

UC Davis

UC Davis Previously Published Works

Title

Complicating Wisdom of Crowds: News Sources Reveal Political Bias in Wikipedia Biographies

Permalink

<https://escholarship.org/uc/item/17p0h7fm>

Journal

Digital Journalism, ahead-of-print(ahead-of-print)

ISSN

2167-0811

Authors

Liang, Zhengyi

Cho, Jaeho

Publication Date

2026-07-07

DOI

10.1080/21670811.2026.2700581

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Complicating Wisdom of Crowds: News Sources Reveal Political Bias in Wikipedia Biographies

Abstract

With the rise of partisan media and ideologically segregated news feeds on social media, balanced information is much needed for citizens to reference when they engage in politics. This paper seeks to explore whether Wikipedia, a peer-production platform, can be an alternative source that offers politically balanced information. By analyzing news references in the Wikipedia biographies of U.S. elected officials, we find evidence of a moderate but significant liberal bias, with liberal media sources cited more extensively than conservative ones across the pages of both Democrats and Republicans. We then discover that the bias in terms of the use of news citations is significantly associated with workload diversity and communication influence during the process of crowd editing. Our findings contribute to the emerging literature on bias in peer-production communities and shed light on the influence of digital media on collaborative knowledge production. Furthermore, our results challenge the belief that political neutrality can be safely assumed from collaborative editing, offering implications for digital journalism on how to produce more balanced, two-sided information.

Keywords: Political bias, Wikipedia, News sources, Crowd editing, Wisdom of crowds, Digital journalism, Peer production

Introduction

In democracies, citizens ideally rely on unbiased and accurate information to form their political judgments and voting decisions, laying a foundation for a robust democratic system. However, the contemporary information landscape increasingly deviates from this ideal. Partisan media outlets present slanted news coverage, which fuels disdain and skepticism toward the opposing side (Levendusky, 2013). Traditional journalism, once characterized as fact-driven and impartial reporting, now often prioritizes narratives over neutrally presented facts, diminishing opportunities for audiences to understand opposing viewpoints (Mutz, 2006). Compounding this, social media platforms utilize algorithms to curate news feeds, situating users in tailored content bubbles and thus reinforcing ideological divides (González-Bailón et al., 2023; Pariser, 2011). This trend appears at odds with the promise of universal access to diverse information by the Internet (Benkler, 2006). In the context of a politically polarized information environment, our study investigates whether Wikipedia, a collaborative peer-production platform, can be an alternative source for more balanced political information.

Wikipedia attracts millions of visits monthly (Wikipedia.org, 2023), establishing itself as a successful peer-production project capitalizing on the collective wisdom of its users (Foote et al., 2023). Its mission extends beyond providing “free access to the sum of all human knowledge”; it also offers individuals the liberty to participate in editing and refining this collective knowledge (Reagle & Koerner, 2020). The platform upholds the principle of a neutral point of view, which calls for the inclusion of “all verifiable points of view that have sufficient

News Sources Reveal Political Bias

due weight” (Wikipedia, n.d.). Unlike traditional news media, which operates under the influence of market dynamics and hierarchical editorial structures, Wikipedia rallies volunteers from diverse backgrounds and motivations to edit and curate its content (Benkler, 2016). This inclusive model allows individuals of varying political perspectives to contribute, ideally leading to information that is presented from a neutral standpoint. Nevertheless, recent studies indicate the presence of potential bias in Wikipedia articles through its collaborative editing model (Kerkhof & Münster, 2019; Langrock & González-Bailón, 2022; Tripodi, 2023). We posit that the collaborative editing process does not necessarily lead to the production of balanced and fair information. Within this context, our study aims to investigate whether political bias exists in Wikipedia’s political content, and if so, to what extent. Furthermore, we examine whether and how the characteristics of crowd editing are associated with political bias on Wikipedia pages.

In the sections that follow, we begin by reviewing existing literature on content bias, methods for measuring bias, and prior studies that have identified potential correlations between bias and characteristics of collaborative editing, including factors such as diversity, crowd size, and the influence of communication. We also discuss the relationship between Wikipedia and digital journalism by elucidating journalistic processes on the peer-production platform. We then perform a computational analysis by probing the Wikipedia biographies of U.S. politicians and the editing metrics at the backstage. Our results suggest that there exists a certain type of political bias as measured by the use of news sources in the Wikipedia articles. Further, diversity and communication influence are found to be significantly associated with political bias. Finally, we conclude this paper with discussions of findings and implications.

Literature Review

Content Bias on Media and Beyond

Content bias refers to the systematic patterns in structuring issues that offer one an advantage over the other (Entman, 2007). This concept has been widely applied to the media context. News outlets often produce one-sided stories that favor chosen entities so as to elicit their preferred actions in readers (Gentzkow & Shapiro, 2010). Haselmayer et al. (2017) found that news coverage tends to align with the political parties that readers support, suggesting the role of media in shaping democratic elections and campaigns. The effects also spill over to the news coverage of politicized issues, such as vaccination (Hart et al., 2020), COVID-19 infections (Spiteri, 2021), and climate change (Schäfer & O'Neill, 2017).

Although the media serves as a significant source of knowledge for Americans (Brenan & Stubbs, 2020), news exposure constitutes only a portion of one's information diet (Wojcieszak et al., 2023). The contemporary information environment is increasingly complex, with individuals frequently turning to platforms outside traditional media, such as social media (Social Media and News Fact Sheet, 2022) and online communities. Wikipedia is one of the top 10 most visited websites (Top Websites Ranking, 2023), enabling it to reach tens of millions worldwide. Wikipedia's success has established it as a leading exemplar of peer-production communities. Given Wikipedia's popularity and accessibility, it is imperative to assess the quality of its resources that are broadly accessible to the public.

Wikipedia is an online encyclopedia that aims to provide free access to “the sum of all human knowledge” (Wikipedia: Prime objective, n.d.). To accomplish the common good, the platform adopts the practice of peer production and encourages volunteers to participate in the

News Sources Reveal Political Bias

editing, discussion, and review processes during massive collaboration (Langrock & González-Bailón, 2022; Reagle & Koerner, 2020). The platform upholds the neutral point of view policy by requiring editors to consider competing views that are supported by credible sources (Wikipedia, n.d.).

However, existing research has identified potential biases in Wikipedia pages. Prior studies have revealed various types of biases in content, such as gender bias (Langrock & González-Bailón, 2022; Reagle & Rhue, 2011), cultural bias (Callahan & Herring, 2011), and underrepresentation from developing nations (Graham et al., 2015) on Wikipedia. Yet, only a handful of these studies have focused on political bias in Wikipedia articles. Greenstein and Zhu (2012) analyzed political articles about US politics on Wikipedia and found that while articles initially exhibited a liberal bias, they eventually progressed toward neutrality and contained an approximately equal number of phrases from the two major parties. Also, political bias might be linked to party affiliations. Pradel (2021) found that there was more personal information about right-leaning German politicians than their left-leaning counterparts in Wikipedia articles. Kerkhof and Münster (2019) detected a coverage bias, noting significant differences in the length of biographies between right-wing and left-wing party members in German. This disparity might arise from partisan contributions as editors from regions favoring a major party often engage more actively in Wikipedia edits and discussions. A potential source of emerging bias is the platform's verifiability policy. Wikipedia espouses that readers should be able to verify the information on the pages, and thereby editors are expected to provide reliable and published sources while editing. The policy aims to eliminate baseless claims, but it overlooks the potential biases present within the publications themselves, possibly leading to skewed content on the

News Sources Reveal Political Bias

platform. By following the principle, contributors may unintentionally replicate and accumulate the bias embedded in the published sources. As a result, the policy about reliability could reduce representation and diversity in content (Koerner, 2020).

Bias Measurements

Two common approaches have been adopted to measure media bias: audience-based and content-based methods. First, an assumption underlying audience-based methods is that the ideological leanings of media outlets align with those of their patrons, and the direction of effects operates both ways. Prior research computed the political position of media sources by averaging the ideological scores of users who shared content from an outlet on Facebook (Bakshy et al., 2015; Moore et al., 2023). The large-scale social media data enables researchers to estimate the ideological leanings of numerous media sources based on user behaviors, such as following and sharing patterns (An et al., 2012; Faris et al., 2017; Wang et al., 2024). The audience-based approach is well-grounded, but some argue that its measurement can be flawed because audience segmentation could be attributed to content with slight differences on media outlets (Budak et al., 2016). Moreover, given the dynamic nature of viewership in a highly competitive media landscape, media bias inferred from audience preferences at one point of time might not hold true later on (Kim et al., 2022).

Second, some researchers directly analyze news coverage to quantify media bias (Gentzkow & Shapiro, 2010; Hamborg et al., 2019). For example, Budak et al. (2016) investigated over 10,000 political articles through supervised machine learning with

News Sources Reveal Political Bias

crowdsourced annotations. Quantifying media bias through text analysis is a computationally challenging task and is prone to misinterpretations and errors. Unsupervised algorithms, such as topic modeling, reveal latent patterns behind texts without prelabeled data, which leaves substantial room for interpretation. As a result, such interpretations can sometimes be arbitrary and lack consensus (Denny & Sprirling, 2018). Moreover, supervised scaling techniques, like machine learning, can introduce human errors through the annotation process.

