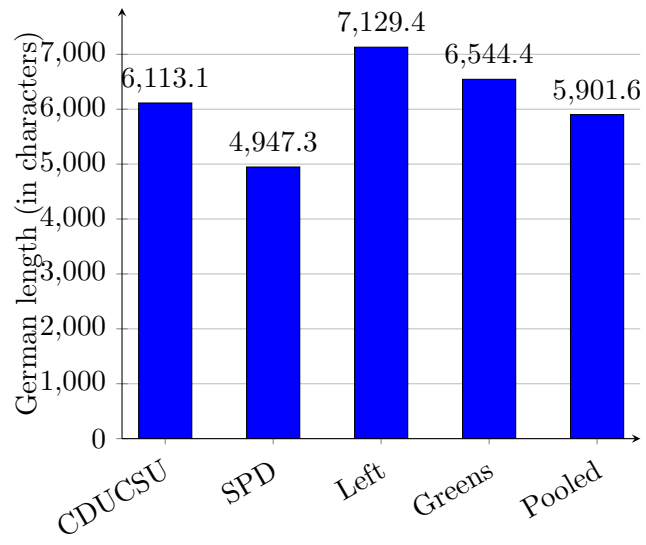


Figure 1: German biography length across German parties in characters

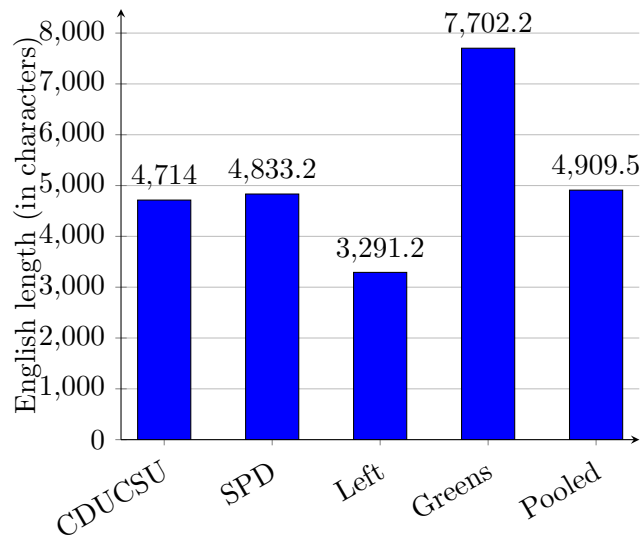


Figure 2: English biography length across German parties in characters

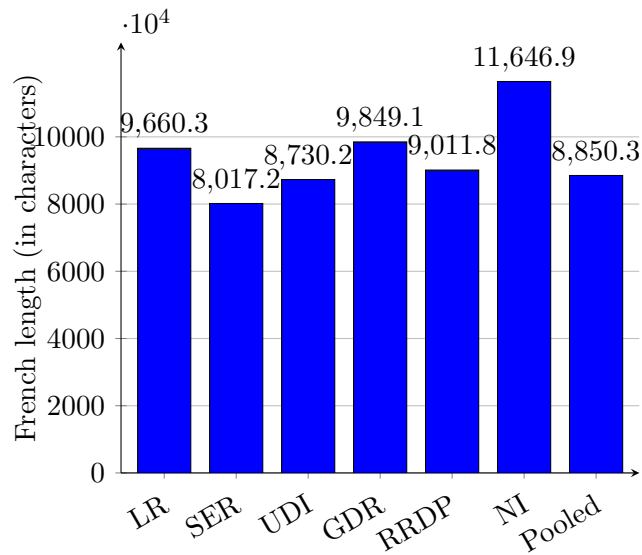


Figure 3: French biography length across French parties in characters

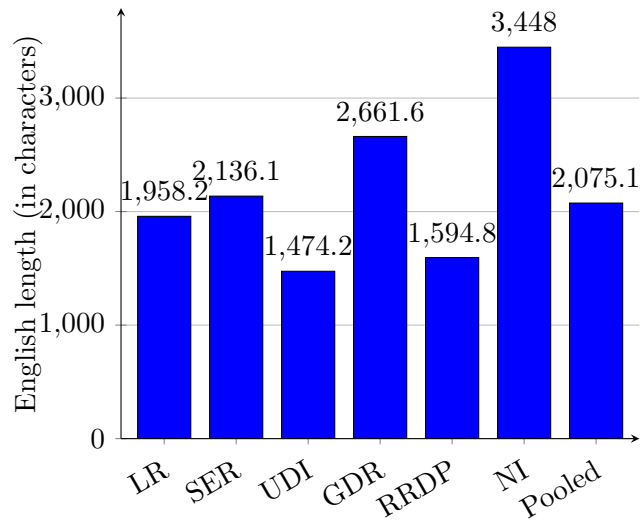


Figure 4: English biography length across French parties in characters

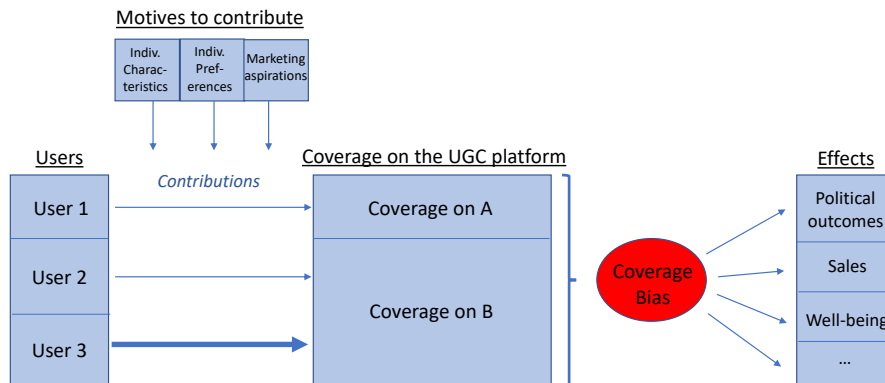


Figure 5: Theoretical considerations of coverage bias

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
<i>German MPs</i>					
German biography length (in characters)	5901.577	4890.87	1118	45316	598
English biography length of German MPs (in characters)	4909.507	4119.552	838	28132	138
CDU/CSU	0.492	0.5	0	1	598
SPD	0.308	0.462	0	1	598
Greens	0.097	0.296	0	1	598
Left	0.104	0.305	0	1	598
Female	0.366	0.482	0	1	598
Former periods in Bundestag	1.577	1.748	0	9	598
Ancillary income (in 1000 Euros)	25.609	106.476	0	1411	596
Directly elected	0.453	0.498	0	1	598
Population density	910.613	1471.26	38	12842.9	271
Fraction population aged 18 to 35	20.114	2.7	15.6	28.9	271
Fraction population with Abitur	36.417	7.473	21.7	59.7	271
Number of international offices	0.209	0.496	0	3	598
Former periods in European Parliament	0.012	0.135	0	2	598
Number of adjectives	65.502	61.972	8	618	598
Number of images	2.475	1.198	0	11	598
Number of categories	9.206	2.446	5	24	598
Number of weblinks	1.487	0.669	0	4	598
Number of unique authors	52.05	59.534	5	580	598
Edited from Bundestag network (German)	0.527	0.5	0	1	598
Number of edits from Bundestag network (German)	4.427	5.399	1	40	309
Characters added from Bundestag network (German)	533.176	919.954	0	7201	306
Characters deleted from Bundestag network (German)	198.373	648.166	0	6744	306
