Unveiling the hidden agenda: Biases in news reporting and consumption

Alessandro Galeazzi [ID][a,b,1], Antonio Peruzzi [ID][c,1], Emanuele Brugnoli [ID][d,e], Marco Delmastro [ID][a,e] and Fabiana Zollo [ID][a,f,*]

[a]Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, 30172 Venice, Italy
[b]Department of Mathematics, University of Padova, 35121 Padova, Italy
[c]Department of Economics, Ca' Foscari University of Venice, 30121 Venice, Italy
[d]Sony Computer Science Laboratories, 00184 Rome, Italy
[e]Centro Ricerche "Enrico Fermi" (CREF), 00184 Rome, Italy
[f]The New Institute Centre for Environmental Humanities, 30123 Venice, Italy
*To whom correspondence should be addressed: Email: fabiana.zollo@unive.it
[1]A.G. and A.P. contributed equally to this work.
Edited By Katherine Ognyanova

Abstract

Recognizing the presence and impact of news outlets' biases on public discourse is a crucial challenge. Biased news significantly shapes how individuals perceive events, potentially jeopardizing public and individual wellbeing. In assessing news outlet reliability, the focus has predominantly centered on narrative bias, sidelining other biases such as selecting events favoring specific perspectives (selection bias). Leveraging machine learning techniques, we have compiled a six-year dataset of articles related to vaccines, categorizing them based on narrative and event types. Employing a Bayesian latent space model, we quantify both selection and narrative biases in news outlets. Results show third-party assessments align with narrative bias but struggle to identify selection bias accurately. Moreover, extreme and negative perspectives attract more attention, and consumption analysis unveils shared audiences among ideologically similar outlets, suggesting an echo chamber structure. Quantifying news outlets' selection bias is crucial for ensuring a comprehensive representation of global events in online debates.

Keywords: media biases, social media, news consumption, news production, misinformation

---

Significance Statement

We thoroughly investigate news biases by analyzing the entire information chain, from the selection of newsworthy events to news consumption, focusing on the often-overlooked selection bias. Using machine learning, we classify six years of news coverage on vaccines and input these data into a Bayesian model to measure selection and narrative biases. Our results demonstrate that third-party reliability classification primarily considers the narrative conveyed by news outlets, neglecting biases in the editorial selection of newsworthy events. Additionally, we analyze the engagement these outlets receive, showing that extreme positions tend to attract more attention. Furthermore, our analysis of news consumption patterns reveals a higher audience similarity among news outlets with similar ideological stances, underscoring the interconnected nature of the information chain.

Introduction

In the public sphere, several perspectives and opinions are frequently exchanged and discussed, leading to the emergence of diverse views and understanding (1). With the advent of social media, a significant portion of the public debate has shifted online (2), where news and views are disseminated through fragmented and ongoing conversations.

The impact of information environments in shaping public opinion has been widely investigated, with researchers addressing different aspects, from the dynamic and consequence of the spread of misinformation (3–9) to the role of suggestion algorithms (10, 11) and the ideological biases of new outlets (12–14).

News coverage can become distorted through two main categories of choices by newsmakers: which events to cover, and how to cover them (15). We refer to systematically distorted decisions in these areas as *selection bias* (also called gatekeeping bias, filtering bias, or agenda bias) and *narrative bias* (or presentation bias, statement bias), respectively.

Selection bias refers to the tendency of a news outlet to choose certain events to cover while ignoring others. For instance, a news

Selection bias	Narrative bias	Perception	Echo chambers
I. Gatekeeping	*II. News reporting*	*III. Engagement*	*IV. Consumption*

Fig. 1. Biases in the information chain.

outlet may focus on adverse events related to vaccinations, while neglecting positive ones, thereby exhibiting a selection bias towards negative coverage. Narrative bias refers to the way in which news events are framed and reported, potentially influencing the reader's perception and interpretation of the events themselves. Together, these two forms of bias can significantly shape the way in which news is consumed and understood by the public.

The former regards the choice of what information will be selected to be marketed as news and what stories will be "deselected" (16). This is the single most fundamental decision in journalism, and it is the most pervasive form of bias (17). However, it is also the least studied form of bias, probably due to the additional research and analysis effort required to examine a news outlet's coverage of all relevant events related to a given topic (18), which needs to include an unobserved population in the analysis. Consequently, the literature has primarily focused on the narration of events (19–21), leaving the selection step of the news production process understudied.