Additionally, social scientists have been exploring alternative ways to measure media bias, aiming to address the limitations of the existing scaling approaches. Examining news characteristics without relying solely on content has emerged as a promising direction. Kim et al. (2022) analyzed TV videos, gauging media bias through the partisanship of political actors on cable news. The political position of a news channel was determined by aggregating the ideological scores of politicians featured on the show. Besides, references and sources used in news reporting were leveraged to estimate media bias. Lin et al. (2011) calculated the ideological positions of media outlets by the number of times congress members were mentioned in news coverage. An outlet that frequently quoted Republicans (or Democrats) was likely to have a conservative (or liberal) slant. Groseclose and Milyo (2005) estimated the ideological positions of media outlets by evaluating the frequency with which think tanks, policy groups, and members of congress with explicit political ideologies were cited by news reports.

These approaches reflect a tradition of audit-style research on media bias, which systematically examines ideological slant (Budak et al., 2016), source selection (Lin et al., 2011), and representational asymmetries (Faris et al., 2017) in news coverage. However, this literature has largely focused on bias within news organizations, with less attention to how media bias

News Sources Reveal Political Bias

migrates beyond newsrooms and becomes embedded in downstream information platforms.

Examining bias on Wikipedia therefore extends prior research by tracing how patterns of news

bias are incorporated into collaboratively produced knowledge. This perspective expands our

understanding of media bias by showing how news slant, as it moves from news production into

shared knowledge infrastructures, can shape political interpretation.

Wikipedia bias is an emerging research area in comparison to media bias studies. A significant body of research quantifies bias using article characteristics, such as length, quality, visibility, and references (Callahan & Herring, 2011; Graham et al., 2015; Langrock & González-Bailón, 2022). Only a few studies have explored content features to gauge media slant. For instance, drawing on Gentzkow and Shapiro's (2010) framework, Greenstein and Zhu (2012) estimated bias by the frequency of phrases from congressional reports in articles. Pradel (2021) developed a bias measure using the percentage of certain words in the text.

Bias in Crowd Editing

Much research examining the causes of media bias has pointed toward news production in the newsroom, such as gatekeeping practice by journalists (Groeling, 2013; Schäfer & O'Neill, 2017). In contrast, Wikipedia has pioneered a transformative approach to knowledge production that relies on volunteers, who are regular people, rather than experts and professionals. Hence, this study aims to investigate how bias emerges from the collaborative editing process. Crowd editing is underpinned by the concept of collective intelligence, also known as the wisdom of crowds. This idea posits that the average estimates of individuals tend to be more accurate than the judgments of singular experts (Lorenz et al., 2011). However,

News Sources Reveal Political Bias

attaining this wisdom of crowds depends on certain conditions (Aral, 2021), and failing to meet the conditions can lead to biased outcomes.

Discussions have focused on how two types of diversity bolster the wisdom of crowds. The first pertains to diversity in the characteristics of crowd members, including aspects like political leanings (Greenstein & Zhu, 2016; Kerkhof & Münster, 2019) and demographics (e.g., gender and race). Research indicates that teams with diverse identities produced better outcomes (Hong & Page, 2004; Shi et al., 2019). Yet, some studies indicate that demographic diversity only marginally enhances group outcomes (De Oliveira et al., 2018; Horwitz & Horwitz, 2007). This has steered the discussion toward the second form of diversity: task-relevant experience. Hong et al. (2020) found a positive correlation between diverse task-related experiences and the precision of crowd predictions. Specifically, workload diversity refers to the extent to which a task is equitably distributed among team members, rather than being dominated by a few people (Hong et al., 2020; Kittur & Kraut, 2008). When a task is approached in a more horizontal fashion—characterized by substantial contributions from a larger number of individuals—collective intelligence is more likely to emerge than when the task is driven by a small subset of contributors (Robert & Romero, 2017). In the context of Wikipedia, equitable participation in article editing increases the likelihood of reducing bias by incorporating a wider range of opinions, expertise, and experiences into the content creation process (Shi et al., 2019). That is, editors contribute diverse cognitive resources and perspectives, thereby presenting more balanced information through the inclusion of a broader range of sources and more representative citations.

News Sources Reveal Political Bias

Crowd size is another critical factor influencing the outcomes of large-scale collaboration. It is likely that larger groups of editors (or larger crowds in general) can enhance content quality and reduce bias (Walter et al., 2022), provided that increased crowd size could potentially bring greater diversity in citations and perspectives (Robert & Romero, 2017). For instance, articles edited by a large number of contributors holding divergent or even polarizing opinions are more likely to be neutral and less biased compared to those edited by a small, homogeneous group (Greenstein et al., 2016). However, research also indicates that a larger group size does not automatically lead to reduced bias (Hong et al., 2020), as the impact of crowd size would be contingent on the diversity within the group (Kittur & Kraut, 2008; Robert & Romero, 2016). Homogenous large groups may end up with reinforcing dominant viewpoints and existing bias. In brief, broad participation from individuals with mixed viewpoints introduces heterogeneous citations, which can offset extreme biases.

Additionally, communication influence within a crowd plays an essential role in shaping crowd performance and mitigating bias. Intragroup communication can positively influence outcomes by facilitating productive deliberation over citation choices (Guth & Brabham, 2017). On Wikipedia, frequent interactions among editors often lead to greater scrutiny of sources, promoting more balanced and neutral citation practices. Prior research suggests that sufficient and detailed discussions tend to introduce ideologically diverse perspectives, thereby yielding balanced outcomes and high-quality collaboration (Shi et al., 2019). However, on the other hand, intense communication among editors may facilitate the emergence of group norms and social hierarchies that suppress diverse viewpoints (Kittur & Kraut, 2008). When experienced editors endorse biased sources during discussions, others may be less inclined to challenge these choices

News Sources Reveal Political Bias

due to a perceived pressure to conform. This dynamic may lead to citation convergence around biased sources due to social conformity pressures (Lorenz et al., 2011).

Journalistic Processes on Wikipedia

A growing body of research has examined the relationship between Wikipedia and digital media (GeiB et al., 2016; Keegan et al., 2013), extending beyond Wikipedia's function as a popular platform for knowledge production. Wikipedia editors engage in a distinctive form of citizen journalism in which ordinary individuals collectively participate in journalistic practices. Two such practices are particularly evident on the platform—gatekeeping and collective editing (Ren & Xu, 2023). Gatekeeping refers to the act of including or excluding news items on a medium (Bro & Wallberg, 2014). Wikipedia contributors participate in gatekeeping by determining an event's newsworthiness (Ren & Xu, 2023). Specifically, editors may rely on criteria of sufficient notability and international importance to evaluate news items (Keegan & Gergle, 2010). Events meeting these criteria are more likely to be incorporated into Wikipedia articles. Furthermore, Wikipedia's gatekeeping differs from that of traditional media. In traditional journalism, gatekeeping has been a top-down, hierarchical process in which information flows from sources through professional editors (Bürger & Schmid-Petri, 2023). By contrast, Wikipedia's gatekeeping is decentralized and collaborative. Rather than a single editor or newsroom deciding the news stories, Wikipedia's content decisions emerge from consensus among multiple editors. This collective gatekeeping is less hierarchical and more transparent, as discussions and edit histories are open to the community.

News Sources Reveal Political Bias

After determining what to discuss through the gatekeeping process, Wikipedia editors collaboratively generate content that reflects their understanding of reality. Existing research has primarily focused on how Wikipedia contributors respond to major real-world events (Fairchild et al., 2016), often emphasizing spikes in editing activity during such periods (Twyman et al., 2017). However, the role of news citations in the editorial process remains comparatively underexplored. News sources are widely cited across Wikipedia articles (Puyu et al., 2024), and reliance on these sources may introduce existing media bias into collaborative editing. The present study investigates patterns in news citations on Wikipedia, thereby revealing how digital journalism influences information production on the platform.

Additionally, Wikipedia influences its audience in a manner distinct from traditional media. While traditional media outlets continue to play a critical role in agenda setting (Meraz, 2009)—often being the first to spotlight issues and determine what is considered newsworthy—Wikipedia generally functions as a downstream information source (Gildersleve et al., 2023). When headlines attract public interest, audiences likely turn to Wikipedia to obtain contextual and background information (Yasseri & Bright, 2016). Unlike newspapers or television news programs that deliver information to largely passive audiences, Wikipedia’s influence operates on an on-demand basis: readers actively choose which entries to consult. Consequently, its impact lies more in deepening public understanding (Gildersleve et al., 2023) than in initiating topics of public discourse.

Research Questions

News Sources Reveal Political Bias

Our study attempts to address whether Wikipedia provides balanced and fair political information in the context of a polarized information environment. We propose to investigate news references, a robust indicator for content bias, in the Wikipedia profiles for American politicians. We argue that news citation patterns in respect to the political leanings of media outlets can provide insights into the political propensity in the articles. Thus, we propose our first research question:

RQ1: What are the citation patterns regarding the use of liberal and conservative news sources on Wikipedia pages for U.S. politicians?