Net character change from Bundestag network (German)	171.321	801.527	-6743	7201	598
Edited from Bundestag network (English)	0.058	0.235	0	1	138
Length (talk pages)	5206.723	13030.45	7	130746	292
Criticizing sentences	0.281	1.041	0	10	598
Number of words	547.518	469.732	76	4326	598
English homepage	0.022	0.146	0	1	598
Translation template	0.017	0.128	0	1	598
<i>French MPs</i>					
French biography length (in characters)	8850.333	4310.317	35	20987	504
English biography length of French MPs (in characters)	2075.115	1586.235	877	11355	296
Les Républicains (LR)	0.341	0.475	0	1	504
Groupe Socialiste (SER)	0.518	0.5	0	1	504
Groupe UDI et indépendants (UDI)	0.042	0.2	0	1	504
Groupe radical, républicain, démocrate et progressiste (RRDP)	0.032	0.175	0	1	504
Groupe de la Gauche démocrate et républicaine (GDR)	0.03	0.17	0	1	504
Non-inscrit (NI)	0.034	0.181	0	1	504
Number of images	6.238	2.18	0	32	504
Number of weblinks	9.813	7.628	0	38	504
Number of categories	21.244	4.894	0	60	504

Notes: This table presents the summary statistics of all variables used in the analysis. The variables *CDU/CSU*, *SPD*, *Greens*, *Left*, *Female*, *Directly elected*, *Edited from Bundestag network (German)*, *Edited from Bundestag network (English)*, *English homepage*, *Translation template*, *Les Républicains*, *Groupe Socialiste*, *Groupe UDI et indépendants*, *Groupe radical, républicain, démocrate et progressiste*, *Groupe de la Gauche démocrate et républicaine*, and *Non-inscrit* are dummy variables. The variables *Population density*, *Fraction population aged 18 to 35*, and *Fraction population with Abitur* are available only for a subset of MPs who are directly elected. The variables *English biography length of German (French) MPs* and *Edited from Bundestag network (English)* are available only for MPs for whom an English Wikipedia biography exists.