In this work, we introduce a methodological framework to address this issue and place it within the broader context of public discussion. First, by focusing on the highly polarized and debated topic of vaccines and immunization, we consider all vaccine-related events reported by a comprehensive selection of Italian news outlets. This ensures our sample is as broad as possible and covers all events deemed newsworthy by any of these sources. The large sample size of news outlets and the plurality of viewpoints considered minimize the probability of overlooking significant newsworthy events.

Our comparative approach does not focus on quantifying the universe of vaccine-related events but rather on detecting differences in their selection among news outlets and their dependence on the classification provided by third-party organizations. Indeed, the classification of news outlets based on their reliability and ideological biases has been extensively used in social media studies to analyze different aspects of online debates (10, 22, 23). Such classification is provided by independent fact-checking organizations (such as MediaBiasFactCheck (24), AllSides (25), or Journalism Trust Initiative (26)) which rates news producers based on journalistic criteria.

Following the literature on media bias (15), events covered in the collected vaccine-related news are categorized as adverse, neutral, or positive based on the harm or benefit they bring to vaccination efforts. For example, negative events may involve the emergence of adverse effects, neutral events may cover periodic data from the vaccination campaign, and positive events may include the discovery of a highly effective new vaccine.

The second step involves classifying articles based on how news outlets report these newsworthy events. The collected vaccine-related news is thus classified according to the narrative conveyed on the subject discussed (27). The categories used are: antivax, neutral, and provax. This classification depends on whether the article emphasizes only negative elements of the news (antivax), whether true or not, provides a balanced account (neutral), or highlights only positive elements (provax), whether true or not. For example, a news article that exaggerates the

magnitude of adverse effects, neglects vaccine effectiveness, or suggests causal relationships between vaccination and adverse effects without evidence is labeled as antivax. A piece of news reporting statistics on vaccine coverage, effectiveness, and adverse effects is considered neutral. In contrast, a news article describing the effectiveness of a new vaccine with sensationalized or emotionally loaded words, exaggerating the beneficial consequences of the event, is labeled as provax.

To provide a comprehensive view of the public debate, it is essential to examine consumer behavior in how they perceive and consume information generated by the news media. Indeed, social media platforms introduce a feedback loop between news producers and consumers. Therefore, our study also investigates engagement and news consumption in relation to selection and narrative biases.

The framework of our analysis is depicted in Figure 1, which summarizes the various phases of the analysis and highlights the possible sources of biases within the information chain: starting from the selection of newsworthy events (I), verifying the narrative of the corresponding news disseminated by information sources (II), studying the engagement they generate among users (III), and examining their consumption by homophilic user groups (IV).

As a case study, we analyze all the news about vaccines produced and disseminated on social media by nearly all Italian news outlets. While we use the Italian vaccine debate as a case study, our framework provides a flexible methodological approach applicable to any context, topic, and language for analyzing the existence and interplay of narrative and selection biases. Nevertheless, given the broad interest in this topic, its social significance, and the high degree of polarization, vaccines serve as an ideal subject to test our methodology. In particular, we analyze the intense vaccine debate that occurred in Italy during the six-year period from 2016 to 2021, which attracted a wide range of views and highly polarized public opinion to the point where politicians from competing parties publicly took opposing positions. During this time frame, the debate focused initially on the design, approval, and enforcement of the legislative framework on mandatory pediatric vaccinations (i.e. Law n.119 of July 31 (28)), which was introduced in mid-2017 and fully implemented only in September 2019. Later on, the debate shifted to COVID-19 vaccines following the onset of the pandemic starting in January 2020.

We compile a comprehensive sample of news providers active in Italy, by relying on assessments provided by well-known organizations, namely NewsGuard, Facta, Pagella Politica, and Butac. Note that Newsguard alone claims to monitor domains covering about 95% of online engagement with news sites (29). These assessments also serve to distinguish between outlets that are *questionable* (i.e. source producing mainly unverified or false content) and those that are *reliable*. Then, we focus on these news outlets to gather all vaccine-related content published on the major social media platforms (Facebook, Instagram, Twitter, and YouTube) during the specified period, ensuring a plurality of viewpoints and thereby minimizing the probability of overlooking significant vaccine-related events.