Additionally, to answer our second research question, we estimate political bias from the use of media sources in the Wikipedia biographies. The estimation technique decreases computational complexity and captures content bias effectively. Moreover, prior studies revealed bias could be generated from crowd editing (Kerkhof & Münster, 2019; Langrock & González-Bailón, 2022). Hence, we delve into the editing process to explore the potential associations between political bias and the vital features of crowd editing. Our second research question is as follows:

RQ2: How are characteristics of crowd editing (i.e., workload diversity, crowd size, communication intensity) associated with political bias on Wikipedia pages for U.S. politicians?

Data Collection

We developed a python crawler to scrape references from Wikipedia pages of U.S. officials who served or are serving in the 114th to 118th Congress (from 2015-2023). In total, we collected data for 875 politicians, including 140 senators and 735 representatives. The Wikipedia

News Sources Reveal Political Bias

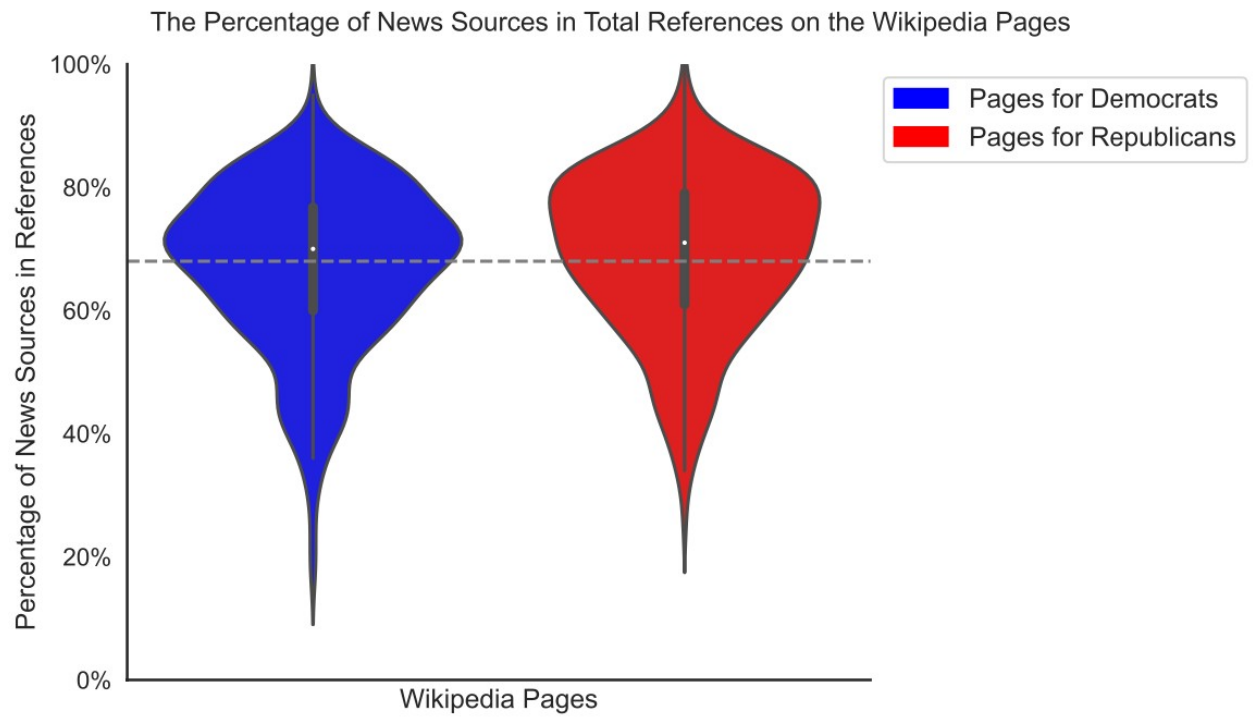
pages of these U.S. politicians contained a total of 60,851 references. Notably, 84.97% of these references are from news sources. Other sources are also cited in the Wikipedia profiles, including websites¹, books (0.86%), magazines (0.83%), press releases (0.48%) and journals (0.33%). Regarding news sources on the pages, our dataset consists of 51,705 news citations, and 566 of them lack information about news sites because the editors did not specify their sources in the references. As a result, the percentage of missing data is 1.09%.

Next, we determined each news outlet's ideological leaning based on Robertson et al. (2018). This research team quantified media bias by analyzing Twitter users' sharing patterns. Initially, over half a million Twitter accounts were matched with voter records to ascertain users' political affiliations. Twitter profiles were retained when names and addresses matched those in the voter record. Next, billions of tweets were scraped, and domain information in URLs, such as 'nytimes.com' in '<https://www.nytimes.com/section/politics>,' was filtered. Bias scores for news websites were calculated based on the assumption that domains shared by Republicans leaned conservative, while those shared by Democrats leaned liberal. These scores ranged from -1 (strongly left-leaning) to 1 (strongly right-leaning). The measurement demonstrated high validity, with results showing strong consistency with previous studies (Bakshy et al., 2015; Budak et al., 2016; Mitchell et al., 2014). Overall, Robertson et al. (2018) classified 19,022 domains along the liberal-conservative ideological spectrum, identifying 9,919 as conservative and 9,103 as liberal.

¹ Websites are a significant source of citations on Wikipedia. Initially, 41.19% of the references were from news sources, and 55.63% of them were from websites. However, we observed an overlap between these two categories, indicating that some references from websites also originated from news outlets. Consequently, we identified this overlap and incorporated it into our analysis. As a result, the percentage of references from news sources increased to 84.97%.

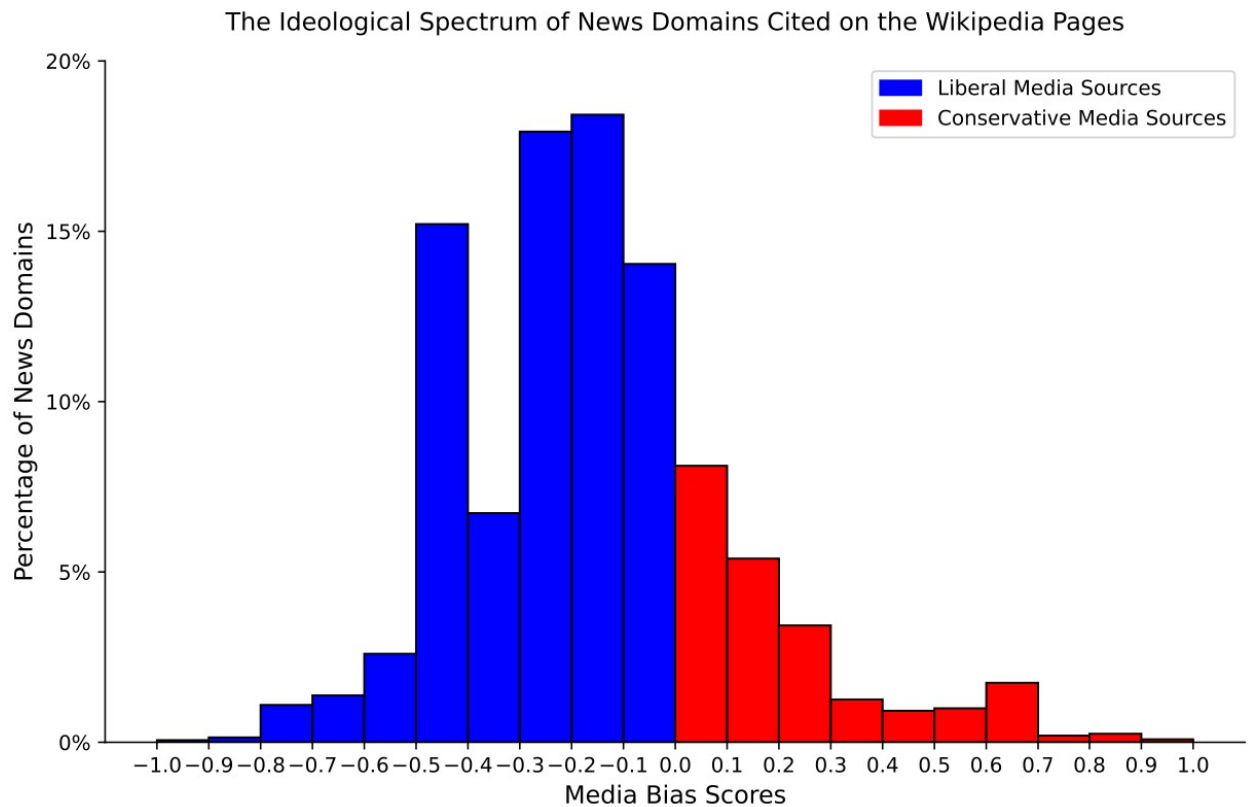
News Sources Reveal Political Bias

a.