Table 2: Literature on measuring media bias

Study	Context	Manifestation	Empirical approach
D'Alession & Allen (2000)	Survey		
Groeling (2013)	Survey		
Puglisi & Snyder (2015)	Survey		
Gentzkow et al. (2015)	Survey		
McCombs and Shaw (1972)	Newspapers, television, magazines	Coverage bias	
Schiffer (2006)	Newspapers	Coverage bias	Control for observable differences between observations
Niven (2003)	Newspapers	Coverage bias	Comparing the media coverage of homogeneous cases within the same media outlet
Puglisi (2011)	Newspapers	Coverage bias	Comparing the media coverage of homogeneous cases within the same media outlet
Brandenburg (2005)	Newspapers	Coverage bias, statement bias	Comparing the media coverage of homogeneous cases within the same media outlet
Morris & Francia (2010)	Television	Coverage bias, statement bias	Comparing the media coverage of homogeneous cases within the same media outlet
Barrett & Peake (2007)	Newspapers	Coverage bias	Comparing the coverage of the same topic in different media outlets
Baum & Groeling (2008)	Online news	Coverage bias	Comparing the coverage of the same topic in different media outlets, study an observable population
Larcinese et al. (2011)	Newspapers	Coverage bias	Comparing the coverage of the same topic in different media outlets
Puglisi & Snyder (2011)	Newspapers	Coverage bias	Comparing the coverage of the same topic in different media outlets
Lott & Hassett (2014)	Newspapers	Coverage bias	Comparing the coverage of the same topic in different media outlets
Fonseca-Galvis et al. (2016)	Newspapers	Coverage bias	Comparing the coverage of the same topic in different media outlets
Groeling (2008)	Television	Coverage bias	Study an observable population
Aday (2010)	Television	Coverage bias	Study an observable population
Soroka (2012)	Newspapers	Coverage bias	Study an observable population
Heinz & Swinnen (2015)	Newspapers	Coverage bias	Study an observable population
Garz (2014)	Newspapers	Coverage bias	Study an observable population
Butler & Schofield (2010)	Newspapers, letters to the editor	Coverage bias	Create an observable population
Dertwinkel-Kalt et al. (2019)	Newspapers, letters to the editor	Coverage bias	Create an observable population
Groseclose & Milyo (2005)	Newspapers, television	Statement bias	
Gentzkow & Shapiro (2010)	Newspapers	Statement bias	
Puglisi & Snyder (2015)	Newspapers	Statement bias	
Ho & Quinn (2008)	Newspapers	Statement bias	
Greenstein & Zhu (2012)	Wikipedia	Statement bias	
Greenstein & Zhu (2018)	Wikipedia	Statement bias	
This paper	User-generated content	Coverage bias	Compare the coverage of the same topic across two language versions of the same platform

Notes: Table 2 gives an overview of the empirical literature on measuring media bias.