Given the large and comprehensive set of articles representing the Italian vaccine debate during 2016–2021, we build machine-learning models capable of accurately classifying the collected content based on the nature of the event being discussed (adverse, neutral, positive) and the narrative being conveyed (antivax, neutral, provax), respectively. This approach provides massive scalability, avoiding the need for a sample selection and extensive manual annotation, which could introduce a bias, and make it suitable for use in many diverse frameworks.

In sum, we consider the news and views on vaccines produced by a diverse and comprehensive set of Italian news outlets that represent the most popular web sources with the largest reach on a wide range of social media platforms (Facebook, Instagram, Twitter, and YouTube). We rely on the assessment granted by independent fact-checking organizations to differentiate between *questionable* and *reliable* sources. Moreover, we exploit our machine learning model to classify news articles based on the narrative conveyed (*antivax*, *neutral*, *provax*) and the nature of the event being discussed (*adverse*, *neutral*, *positive*). Using our classification models, we quantitatively measure and analyze how selection and narrative biases shape the production and consumption of vaccine news.

Hence, our contribution to the literature is three-fold: (i) First, we propose a methodology to measure not only narrative bias but also selection bias and compare them to the reliability classification provided by third parties. (ii) Second, we analyze the relationship between these news-related biases and the online environment considering the spreading patterns and the engagement generated by the news outlets. (iii) Third, we provide a method to evaluate whether the ratings provided by external entities suffer from ideological bias. In other words, we pose the fundamental issue that can be epitomized in the Latin phrase of the Roman poet Juvenal: "Quis custodiet ipsos custodes?" (i.e. Who will guard the guardians?).

Our findings indicate that the classifications performed by third-party entities predominantly align with the narrative bias dimension. However, they exhibit reduced accuracy in assessing the reliability of news outlets based on selection bias and are skewed towards a provax narrative and a selection of positive events. We showed that highly biased news outlets tend to generate greater engagement. Moreover, the consumption patterns analysis revealed a significant overlap in audience among outlets with similar biases, hinting at the presence of an echo chamber effect.

Results
Selection bias

We start our analysis by addressing the issue of selection bias, which takes precedence over other forms of informational distortion. This bias plays a pivotal role in determining the pool of newsworthy events that are not reported and narrated by news outlets. At its core, selection bias molds the perception of what qualifies as noteworthy, thereby shaping the subsequent narrative presented to the public. This inherent bias in the selection process can exert a profound influence on the public's understanding of reality, leading to the over-representation or under-representation of certain events and contributing to a skewed worldview. Recognizing and mitigating selection bias is imperative for cultivating a more accurate and impartial portrayal of events within the media landscape.

The methodology for measuring news outlets' selection bias involved first identifying all events covered by at least one outlet during the reference period. Since the sample includes all national news outlets and most local ones, the analysis comprehensively captures all relevant events related to vaccinations without evident structural distortions in event coverage. In other words, incorporating sources with opposing viewpoints and diverse editorial strategies ensured a set of newsworthy events that fairly represents all types of events. Once identified, these events were classified—using the outlined machine learning methodology—into positive, neutral, or adverse categories. Each news source was then linked to the events it reported. A news outlet is considered to have selection bias (either positive or adverse) if it systematically emphasizes a particular type of event in a way that diverges from the overall distribution of newsworthy events. This bias can distort the representation of reality by disproportionately emphasizing certain types of events, even if the outlet's reporting remains neutral in tone.

This bias can skew the representation of the world by over-representing certain event types, even if the outlet's reporting style remains neutral.

To quantify this bias, i.e. the tendency of a news outlet to over- or under-discuss one type of newsworthy event, we fit a latent space model (30, 31) that estimates latent factor conveying information about the news outlets' narrative when dealing with adverse and positive events. This model takes as input the publishing behavior (i.e. information about the narratives and types of published articles) of news outlets and maps each outlet on a latent dimension representing the narrative stance (see section Materials and Methods). We set up our model so that the more negative (positive) the latent factor the stronger the antivax (provax) narrative for both adverse and positive events. Then, we use the outlet-specific intercepts of the model (as described in section Materials and Methods) to quantify the selection bias. In our model, the intercept parameters $\alpha_{i,k=Adv}$ and $\alpha_{i,k=Pos}$, obtained as a byproduct of the latent narrative estimation, represent the propensity of news outlet i to report on adverse ($k = Adv$) and positive events ($k = Pos$), respectively. The higher α_{ik}, the stronger the propensity to report on that type of event.