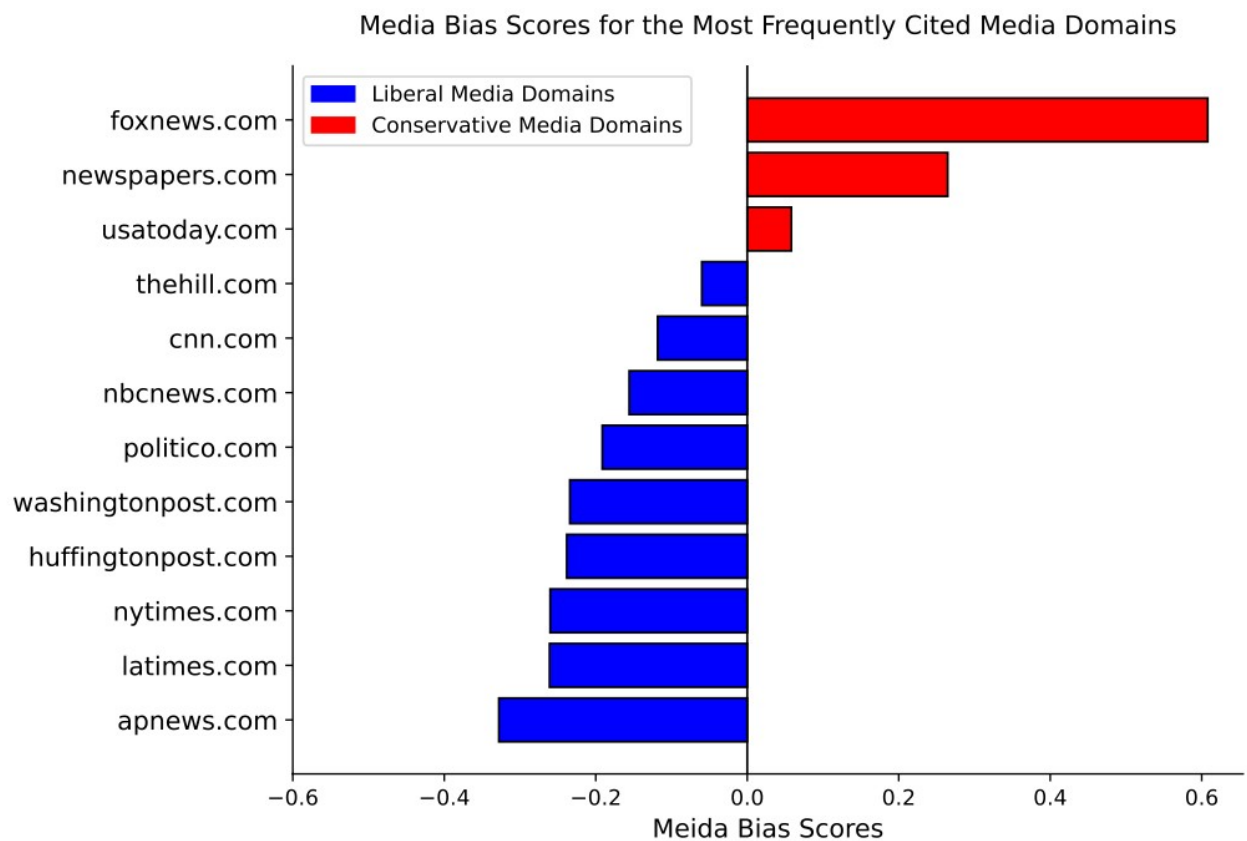


b.

News Sources Reveal Political Bias



c.



News Sources Reveal Political Bias

Figure 1 | News sources cited on the Wikipedia pages of U.S. elected officials. a. The percentage of news sources in total references on pages for Republicans and Democrats. b. The ideological spectrum of news domains cited on the Wikipedia pages. c. Media bias scores for the most frequently cited media domains.

Figure 1a displays the percentage of news sources within total references on Wikipedia profiles for U.S. elected officials. The figure illustrates the extensive use of news resources in those profiles. On average, 68.26% of the citations on the pages were attributed to news outlets with known political leanings, as indicated by the dotted horizontal line. This pattern is consistent for profiles of both Republicans and Democrats, highlighting the heavy reliance on news sources. This overwhelming use of news sources justifies exploring bias in Wikipedia by examining bias in the media references cited. Additionally, Figure 1b presents the ideological spectrum of news domains cited in the Wikipedia articles. The X-axis represents media bias scores, and the Y-axis indicates the percentage of news domains falling into each bin on the X-axis. As shown in the figure, liberal media sources were used more extensively than conservative media sources in composing the Wikipedia profiles of U.S. politicians. Specifically, 85.72% of the media sources lean liberal, while 14.28% of the media sources lean conservative. Besides, Figure 1c illustrates the media bias scores for the 12 most frequently cited media domains. As the figure indicates, liberal media domains are more popular than conservative media domains among the most cited, and the well-cited ones tend to be major media outlets.

In total, our dataset included 2,572 news domains. After applying the classification scheme, we were able to determine the political leanings of 1,405 news outlets, comprising 724 liberal outlets and 681 conservative outlets. As these media outlets were both popular and widely cited, they accounted for 43,613 references, representing 71.67% of all citations in the dataset.

News Sources Reveal Political Bias

Subsequent analyses were based on the references from news domains with known political leanings.

In addition to references on the Wikipedia articles, we gathered article characteristics and editing statistics including character counts, page quality, pageviews, reference counts, monthly edit counts, editor counts, and the number of words and edits on talk pages. Wikipedia contributors utilize talk pages as a forum to discuss the content and structure of articles (Kittur et al., 2007). It is common for members to ask for advice or opinions, invite others to participate in discussions and share relevant information on talk pages (Morgan et al., 2013). Most data features were collected through Wikipedia API at <https://xtools.wmcloud.org/api>, and the comments on talk pages were harvested through a python package, pywikibot.

Results

To address RQ1, we developed a metric of political bias to quantify the ideological tilt in the use of news sources cited on Wikipedia pages, positioning them along a liberal–conservative continuum. Political bias is operationalized as the sum of the bias scores of all cited sources on a given page, divided by the total number of sources cited. For this study, we collected 875 Wikipedia pages of U.S. politicians and calculated the political bias score B for each page using the following formula:

$$B = \frac{\sum_{i=1}^n (S_i * C_i)}{N} \quad (1)$$

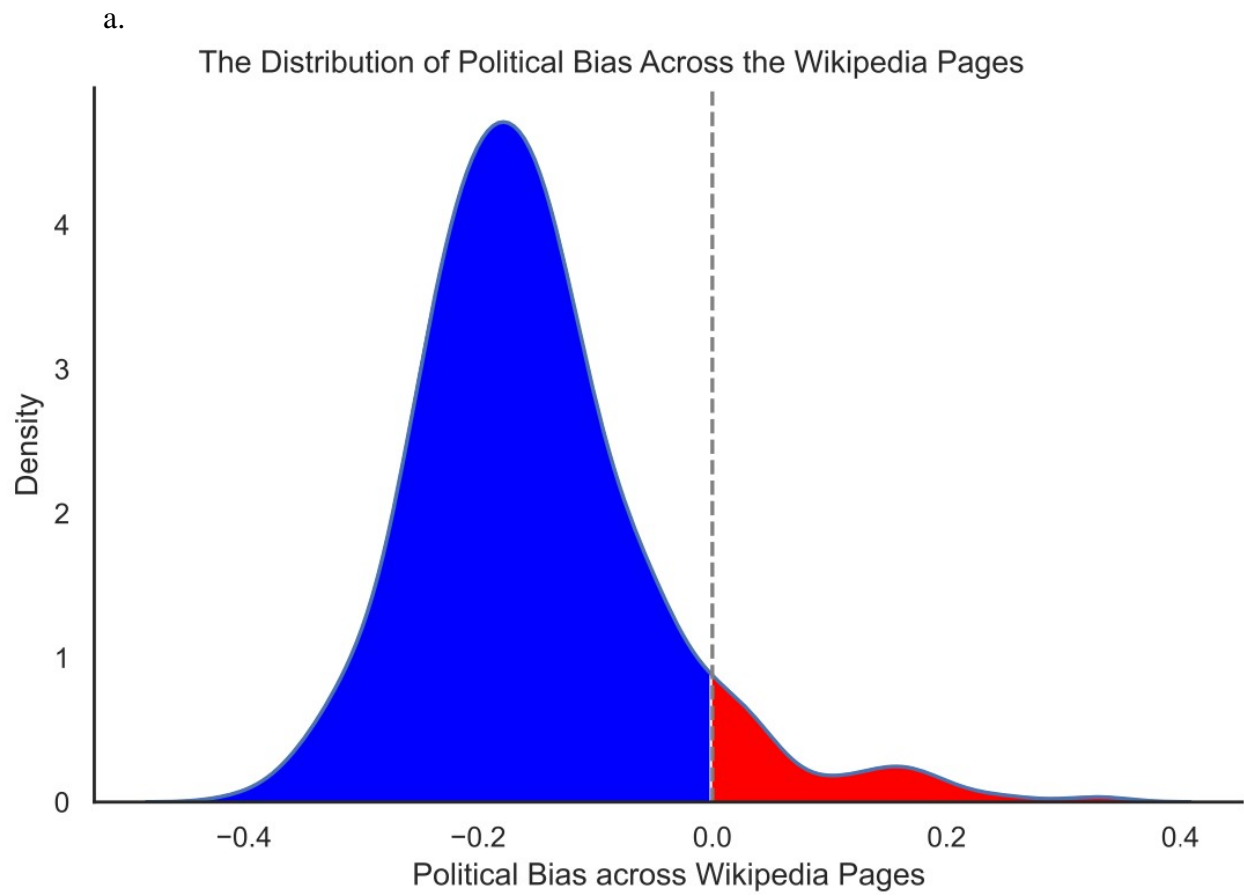
where S_i is the bias score of each source i from Robertson et al. (2018) ranging from -1 to 1 .