Table 3: Literature on user-generated content

Study	Context	Platform	Subject of interest
Luca (2015)	Survey		
Mesgari et al. (2015)	Survey	Wikipedia	
Mayzlin (2006)	Consumer reviews, chat	Theoretic model	
Dellarocas (2006)	Consumer reviews	Theoretic model	
Mayzlin et al. (2014)	Consumer reviews	Expedia, TripAdvisor	Detection of promotional reviews
Luca & Zervas (2016)	Consumer reviews	Yelp	Detection of promotional reviews
Huang & Chen (2006)	Consumer reviews		Effect of promotional reviews
Melnik & Alm (2002)	Consumer reviews	eBay	Effect of promotional reviews
Clemons et al. (2006)	Consumer reviews	ratebeer.com	Effect of promotional reviews
Chevalier & Mayzlin (2006)	Consumer reviews	Amazon, Barnesandnoble.com	Effect of promotional reviews
Anderson & Magruder (2012)	Consumer reviews	Yelp	Effect of promotional reviews
Lu et al. (2013)	Consumer reviews	ABC.com (restaurant review site in China)	Effect of promotional reviews
Wang (2010)	Consumer reviews	Yelp, Citysearch, Yahoo Local	Motives to contribute
Chen et al. (2010)	Consumer reviews	MovieLens	Motives to contribute
Zhang & Zhu (2011)	Articles	Wikipedia	Motives to contribute
Brown (2011)	Articles	Wikipedia	Wikipedia quality
Halavais & Lackaff (2008)	Articles	Wikipedia	Wikipedia contributors
Hinnosaar (2019)	Articles	Wikipedia	Wikipedia contributors
Rosenzweig (2006)	Articles	Wikipedia	Wikipedia contributors
Callahan & Herring (2011)	Articles	Wikipedia	Wikipedia coverage
This paper	MP biographies	Wikipedia	Detection of coverage bias

Notes: Table 3 gives an overview of the literature on user-generated content.

Table 4: Literature on social media and political outcomes

Study	Context	Outcome of interest
Goldfarb & Tucker (2019)	Survey	
Zhuravskaya et al. (2020)	Survey	
Falck et al. (2014)	Broadband internet diffusion	election outcomes
Gavazza et al. (2019)	Broadband internet diffusion	election outcomes
Campante et al. (2018)	Broadband internet diffusion	political participation
Bond et al. (2012)	Facebook	individual voting behavior
Liberini et al. (2020)	Facebook	individual voting behavior
Aral & Walker (2012)	Facebook	product adoption
Enikolopov et al. (2020)	VK	political participation
This paper	Wikipedia	coverage bias

Notes: Table 4 gives an overview of the literature on social media and political outcomes.

Table 5: Basic results

	(1)	(2)	(3)	(4)	(5)
	Length	Length	Length	Length	Length
CDU/CSU	1165.8*** (374.1)	1057.0*** (385.9)	971.3*** (366.6)	870.4** (353.1)	1614.8** (738.7)
Left	2182.2*** (728.7)	2297.7*** (730.6)	2388.3*** (694.6)	2396.4*** (656.1)	
Greens	1597.2 (972.0)	1692.7* (985.2)	1736.4* (906.8)	1733.5* (917.7)	
Female		-689.1 (435.6)	-474.8 (403.9)	-216.0 (402.0)	299.0 (749.8)
Former periods in BT			869.6*** (161.5)	875.5*** (160.2)	723.1*** (197.6)
Ancillary Income				2.002 (1.702)	1.681 (1.621)
PhD				1915.3*** (608.9)	886.2 (868.8)
Population density					0.896*** (0.278)
Fraction pop. 18–35					255.2 (155.5)
Fraction pop. with Abitur					-1.997 (48.21)
Constant	4947.3*** (232.7)	5231.9*** (305.5)	3810.6*** (347.0)	3327.6*** (373.3)	-2578.2 (3436.9)
N	598	598	598	596	266
R^2	0.021	0.025	0.121	0.150	0.185

Notes: Robust standard errors in parentheses. The dependent variable $length_i^G$ measures Wikipedia coverage of MP i in terms of her biography length in characters. CDU/CSU_i , $Left_i$, and $Greens_i$ are dummy variables equal to 1 if MP i is affiliated to that party; SPD is the omitted category. $Female_i$ is equal to 1 if MP i is a woman. $Former\ periods\ in\ BT_i$ counts the election terms that MP i has been in parliament. $Ancillary\ income_i$ is the mean ancillary income of MP i during the 18th election term in 1,000 Euros based on the estimation of `abgeordnetenwatch.de`. PhD_i is equal to 1 if MP i has a PhD. $Population\ density_i$, $Fraction\ pop.\ 18 - 35_i$, and $Fraction\ pop.\ with\ Abitur_i$ refer to i 's constituency demography. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Difference-in-difference results