To quantify how balanced a news outlet's reporting is between positive and negative events, we have defined a *Selection Index* (see section Materials and Methods). This index is calculated by measuring the distance between the news outlet's position and the diagonal in the adverse-positive propensity factor plane. The farther a point is from the diagonal, the more the corresponding news outlet favors one type of events in its articles. It is important to note that selection bias requires joint consideration of the propensity to write about positive and negative events in order to have a single index that provides information on the imbalance of reporting.

Figure 2A displays the propensity factors for positive and adverse events. The 45-degree dashed line represents the perfect balance between the propensity to report on positive and negative events. The figure shows that questionable news outlets have a strong propensity to report on adverse events and a weak propensity to report on positive events. The vast majority of reliable outlets show a mostly balanced approach in their reporting, with a slight preference for positive events on average. Noticeably, a small fraction of reliable outlets, known for their strong pro-science position, exhibit a strong propensity to report on positive events and a weak propensity to report on adverse events. The different propensities of questionable and reliable news outlets in reporting positive and adverse events are confirmed by the distributions of their distances from the balanced-selection line and the angles in polar coordinates, as depicted in Figure 2B and
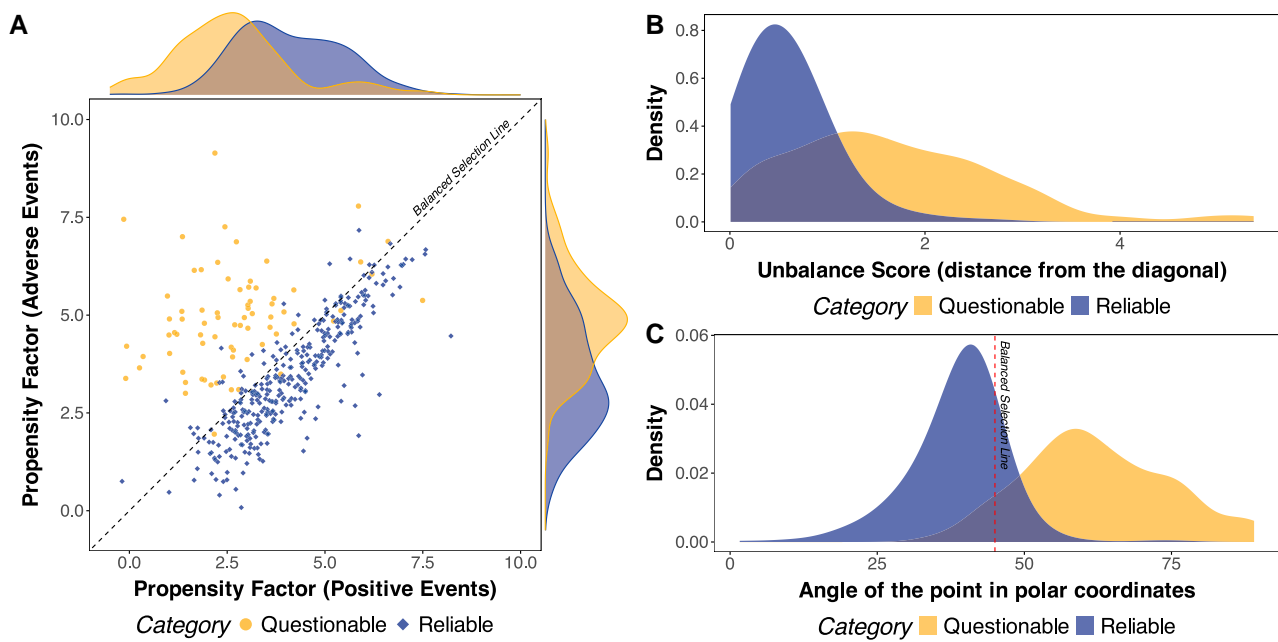
Fig. 2. Estimated news outlets' propensity to report on positive events against adverse events (left) and distributions of distances and angles of the point from the balanced selection line (right). The 45-degree line represents the set of all points showing a balanced selection of news, i.e. equal propensity of reporting on positive and adverse events.