News Sources Reveal Political Bias

C_i is the number of times each source i is cited on a given page, and N is the total number of sources (including both liberal and conservative news sources) cited on the page. The resulting value B ranges from -1 to 1 , with negative scores indicating a lean toward liberal media bias and positive scores indicating a lean toward conservative media bias.

Our measurement of political bias aligns with Wikipedia's stated commitment to neutrality. To present a neutral point of view (Wikipedia, n.d.), a politically neutral article should incorporate both liberal and conservative media sources to ensure a balanced perspective. In our framework, when the bias scores of conservative and liberal sources cancel each other out—i.e., are equal in magnitude but opposite in direction—the resulting political bias score B equals zero, indicating a balanced use of news sources. Conversely, if a Wikipedia page disproportionately cites liberal or conservative sources, the page is considered to exhibit a corresponding ideological bias (Tan & Weaver, 2010), suggesting a tilt toward one point of view over the other.

News Sources Reveal Political Bias



b.

News Sources Reveal Political Bias

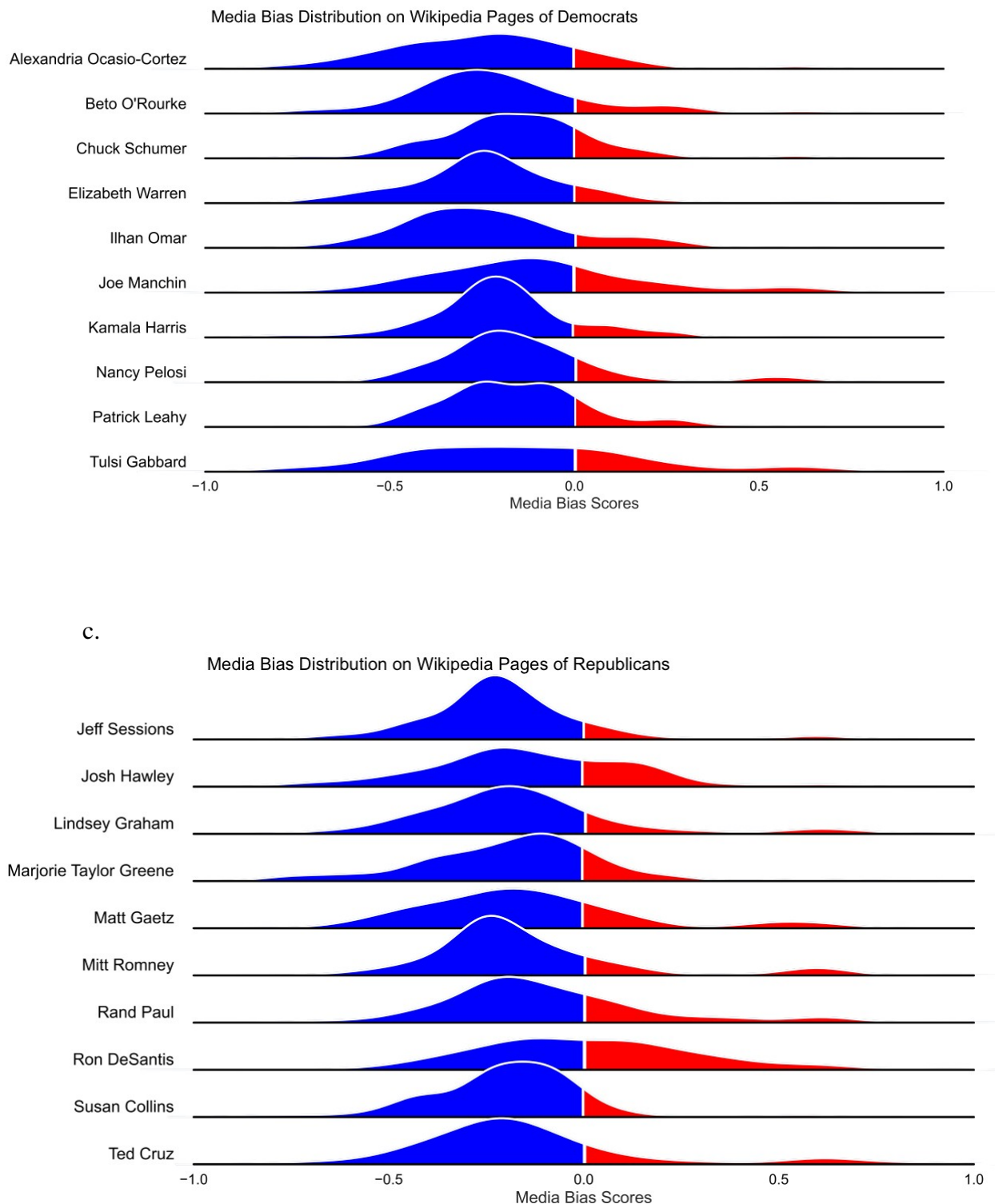


Figure 2 | Political bias on the Wikipedia Pages of the U.S. elected officials. a. The distribution of political bias across the Wikipedia pages. b. The distributions of media bias scores for Wikipedia pages among prominent Democrats. c. The distributions of media bias scores for Wikipedia pages among prominent Republicans.

Figure 2 presents political bias across the Wikipedia pages for U.S. politicians. As shown in Figure 2a, the distribution of political bias for each page is centered around -0.15 with most bias scores concentrating on the left. This right-skewed distribution reveals that a moderate but significant liberal bias is embedded in the Wikipedia pages of U.S. elected officials, regardless of their political affiliations (*Welch t statistic* = -44.61 , $p < 0.001$, $N = 875$). Our findings demonstrate a preference for liberal news sources in the profiles of U.S. politicians on Wikipedia. Furthermore, Figure 2b and 2c plot the distributions of media bias scores for news references on the Wikipedia pages among influential Democrats and Republicans. As shown in the two subplots, liberal news sources were extensively more cited than their conservative counterparts across the Democratic and Republican pages. Only the Wikipedia page for Ron DeSantis demonstrates a relatively balanced use of liberal and conservative media sources. This pattern mirrors the findings in Figure 2a, highlighting the heavy reliance on liberal media sources in the Wikipedia profiles for U.S. politicians irrespective of their party affiliations.

While our analysis relied on news sources to estimate political bias on the Wikipedia pages, we also sought to investigate the editing processes through which this bias emerges. To answer RQ2, we developed three key measurements (i.e., crowd size, workload diversity, and communication intensity) that significantly influence crowd performance (Robertson et al., 2018; Yan et al., 2021). These measurements help us analyze the impact of crowd editing on political bias. Crowd size refers to the number of editors collaborating on a Wikipedia article, and it increases as more editors get involved. Moreover, the edits are not evenly distributed, with a few authors contributing to a large portion of edits. To capture this, we measured workload diversity

News Sources Reveal Political Bias

using the Gini coefficient, which assesses the inequality in contributions among editors on Wikipedia pages. The formula to compute workload diversity is as follows:

$$G = \frac{\sum_{i=1}^n (2i - n - 1) x_i}{n \sum_{i=1}^n x_i} \quad (2)$$

where x represents the edit count, n is the number of edit counts observed, and i is the rank of x among all values in ascending order. A low G value indicates that all editors made similar contributions, while a high G value suggests that a few editors performed the majority of edits.

To simplify interpretation, we multiplied G values by -1 , so as the G value increases, the diversity of workload also increases. Next, communication intensity measures the extent of discussions among editors on talk pages. Wikipedia editors contribute to articles on the front stage and engage in discussions about the topics on the backstage (i.e., talk pages) (Shi et al., 2019). Communication intensity is a function of talk page length and edit counts that demonstrates the volume of conversations. The formula is defined as follows:

$$C = \frac{L}{N_{\text{edits}}} \quad (3)$$

where L denotes talk page length in terms of the number of words, and N_{edits} represents the number of edits on the talk page. As the C value increases, editors engage in longer and more intense debates on the discussion platform.

Figure 3 plots estimated coefficients from linear models predicting political bias on Wikipedia pages for U.S. politicians, along with 95% confidence intervals. The variables were classified into three categories: editors' attributes, Wikipedia subjects' attributes, and information quality attributes. Control variables included information quality characteristics such as character

News Sources Reveal Political Bias

count, page quality, pageviews, reference count, and monthly edit counts. We found that an increase in workload diversity correlated with a more liberal bias in articles ($B = -0.009$, $t = -2.748$, $p < .05$). In other words, when Wikipedians contributed relatively equally to editing an article, they tended to cite more liberal news sources, resulting in a liberal-leaning article. Additionally, higher communication intensity had a significant effect on political bias ($B = -0.007$, $t = -2.103$, $p < .05$)². Longer and more elaborate debates among editors were associated with a preference for liberal news sources in articles, leading to a liberal bias. However, crowd size was not significantly associated with political bias ($B = -0.002$, $t = -0.244$, $p = .81$).