	(1)	(2)	(3)	(4)	(5)
	OLS	Two-step	MEM Two-step	ML	MEM ML
CDU/CSU	2230.8 (1378.3)	3165.4* (1749.1)	2275.5	6010.6*** (1720.4)	2963.0
Left	6447.0*** (1627.8)	7892.4*** (2489.3)	6488.2	11842.6*** (2365.9)	6843.8
Greens	2901.8 (2805.8)	4365.4* (2510.7)	3183.5	7941.9*** (2432.1)	3385.0
Female	-1237.8 (1265.4)	-1409.9 (1284.4)	-1184.7	-1443.3 (1419.8)	-356.8
Former periods in BT	670.0** (332.0)	1030.4** (472.2)	577.0	1902.7*** (370.7)	187.7
PhD	2509.2* (1477.9)	3180.6** (1518.9)	2536.5	3845.3** (1599.4)	849.9
Ancillary Income	10.18 (9.853)	9.837 (8.039)	11.32	1.398 (7.474)	5.105
Constant	-245.1 (1607.3)	-5430.1 (5347.4)		-18997.1*** (2375.3)	
Mill's Lambda		2837.6 (2792.9)		10558.6*** (906.1)	
N	136	596	596	596	596
R^2	0.124				

Notes: Robust standard errors in parentheses. The dependent variable in columns (1), (2), and (4) is $length_i^G - length_i^E$ that measures the differences in Wikipedia coverage of MP i in terms of her biography length in characters. CDU/CSU_i , $Left_i$, and $Greens_i$ are dummy variables equal to 1 if MP i is affiliated to that party; SPD is the omitted category. $Female_i$ is equal to 1 if MP i is a woman. $Former\ periods\ in\ BT_i$ counts the election terms that MP i has been in parliament. $Ancillary\ income_i$ is the mean ancillary income of MP i during the 18th election term in 1,000 Euros based on the estimation of abgeordnetenwatch.de. PhD_i is equal to 1 if MP i has a PhD. $Population\ density_i$, $Fraction\ pop.\ 18-35_i$, and $Fraction\ pop.\ with\ Abitur_i$ refer to i 's constituency demography. The MEMs in columns (3) and (5) show the coefficients' marginal effect at the mean, i.e., the change in the difference in biography length, given that it is observed, and holding all other factors at their mean. MEMs for dummy variables show the effect in the dependent variable for a discrete change from 0 to 1. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Further dimensions of coverage

	(1)	(2)	(3)	(4)	(5)
	Adjectives	Adj / Words	Images	Categories	Weblinks
CDU/CSU	10.35** (4.376)	0.00123 (0.00170)	0.344*** (0.0987)	0.648*** (0.195)	0.408*** (0.0567)
Left	28.08*** (8.294)	-0.00184 (0.00244)	0.454*** (0.171)	2.772*** (0.325)	0.643*** (0.103)
Greens	19.24* (11.55)	-0.00563** (0.00237)	0.735*** (0.257)	0.510 (0.360)	0.399*** (0.0946)
Female	-4.064 (4.992)	-0.00237 (0.00148)	-0.173* (0.101)	-0.0350 (0.184)	-0.0573 (0.0555)
Former periods in BT	11.46*** (2.039)	0.000925** (0.000396)	0.131*** (0.0422)	0.469*** (0.0581)	0.0620*** (0.0139)
Ancillary Income	0.0198 (0.0211)	-0.000011** (0.000005)	0.00009 (0.00004)	0.0002 (0.00123)	-0.0003 (0.0002)
PhD	22.35*** (8.269)	-0.00406** (0.00192)	0.284** (0.124)	0.433* (0.254)	0.00588 (0.0714)
Constant	34.12*** (4.722)	0.118*** (0.00175)	1.983*** (0.109)	7.737*** (0.188)	1.111*** (0.0559)
N	596	596	596	596	596
R^2	0.149	0.039	0.087	0.212	0.133

Notes: Robust standard errors in parentheses. The dependent variable in column (1) is $Adjectives_i$, which measures the number of adjectives in the Wikipedia biography of MP i . The dependent variable in column (2) is $Adjectives/Words_i$, which measures the adjective-to-word ratio of MP i . The dependent variable in column (3) is $Images_i$, which measures the number of images in the biography of MP i . The dependent variable in column (4) is $Categories_i$, which measures the number of categories that the biography of MP i is assigned to. The dependent variable in column (5) is $Weblinks_i$, which measures the number of weblinks under party control underneath the biography of MP i . CDU/CSU_i , $Left_i$, and $Greens_i$ are dummy variables equal to 1 if MP i is affiliated to that party; SPD is the omitted category. $Female_i$ is equal to 1 if MP i is a woman. $FormerperiodsinBT_i$ counts the election terms that MP i has been in parliament. $Ancillaryincome_i$ is the mean ancillary income of MP i during the 18th election term in 1,000 Euros based on the estimation of `abgeordnetenetwatch.de`. PhD_i is equal to 1 if MP i has a PhD. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Negative coverage