C, respectively. Results show that reliable outlets tend to be less unbalanced than questionable ones, since the distribution of reliable news outlets in Figure 2B peaks closer to 0 with respect to the questionable one. Moreover, Figure 2C shows how most of the reliable news outlets lie below the angle of the balanced-selection line (45 degrees), implying a propensity to report more on positive than negative events. Conversely, most questionable outlets are placed above the line, thus indicating a propensity to report more on negative than positive events. Our analysis underscores a significant disparity in the selection process of newsworthy events between questionable and reliable news outlets. However, it also shows that both reliable and questionable news outlets are influenced by selection bias, albeit to varying degrees.

Narrative bias

Based on the set of selected newsworthy events, information sources determine the narrative through which these events are reported. The process involves a deliberate choice in framing and presenting the events, shaping the way they are perceived by the audience.

We quantify the narrative bias of each source by exploiting the distribution of the narratives (antivax, neutral, and provax) of the articles released by the source on a given type of event. The more a news outlet adheres to one narrative (the closer its latent position to the ideal points corresponding to one of the three narratives) the higher the bias in favor of that narrative (see section Materials and Methods).

Figure 3 shows the results of the latent factor estimation, where each dot represents the coordinates of a news outlet's narrative when dealing with positive (*x*-axis) or adverse (*y*-axis) events. Each dot is colored according to the reliability of the news outlet, as derived from third-party data (see section Materials and Methods for further details).

As shown in Figure 3, the estimated narrative bias aligns well with third-party classifications of news outlets as reliable or questionable.

Indeed, the distinction between questionable and reliable outlets is reflected by their differing reporting styles. Questionable outlets tend to have a negative stance when reporting on both positive and adverse events, while reliable outlets have a milder position when reporting on adverse events and are more positive when reporting on positive events. This can be also observed in the different modes of the marginal distributions. Noticeably, the set of reliable outlets includes not only those with moderate positions, but also some of those with a strongly positive narrative. Through manual inspection, we find the presence of news outlets historically known for a strong conspiracy component in the bottom-left corner of the plot, while those with a strong pro-science position are located in the top-right corner, confirming the soundness of our estimations.