Furthermore, we explored the impact of politicians' demographics on political bias. Articles for senators exhibited more conservative bias compared to those of representatives ($B = 0.027$, $t = 2.628$, $p < .01$). Republican pages were more likely to cite conservative media sources than Democratic pages ($B = 0.059$, $t = 8.648$, $p < .001$). Additionally, as politicians' age increased, their Wikipedia profiles tended to display a more liberal bias in their use of news sources ($B = -0.021$, $t = -5.196$, $p < .001$). However, gender did not significantly correlate with political bias ($B = -0.01$, $t = -1.258$, $p = .21$). We note that our regression results have robust standard errors, and the regression coefficients are relatively small because the dependent variable political bias has small values, ranging from -1 to 1 . The regression table can be found at the appendix.

² Communication intensity is determined by talk page length and the number of edits. Wikipedia editing is carried out by both humans and bots. We conducted two separate calculations, one based on edits made by humans and bots, and another based solely on human edits. Both measures produced similar significant regression results. The parameters discussed in the main text are driven from the second (human-edit only) calculation.

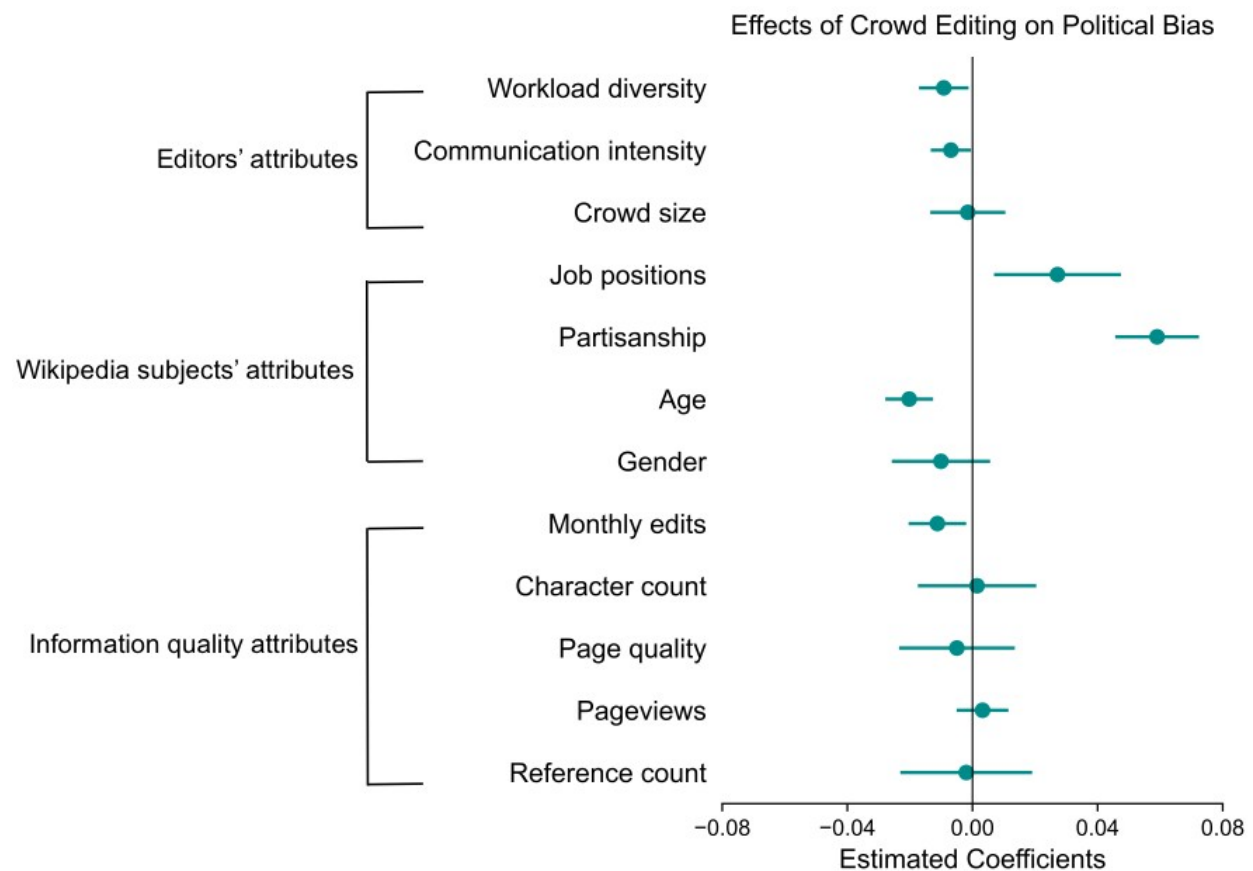


Figure 3 | Effects of Crowd Editing on Political Bias.

Conclusion

This study examined political bias on Wikipedia in relation to the use of liberal and conservative news references, with two primary goals. First, we aimed to determine whether the popular online encyclopedia could serve as an alternative source of balanced information amid an increasingly polarized media landscape. Second, we sought to explore whether, and how, editing patterns and characteristics on the platform are associated with bias, if present. By

News Sources Reveal Political Bias

analyzing the news references on the Wikipedia pages for U.S. elected officials, we found that news sources were extensively cited in the profiles with 68.26% citations coming from media outlets. Furthermore, our analysis revealed a moderate but significant liberal bias, with liberal news sources cited more frequently than conservative ones in the Wikipedia biographies for the politicians, regardless of their political affiliations. The results indicate a preference for the use of liberal news sources, hinting at potential content bias in Wikipedia articles. The skewed patterns of news citations imply that media bias might permeate the online encyclopedia, potentially affecting the neutrality of the articles. This calls into question Wikipedia's role as a neutral platform offering balanced political information.

The preference for liberal media sources likely reflects a form of framing bias: amplifying voices from one ideology, which may affect audience trust and discourage engagement. When Wikipedia editors consistently rely on one side's sources, they risk misrepresenting a politician's profile by favoring one perspective over others. This selective sourcing may lead to skewed public opinions (Tan & Weaver, 2010), prompting audiences to prioritize the liberal narrative over the conservative one and contributing to a lack of understanding of alternative viewpoints. As a result, the moderate liberal bias may compromise the neutrality and objectivity of articles on Wikipedia, thereby diminishing public trust in the encyclopedia. Because Wikipedia attracts ideologically mixed readership (Faris et al., 2017; Robertson et al., 2018), some viewers, specifically from the conservative side, may perceive its content as biased and become less inclined to engage with the platform.

Beyond the impact on trust and engagement, the preference for liberal media sources may also contribute to the emergence of left-leaning agendas on Wikipedia, which in turn can

News Sources Reveal Political Bias

influence public perception. As a downstream information source, Wikipedia provides contextual and background knowledge on topics that capture public interest (Gildersleve et al., 2023). When individuals search for a politician online, the way Wikipedians frame that politician's profile can shape how audiences evaluate the official. In this sense, Wikipedia functions as part of a broader, decentralized network of actors involved in sense-making, contributing to the evolving landscape of digital journalism and enriching our understanding of how public opinion is formed.

Furthermore, this study advances understanding of how digital media influence peer-production communities. Our analysis suggests that Wikipedia's citation practices may rely on a shared pool of sources that lean moderately liberal, regardless of the political affiliation of the article's subject. This pattern likely stems from Wikipedia's structural dependence on mainstream media outlets, which often reproduce the biases of the broader digital news ecosystem. When certain perspectives—such as conservative viewpoints—are underrepresented in these outlets, Wikipedia may mirror those omissions or slants. Through the repeated citation of liberal-leaning sources, editors may transmit left-leaning perspectives and values into the construction of politicians' pages, thereby marginalizing conservative viewpoints in the editorial process. In short, the preference for liberal media sources shapes Wikipedia's journalistic practices and facilitates the transfer of digital media bias into the platform's processes of knowledge production.

Wikipedia primarily shapes interpretation rather than agenda setting, underscoring the normative importance of neutrality on the platform. Traditional media outlets determine which issues receive public attention, whereas Wikipedia helps audiences contextualize and understand issues they have already encountered elsewhere (Gildersleve et al., 2023). As a result, biased

News Sources Reveal Political Bias

content does not merely skew exposure but can affect how readers interpret and evaluate political actors and events. When content lacks balance and diversity, readers may be less able to make informed political judgments. In this context, deviations from neutrality are particularly detrimental in a digital journalism environment where Wikipedia serves as a reference point for sense-making rather than topic initiation.

Our finding also has implications for how digital journalists approach collaborative editing practices. First, it is essential to cite a diverse range of sources to enhance the objectivity of news articles. When a collaborative process relies too heavily on a single type of source—such as liberal-leaning media—it risks introducing systematic bias into the content. Therefore, digital journalists should actively seek a balanced range of source materials and include voices from different perspectives to counteract any possible one-sided tilt, such as including both liberal and conservative media sources. Second, digital journalists should maintain transparency in their citation practices, for instance by providing hyperlinks or reference lists at the end of news stories (De Maeyer, 2014). Such practices not only enable readers to verify the accuracy of claims but also foster accountability among authors, which can, in turn, encourage the use of more balanced sources.