	(1)	(2)	(3)	(4)	(5)	(6)
	Length	Length	Length	Length	Length	ML
CDU/CSU	510.0** (257.5)	451.1* (255.0)	439.5* (247.3)	454.7* (238.3)	704.2 (493.0)	2927.9*** (1103.2)
Left	575.3 (356.9)	634.3* (362.6)	728.9** (349.0)	814.6** (338.5)		3973.5** (1807.3)
Greens	131.9 (312.0)	185.5 (309.2)	285.4 (300.6)	323.8 (295.7)		1424.4 (1718.0)
Female		-396.5* (226.5)	-312.1 (219.1)	-179.7 (214.4)	-132.9 (336.7)	-1091.6 (995.8)
Former periods in BT			370.2*** (79.43)	387.1*** (78.72)	354.1*** (118.7)	531.5* (276.0)
Ancillary Income				1.662 (1.117)	1.656 (1.201)	5.578 (6.731)
PhD				1006.7*** (344.0)	446.8 (564.5)	2016.5* (1117.0)
Population density					0.719*** (0.200)	
Fraction pop. 18-35					39.03 (117.9)	
Abitur					-35.23 (30.28)	
Constant	4539.5*** (175.4)	4703.6*** (196.1)	4121.9*** (198.6)	3810.0*** (188.9)	3781.0* (1951.8)	-8857.4*** (1993.7)
Mills						1.662***
Lambda						(0.275)
N	533	533	533	532	236	532
R^2	0.009	0.014	0.069	0.095	0.132	

Notes: Robust standard errors in parentheses. The results are based on a subsample of observations whose biographies do not contain any criticizing sentences. The dependent variable in columns (1) to (5) is $length_i^G$ which measures Wikipedia coverage of MP i in terms of her biography length in characters. The dependent variable in column (6) is $length_i^G - length_i^E$, which measures the differences in Wikipedia coverage of MP i in terms of her biography length in characters. CDU/CSU_i , $Left_i$, and $Greens_i$ are dummy variables equal to 1 if MP i is affiliated to that party; SPD is the omitted category. $Female_i$ is equal to 1 if MP i is a woman. $Former\ periods\ in\ BT_i$ counts the election terms that MP i has been in parliament. $Ancillary\ income_i$ is the mean ancillary income of MP i during the 18th election term in 1,000 Euros based on the estimation of abgeordnetenwatch.de. PhD_i is equal to 1 if MP i has a PhD. $Population\ density_i$, $Fraction\ pop.\ 18 - 35_i$, and $Fraction\ pop.\ with\ Abitur_i$ refer to i 's constituency demography. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Evidence from the French National Assembly

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Length	Images	Weblinks	Categories	DiD OLS	DiD ML	MEM
LR	1526.9*** (412.6)	-0.187 (0.230)	1.853*** (0.710)	2.153*** (0.472)	858.5* (489.4)	2802.3*** (608.2)	1401.8
UDI	589.9 (819.9)	-0.403 (0.366)	1.807 (1.631)	1.832* (1.045)	829.0 (903.0)	3217.9** (1336.4)	1579.4
GDR	1726.6 (1169.4)	0.837 (0.588)	1.238 (1.700)	3.871*** (1.239)	1232.2 (1078.8)	4006.2*** (1548.1)	2366.4
RRDP	375.2 (1260.7)	0.224 (0.446)	3.046 (2.142)	-0.366 (0.822)	507.7 (1607.9)	194.1 (1592.5)	193.30
NI	3524.4** (1367.3)	-0.297 (0.355)	6.238** (2.668)	0.0244 (0.820)	4094.1** (1824.7)	3909.6** (1582.1)	3119.7
Constant	8122.5*** (247.2)	6.297*** (0.119)	8.762*** (0.459)	20.33*** (0.291)	7121.0*** (343.8)	2753.1*** (455.1)	
N	504	504	504	504	296	504	
R^2	0.043	0.008	0.034	0.055	0.037		

Notes: Robust standard errors in parentheses. The results are based on the MPs from the National Assembly of the 14th legislature of the French Fifth Republic. The dependent variable in column (1) is $length_i^F$ which measures Wikipedia coverage of MP i in terms of her biography length in characters. The dependent variable in column (2) is $Images_i$, which measures the number of images in the biography of MP i . The dependent variable in column (3) is $Weblinks_i$, which measures the number of external weblinks underneath the biography of MP i . The dependent variable in column (4) is $Categories_i$, which measures the number of categories that the biography of MP i is assigned to. LR , UDI , GDR , $RRDP$, and NI are dummy variables equal to 1 if MP i is affiliated to that party; SER is the omitted category. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Survey evidence