Interplay between selection and narrative biases

After analyzing selection bias and narrative bias separately, we focus on the interplay between them. We address this question by studying the dependence between propensity values, from which we compute selection bias, and narrative values computed with the model. Figure 4 shows the scatterplot of narrative vs propensity for the three types of events. The results reveal some interesting insights. First, there is a significant yet moderate correlation between the two types of news production biases (i.e. propensity and narrative), which is positive for positive events (Pearson's coefficient: 0.420, P-value < 0.001), negative for adverse events (Pearson's coefficient: -0.470, P-value < 0.001), and weakly positive for neutral events (Pearson's coefficient: 0.269, P-value < 0.001). This means that, although correlated, a news source more inclined to report positive events is likely to have a provax narrative, and one that focuses on negative events
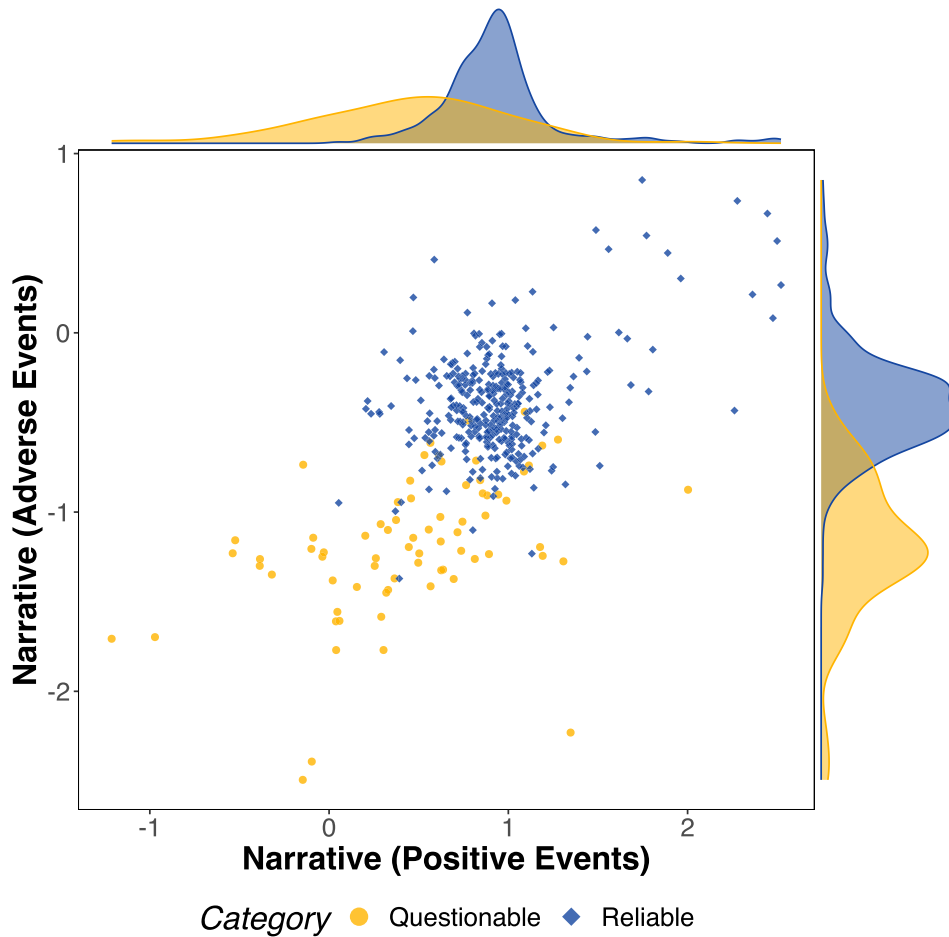
Fig. 3. News outlets' narrative bias in reporting positive events compared to their estimated stance in reporting adverse events, as estimated by the Latent Space Bayesian Model. Points are colored according to the classification retrieved from third-party data. The asymmetry in axis values is due to different framing strategies adopted when reporting events of different natures (positive or negative).

tends to have an antivax narrative. However, these are two distinct stages in the information chain that do not overlap. Moreover, when we look at questionable and reliable sources separately, they show very different editorial approaches: questionable sources have stronger correlations (Pearson's coefficient: 0.460, −0.514, 0.429, P-value < 0.001), whereas reliable sources show weaker correlations (Pearson's coefficient: 0.266, −0.274, −0.198, P-value < 0.001), indicating that for questionable sources these two stages are more intertwined.

Biases and engagement

In the previous sections, we examined the behavior of news outlets and emphasized the biases present in their published content. However, online social media offers us the opportunity to further analyze how content is received and interpreted by the public. Our next objective is to investigate the relationship between the strategies adopted by news outlets in terms of narrative and selection bias, and their level of user engagement. To do this, we must control for any scaling effects that may be present due to well-established news outlets having a larger audience and more resources for coverage. Therefore, we define an adjusted measure of engagement $E(s; k; T)$:

$$E(s; k; T) = \frac{I(s; k; T)}{C(s; k; T) \cdot F(s; T)},$$

where $C(s; k; T)$ denotes the number of contents (i.e. articles) published by the news outlet $s \in S$ on events of type $k \in \{Adv, Neu, Pos\}$ in the time span T and $I(s; k; T)$ represents the corresponding number of user interactions (e.g. likes, shares, comments) received by s on articles about events of type k in the time span T, while $F(s; T)$ represents the average number of followers of the social media accounts of news outlet s that were active during T.