We examined the editing process to better understand how political bias emerges by analyzing key aspects of crowd collaboration. Our findings indicate that higher workload diversity—where contributions are more evenly distributed among editors—is associated with a stronger liberal bias in the selection of news sources. In other words, Wikipedia articles in which editors contribute similar amounts of work tend to exhibit a left-leaning bias. This result appears to contradict prior research suggesting that high workload diversity enhances collaborative

News Sources Reveal Political Bias

outcomes and mitigates bias by incorporating heterogeneous perspectives and experiences (Robert & Romero, 2016). The discrepancy implies that workload diversity does not necessarily lead to ideological diversity in cited sources. Instead, Wikipedia editing may be carried out by a politically homogeneous group of contributors (Shi et al., 2019), whose similar levels of participation reinforce shared biases rather than offset them.

Our research also reveals that social influence may contribute to the emergence of political bias in Wikipedia profiles. Specifically, articles tend to show a stronger liberal bias when editors engage in longer and more intense discussions. This finding suggests that increased task-related communication does not always lead to a more balanced selection of citations. Instead, such communication may be shaped by social dynamics. When dominant voices prevail within the editorial network, they can influence others' perspectives, potentially steering the article away from a neutral stance (Becker et al., 2017). On Wikipedia, experienced editors may exert disproportionate influence over article development. If these individuals promote biased citations during discussions, social pressure may discourage less experienced editors from expressing dissenting views. Interestingly, our analysis did not identify a significant relationship between crowd size and political bias. This lack of a direct effect suggests that the influence of crowd size on collaborative outcomes may be more complex and contingent on other factors such as diversity of perspectives or communication dynamics (Kittur & Kraut, 2008; Robert & Romero, 2016).

Taken together, this paper reveals that political information on Wikipedia may exhibit bias, particularly in the choice of news sources. This bias is associated with both workload diversity and the intensity of communication in the process of collaborative editing. Quality

News Sources Reveal Political Bias

information plays a critical role in a well-functioning democracy because citizens rely on it to make political decisions. Unlike traditional media, Wikipedia is a popular knowledge source that invites individuals with varying political perspectives to contribute to its content. However, the nature of collaborative editing does not inherently guarantee a neutral perspective in its final products. It requires caution to engage in political deliberation based on the political information on Wikipedia. Furthermore, there is a need for more effective platform policies to guide the citation of media sources and reduce any inherent bias in the editing process.

Limitations and Suggestions for Future Research

News references, while distinct from the content itself, do not necessarily provide a comprehensive indication of the overall bias present within Wikipedia pages. Current methodologies to discern media bias from content include the application of supervised machine learning algorithms as suggested by Budak et al. (2016) and the analysis of the prevalence of ideological phrases in news articles, as discussed by Greenstein & Zhu (2012). The former approach is resource-intensive, demanding substantial human intervention for data labeling. The latter, on the other hand, focuses only on specific sentences and omits others, leading to a possibly imprecise estimation of content bias. Our study attempts to balance computational complexity and accuracy by inferring content bias through the examination of news source usage within Wikipedia articles. Although our findings reveal an imbalance in the ideological leanings of cited news sources, they do not directly confirm the presence of content bias within Wikipedia articles. Future research should therefore explore whether and how citation bias translates into actual content bias in the textual narratives of the platform.

News Sources Reveal Political Bias

In addition, our study cannot separate news bias from political bias on the platform. Ideally, we would compare political bias with news bias—as a baseline—to determine whether the editing process itself contributes to bias. However, there is no single universally accepted standard for how biased the news is. The absence of a reasonable baseline hinders our ability to assess the extent of political bias on Wikipedia.

Funding

No funding was received.

Disclosure Statement

The authors have no competing interests to declare.

References

- An, J., Cha, M., Gummadi, K., Crowcroft, J., & Quercia, D. (2012). Visualizing media bias through Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 6, No. 2, pp. 2-5).
- Aral, S. (2021). *The hype machine: How social media disrupts our elections, our economy, and our health--and how we must adapt*. Crown Currency.
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130-1132.
- Becker, J., Brackbill, D., & Centola, D. (2017). Network dynamics of social influence in the wisdom of crowds. *Proceedings of the national academy of sciences*, 114(26), E5070-E5076.
- Benkler, Y. (2006). *The wealth of networks: how social production transforms markets and freedom*. Yale University Press.
- Benkler, Y. (2016). Peer production and cooperation. In J. M. Bauer & M. Latzer (Eds.), *Handbook on the economics of the internet* (pp.91–119). Edward Elgar
- Brenan, M., & Stubbs, H. (2020). News Media Viewed as Biased but Crucial to Democracy. <https://news.gallup.com/poll/316574/news-media-viewed-biased-crucial-democracy.aspx>
- Bro, P., & Wallberg, F. (2014). Digital gatekeeping: News media versus social media. *Digital journalism*, 2(3), 446-454

News Sources Reveal Political Bias

- Budak, C., Goel, S., & Rao, J. M. (2016). Fair and balanced? Quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1), 250-271.
- Bürger, M., Schlögl, S., & Schmid-Petri, H. (2023). Conflict dynamics in collaborative knowledge production. A study of network gatekeeping on Wikipedia. *Social Networks*, 72, 13–21. <https://doi.org/10.1016/j.socnet.2022.08.002>
- Callahan, E. S., & Herring, S. C. (2011). Cultural bias in Wikipedia content on famous persons. *Journal of the American society for information science and technology*, 62(10), 1899-1915.
- De Oliveira, Stephanie, & Richard E Nisbett (2018). “Demographically Diverse Crowds Are Typically Not Much Wiser Than Homogeneous Crowds.” *Proceedings of the National Academy of Sciences of the United States of America*. 115.9 : 2066–2071. Web.
- De Maeyer, J. (2014). Citation needed: Investigating the use of hyperlinks to display sources in news stories. *Journalism Practice*, 8(5), 532-541.
- Denny, M. J., & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*, 26(2), 168-189.
- Entman, R. M. (2007). Framing bias: Media in the distribution of power. *Journal of communication*, 57(1), 163-173.
- Fairchild, G., De Silva, L., Del Valle, S. Y., & Segre, A. M. (2016). Eliciting Disease Data from Wikipedia Articles. *Online Journal of Public Health Informatics*, 8(1). <https://doi.org/10.5210/ojphi.v8i1.6526>
- Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., & Benkler, Y. (2017). *Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election*. Berkman Klein Center Research Publication, 6.
- Foote, J., Shaw, A., & Hill, B. M. (2023). Communication networks do not predict success in attempts at peer production. *Journal of Computer-Mediated Communication*, 28(3). <https://doi.org/10.1093/jcmc/zmad002>
- GeiB, S., Leidecker, M., & Roessing, T. (2016). The interplay between media-for-monitoring and media-for-searching: How news media trigger searches and edits in Wikipedia. *New Media & Society*, 18(11), 2740–2759. <https://doi.org/10.1177/1461444815600281>
- Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? Evidence from US daily newspapers. *Econometrica*, 78(1), 35-71.
- Gildersleve, P., Lambiotte, R., & Yasseri, T. (2023). Between news and history: identifying networked topics of collective attention on Wikipedia. *Journal of Computational Social Science*, 6(2), 845-875.
- González-Bailón, S., Lazer, D., Barberá, P., Zhang, M., Allcott, H., Brown, T., ... & Tucker, J. A. (2023). Asymmetric ideological segregation in exposure to political news on Facebook. *Science*, 381(6656), 392-398.
- Graham, M., Straumann, R. K., & Hogan, B. (2015). Digital Divisions of Labor and Informational Magnetism: Mapping Participation in Wikipedia. *Annals of the Association of American Geographers*, 105(6), 1158–1178. <https://doi.org/10.1080/00045608.2015.1072791>
- Greenstein, S., & Zhu, F. (2012). Is Wikipedia Biased?. *American Economic Review*, 102(3), 343-348.

News Sources Reveal Political Bias

- Greenstein, S., Gu, Y., & Zhu, F. (2016). Ideological segregation among online collaborators: evidence from Wikipedians (No. w22744). *National Bureau of Economic Research*.
- Groeling, T. (2013). Media bias by the numbers: Challenges and opportunities in the empirical study of partisan news. *Annual Review of Political Science*, 16(1), 129–151.
- Groseclose, T., & Milyo, J. (2005). A measure of media bias. *The quarterly journal of economics*, 120(4), 1191–1237.
- Guth, K. L., & Brabham, D. C. (2017). Finding the diamond in the rough: Exploring communication and platform in crowdsourcing performance. *Communication Monographs*, 84(4), 510–533.
- Hamborg, F., Zhukova, A., & Gipp, B. (2019). Automated identification of media bias by word choice and labeling in news articles. In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (pp. 196–205). IEEE.
- Hart, P. S., Chinn, S., & Soroka, S. (2020). Politicization and polarization in COVID-19 news coverage. *Science communication*, 42(5), 679–697.
<https://doi.org/10.1177/1075547020950735>
- Haselmayer, M., Wagner, M., & Meyer, T. M. (2017). Partisan bias in message selection: Media gatekeeping of party press releases. *Political Communication*, 34(3), 367–384.
<https://doi.org/10.1080/10584609.2016.1265619>
- Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, 101(46), 16385–16389.
- Hong H., Ye Q., Du Q., Wang G. A., Fan W. (2020). Crowd characteristics and crowd wisdom: Evidence from an online investment community. *Journal of the Association for Information Science and Technology*, 71, 423–435. Crossref
- Horwitz, S.K., & Horwitz, I.B. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management*, 33, 987–101
- Keegan, B. C. (2013). A history of newswork on Wikipedia. *Proceedings of the 9th International Symposium on Open Collaboration*, 1–10.
<https://doi.org/10.1145/2491055.2491062>
- Keegan, B., & Gergle, D. (2010). Egalitarians at the gate: One-sided gatekeeping practices in social media. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 131–134).
- Kerkhof, A., & Münster, J. (2019). Detecting coverage bias in user-generated content. *Journal of Media Economics*, 32(3–4), 99–130. <https://doi.org/10.1080/08997764.2021.1903168>
- Kim, E., Lelkes, Y., & McCrain, J. (2022). Measuring dynamic media bias. *Proceedings of the National Academy of Sciences - PNAS*, 119(32), 1–3.
<https://doi.org/10.1073/pnas.2202197119>
- Kittur, A., & Kraut, R.E. (2008). Harnessing the wisdom of crowds in Wikipedia: Quality through coordination. *Proceedings of the 2008 conference on computer supported cooperative work* (pp. 37–46). New York, NY: ACM.
- Kittur, A., Suh, B., Pendleton, B., & Chi, E. (2007). He says, she says: conflict and coordination in Wikipedia. Conference on Human Factors in Computing Systems: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*; 28 Apr.-03 May 2007, 453–462. <https://doi.org/10.1145/1240624.1240698>