	Classroom survey ($N = 60$)				Rep. Survey Age < 37 ($N = 159$)				Rep. Survey Age 37 - 55 ($N = 169$)				Rep. Survey Age > 55 ($N = 172$)				Rep. Survey Pooled ($N = 500$)			
	(1) Three-page biography	(2) One-page biography	(3) Don't know	(4) Difference (1)-(2)	(5) Three-page biography	(6) One-page biography	(7) Don't know	(8) Difference (5)-(6)	(9) Three-page biography	(10) One-page biography	(11) Don't know	(12) Difference (9)-(10)	(13) Three-page biography	(14) One-page biography	(15) Don't know	(16) Difference (13)-(14)	(17) Three-page biography	(18) One-page biography	(19) Don't know	(20) Difference (17)-(18)
Knowledge	0.417	0.1	0.483	0.316***	0.359	0.226	0.415	0.133**	0.254	0.302	0.444	-0.066	0.198	0.413	0.390	-0.218***	0.268	0.316	0.416	-0.048
Intelligence	0.117	0.183	0.7	-0.07	0.233	0.258	0.509	-0.025	0.148	0.349	0.503	-0.201***	0.111	0.424	0.465	-0.313***	0.162	0.346	0.492	-0.184***
Provides strong leadership	0.283	0.2	0.517	0.083	0.308	0.289	0.403	0.019	0.207	0.337	0.456	-0.130**	0.151	0.517	0.331	-0.366***	0.220	0.384	0.396	-0.164 ***
Sets a good moral example	0.217	0.167	0.617	0.05	0.283	0.296	0.421	-0.013	0.148	0.385	0.468	-0.237***	0.145	0.436	0.419	-0.291***	0.190	0.374	0.436	-0.184***
Strength as public servant	0.367	0.133	0.5	0.233***	0.396	0.245	0.359	0.151**	0.260	0.325	0.414	-0.065	0.250	0.424	0.326	-0.174***	0.300	0.334	0.366	-0.034
Empathy	0.25	0.117	0.633	0.133*	0.327	0.182	0.491	0.145**	0.207	0.278	0.515	-0.071	0.215	0.361	0.424	-0.146**	0.248	0.276	0.476	-0.028
Inspiring	0.4	0.117	0.483	0.283***	0.386	0.245	0.371	0.141**	0.296	0.260	0.448	0.036	0.215	0.442	0.343	-0.227***	0.296	0.318	0.386	-0.022
Decency	0.167	0.15	0.683	0.017	0.220	0.315	0.465	-0.095	0.160	0.302	0.538	-0.142***	0.151	0.401	0.448	-0.250***	0.176	0.340	0.490	-0.164***
Honesty	0.083	0.183	0.73	-0.100	0.258	0.352	0.390	-0.094	0.231	0.331	0.438	-0.100*	0.192	0.419	0.390	-0.227***	0.226	0.368	0.406	-0.142***

Notes: This table summarizes the results of our survey. Column (1) shows the nine valence characteristics under consideration. Column (2) shows the share of participants that opted for the three-page biography. Column (3) shows the share of participants that opted for the one-page biography. Column (4) shows the difference between columns (2) and (3) along with its standard error. We test against the null hypothesis that (2) = (3). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Robustness checks I

	(1)	(2)	(3)	(4)
	Length	Length	Length	Length
CDU/CSU	1106.7** (529.0)	1903.7** (850.0)	1047.9*** (381.4)	2167.2*** (780.8)
Left	2554.5*** (758.7)		2171.8*** (662.4)	
Greens	1933.0* (1000.1)		1599.9* (911.8)	
Female	-694.2 (501.3)	330.5 (866.7)	-421.7 (408.1)	135.5 (740.5)
Former periods in BT	679.5*** (173.4)	411.5* (217.6)	905.5*** (159.4)	747.0*** (197.5)
PhD	2085.3*** (691.5)	1562.8 (1021.7)	1917.1*** (613.3)	778.6 (883.1)
Party head	13689.0*** (4381.3)	7010.2 (4710.3)		
Minister during 18th BT	9316.7** (3732.3)	11781.8*** (4530.0)		
Former periods as minister	10701.3*** (2517.3)	11764.1*** (2765.1)		
Early resign			7629.5** (3397.5)	13072.9*** (4265.3)
Population density		0.732** (0.306)		1.515*** (0.201)
Fraction pop. 18 - 35		305.4 (190.5)		92.70 (156.3)
Fraction pop. with Abitur		-36.26 (55.49)		5.823 (47.63)
Constant	3805.8*** (594.6)	-1841.2 (3575.3)	3402.3*** (372.1)	-433.3 (3410.8)
N	633	289	607	275
R^2	0.437	0.542	0.173	0.344