News Sources Reveal Political Bias

- Koerner, J. (2020). Wikipedia Has a Bias Problem. In *Wikipedia @ 20: Stories of an incomplete revolution* (pp. 322-332). MA: MIT Press.
- Langrock, I., & González-Bailón, S. (2022). The Gender Divide in Wikipedia: Quantifying and Assessing the Impact of Two Feminist Interventions. *Journal of Communication*, 72(3), 297-321.
- Levendusky, M. (2013). *How partisan media polarize America*. University of Chicago Press.
- Lin, Y. R., Bagrow, J., & Lazer, D. (2011). More voices than ever? quantifying media bias in networks. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 5, No. 1, pp. 193-200).
- Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the national academy of sciences*, 108(22), 9020-9025.
- Meraz, S. (2009). Is there an elite hold? Traditional media to social media agenda setting influence in blog networks. *Journal of computer-mediated communication*, 14(3), 682-707.
- Moore, R. C., Dahlke, R., & Hancock, J. T. (2023). Exposure to untrustworthy websites in the 2020 US election. *Nature Human Behaviour*, 7(7), 1096-1105.
- Morgan, J., Gilbert, M., Zachry, M., & McDonald, D. (2013). A content analysis of wikiproject discussions: toward a typology of coordination language used by virtual teams. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion*, 231-234. <https://doi.org/10.1145/2441955.2442011>
- Mutz, D. C. (2006). *Hearing the other side: deliberative versus participatory democracy*. Cambridge University Press.
- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- Puyu, Y., Colavizza, G., & Yang, P. (2024). Polarization and reliability of news sources in Wikipedia. *Online Information Review*, 48(5), 908-925. <https://doi.org/10.1108/OIR-02-2023-0084>
- Pradel, F. (2021). Biased representation of politicians in Google and Wikipedia search? The joint effect of party identity, gender identity and elections. *Political Communication*, 38(4), 447-478.
- Reagle, J. M., & Koerner, J. L. (Eds) (2020). *Wikipedia @ 20: Stories of an incomplete revolution*. Cambridge, MA: MIT Press.
- Reagle, J., & Rhue, L. (2011). Gender bias in Wikipedia and Britannica. *International Journal of Communication*, 5, 21.
- Ren, R., & Xu, J. (2023). It's not an encyclopedia, it's a market of agendas: Decentralized agenda networks between Wikipedia and global news media from 2015 to 2020. *New Media & Society*, 26(11), 6235-6259. <https://doi.org/10.1177/14614448221149641>
- Robert, L. P., & Romero, D. M. (2017). The influence of diversity and experience on the effects of crowd size. *Journal of the Association for Information Science and Technology*, 68(2), 321-332. <https://doi.org/10.1002/asi.23653>
- Robertson, R. E., Jiang, S., Joseph, K., Friedland, L., Lazer, D., & Wilson, C. (2018). Auditing partisan audience bias within google search. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1-22.

News Sources Reveal Political Bias

- Schäfer, M. S., & O'Neill, S. (2017). *Frame Analysis in Climate Change Communication: Approaches for Assessing Journalists' Minds, Online Communication and Media Portrayals*. Oxford University Press. <https://doi.org/10.5167/uzh-148182>
- Shi, F., Teplitskiy, M., Duede, E., & Evans, J. A. (2019). The wisdom of polarized crowds. *Nature human behaviour*, 3(4), 329-336.
- Social Media and News Fact Sheet. 2022. <https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/>
- Spiteri, J. (2021). Media bias exposure and the incidence of COVID-19 in the USA. *BMJ Global Health*, 6(e006798), 1–9. <https://doi.org/10.1136/bmjgh-2021-006798>
- Tan, Y., & Weaver, D. H. (2010). Media Bias, Public Opinion, and Policy Liberalism from 1956 to 2004: A Second-Level Agenda-Setting Study. *Mass Communication & Society*, 13(4), 412–434. <https://doi.org/10.1080/15205430903308476>
- Tripodi, F. (2023). Ms. Categorized: Gender, notability, and inequality on Wikipedia. *New Media & Society*, 25(7), 1687–1707. <https://doi.org/10.1177/14614448211023772>
- Top Websites Ranking. 2023. <https://www.similarweb.com/top-websites/>
- Twyman, M., Keegan, B. C., & Shaw, A. (2016). Black Lives Matter in Wikipedia: Collaboration and Collective Memory around Online Social Movements. *arXiv.Org*. <https://doi.org/10.48550/arxiv.1611.01257>
- Wang, Y., Kim, S. J., Shan, Y., Sun, Y., Jiang, X., Lee, H., ... & Shah, D. (2024). Slant, Extremity, and Diversity: How the Shape of News Use Explains Electoral Judgments and Confidence. *Public Opinion Quarterly*, 88(SI), 708-734.
- Walter, V., Kölle, M., & Collmar, D. (2022). Measuring the Wisdom of the Crowd: How Many is Enough? *Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 90(3), 269–291. <https://doi.org/10.1007/s41064-022-00202-2>
- Wikipedia.(n.d.). “Neutral point of view.” https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view
- Wikipedia: Prime objective. (n.d.). [https://en.wikipedia.org/wiki/Wikipedia:Prime_objective#:~:text=reading%20\(external%20links\)-,Vision%20statement,what%20we're%20doing.%22](https://en.wikipedia.org/wiki/Wikipedia:Prime_objective#:~:text=reading%20(external%20links)-,Vision%20statement,what%20we're%20doing.%22)
- Wikipedia.org.(2023). <https://www.similarweb.com/website/wikipedia.org/#overview>
- Wojcieszak, M., de Leeuw, S., Menchen-Trevino, E., Lee, S., Huang-Isherwood, K. M., & Weeks, B. (2023). No polarization from partisan news: Over-time evidence from trace data. *The International Journal of Press/Politics*, 28(3), 601-626.
- Yan, B., Jian, L., Ren, R., Fulk, J., Sidnam-Mauch, E., & Monge, P. (2021). The Paradox of Interaction: Communication Network Centralization, Shared Task Experience, and the Wisdom of Crowds in Online Crowdsourcing Communities. *Communication Research*, 48(6), 796–818. <https://doi.org/10.1177/0093650220915033>
- Yasseri, T., & Bright, J. (2016). Wikipedia traffic data and electoral prediction: towards theoretically informed models. *EPJ Data Science*, 5(1), 22.

Appendix

Table 1 | OLS Regressions between Political Bias and Crowd Editing

	Political bias B (SE)
Intercept	−0.1825*** (.007)
Workload diversity	−0.009*(.004)
Communication intensity	−0.007*(.003)
Crowd size	−0.002 (.006)
Covariates	
Job positions (Senators=high)	0.027** (.010)
Party affiliation (Rep=high)	0.059***(.007)
Age	−0.02***(.004)
Gender	−0.01 (.008)
Character count	0.002 (.01)
Page quality	−0.005 (.009)
Pageviews	0.003 (.004)
Reference count	−0.002 (.001)
Adjusted R2	0.138
F	(12, 857) = 12.57

Notes:

1. Three independents were excluded as they did not have party affiliations.
2. * $p < .05$, ** $p < .01$, *** $p < .001$.
3. To represent gender, party affiliation, and job positions, we used dummy variables. Specifically, we assigned 0 to females and 1 to males for gender, 0 to Democrats and 1 to Republicans for party affiliation, and 0 to representatives and 1 to senators for job positions. In addition, we categorized page quality into six levels: Stub (very basic description), Start (incomplete description), C (missing substantial content), B (relatively good articles), GA (good articles), and FA (featured articles). We created a dummy variable by assigning 0 to the categories of Stub, Start, and C, and 1 to B, GA, and FA. We assigned null values a 0 because the average category for page quality was C which we encoded as zero.
4. We standardized the independent variables using z scores, ensuring that the regression coefficients across variables were comparable on the same baseline.