Notes: Robust standard errors in parentheses. The dependent variable $length_i^G$ measures Wikipedia coverage of MP i in terms of her biography length in characters. CDU/CSU_i , $Left_i$, and $Greens_i$ are dummy variables equal to 1 if MP i is affiliated to that party; SPD is the omitted category. $Female_i$ is equal to 1 if MP i is a woman. $Former\ periods\ in\ BT_i$ counts the election terms that MP i has been in parliament. $Ancillary\ income_i$ is the mean ancillary income of MP i during the 18th election term in 1,000 Euros based on the estimation of abgeordnetenwatch.de. PhD_i is equal to 1 if MP i has a PhD. $Population\ density_i$, $Fraction\ pop.\ 18 - 35_i$, and $Fraction\ pop.\ with\ Abitur_i$ refer to i 's constituency demography. $Party\ head_i$ is equal to 1 if MP i is chairman of her party or its parliamentary group. $Minister\ during\ 18th\ BT_i$ is equal to 1 if i was a minister in the 18th election term. $Former\ periods\ as\ minister_i$ counts the election terms during which MP i was minister before the 18th election term. $Early\ resign_i$ is equal to 1 if MP i left the Bundestag early. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Robustness checks II

	(1)	(2)	(3)	(4)	(5)
	DiD	DiD	DiD	DiD	DiD
CDU/CSU	3836.8*** (1212.8)	4767.6*** (1405.9)	5032.0*** (1466.2)	5533.5*** (1590.2)	0.314 (0.233)
Left	8622.4*** (1714.2)	9979.4*** (1944.0)	10373.3*** (2023.7)	11128.3*** (2189.9)	1.035*** (0.345)
Greens	4359.3** (1745.9)	5818.2*** (1993.2)	6255.7*** (2076.7)	7111.0*** (2249.0)	0.00663 (0.347)
Female	-1570.5 (999.5)	-1506.5 (1148.4)	-1499.7 (1200.0)	-1483.2 (1306.8)	-0.232 (0.142)
Former periods in BT	987.7*** (265.4)	1348.2*** (303.4)	1458.8*** (316.2)	1678.5*** (342.9)	0.0645 (0.0737)
PhD	2565.7** (1123.5)	3007.0** (1293.2)	3167.8** (1351.8)	3501.7** (1473.3)	0.138 (0.193)
Ancillary Income	0.802 (5.490)	0.832 (6.145)	0.938 (6.392)	1.180 (6.916)	- 0.0000 (0.0009)
Constant	-13683.6*** (1909.9)	-15861.5*** (2043.0)	-16490.7*** (2096.4)	-17738.0*** (2220.4)	-0.0713 (0.952)
Mills	6793.5***	8244.2***	8699.7***	9621.6***	0.284
Lambda	(812.0)	(807.0)	(820.1)	(854.2)	(0.510)
<i>N</i>	596	596	596	596	596

Notes: Robust standard errors in parentheses. The results are based on an ML estimation of the DiD selection model. The dependent variable is $length_i^G - length_i^E$, which measures the differences in Wikipedia coverage of MP i in terms of her biography length in characters. In columns (1) to (4) the German length $length_i^G$ was scaled with the factors 0.6, 0.75, 0.8, and 0.9, respectively, before taking the difference. CDU/CSU_i , $Left_i$, and $Greens_i$ are dummy variables equal to 1 if MP i is affiliated to that party; SPD is the omitted category. $Female_i$ is equal to 1 if MP i is a woman. $Former\ periods\ in\ BT_i$ counts the election terms that MP i has been in parliament. $Ancillary\ income_i$ is the mean ancillary income of MP i during the 18th election term in 1,000 Euros based on the estimation of abgeordnetenwatch.de. PhD_i is equal to 1 if MP i has a PhD. $Population\ density_i$, $Fraction\ pop.\ 18 - 35_i$, and $Fraction\ pop.\ with\ Abitur_i$ refer to i 's constituency demography. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.