To gain insight into the relationship between engagement metrics and both narrative bias and selection bias proxies, we present scatter plots in Figure 5. The top panels display the relationship between the narrative bias factor and the engagement metric computed for adverse, neutral, and positive events, respectively. Similarly, the bottom panels show the relationship between the selection bias metric and the engagement metric. In both cases, the relationship appears strongly nonlinear. A U-shaped relationship emerges to exist between engagement and narrative bias, with more extreme narratives seeming to be associated with higher engagement, while lower engagement is associated with moderate positions about the topic. However, this relationship appears to be mostly driven by the fact that questionable outlets, characterized by a more negative outlook on the topic, are more successful in generating engagement (see also Figures S8–S10). Also, a convex relationship seems to be in place when considering the selection bias metric. We further investigate these relationships through linear regression, which confirms the existence of
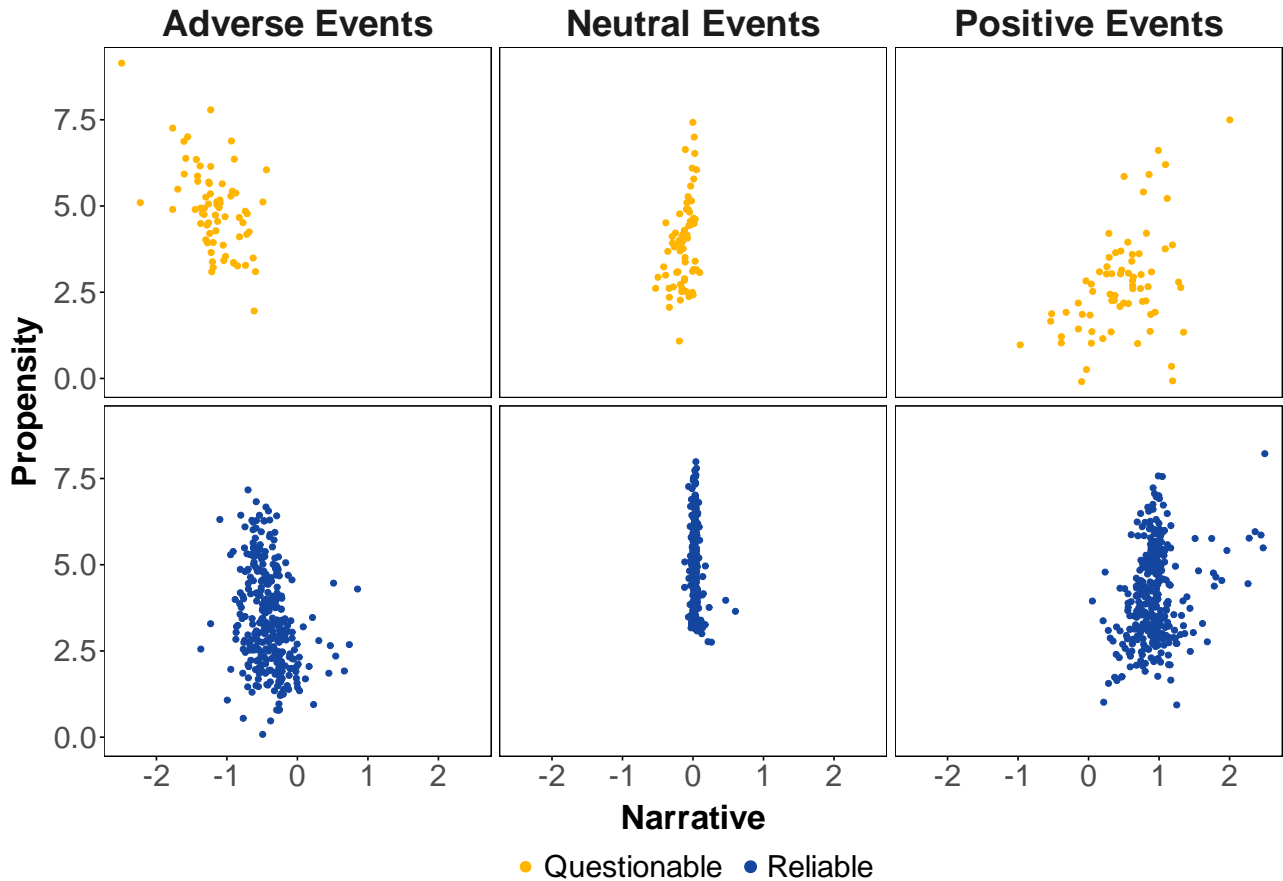
Fig. 4. Propensity vs narrative values for positive, adverse, and neutral events. Points represent news outlets' scores for narrative and propensity computed with the Latent Space Bayesian Model. Questionable outlets (top row) exhibit a moderate correlation (Pearson's coefficients: 0.460, −0.514, 0.429, *P*-value < 0.001) between propensity and narrative for all three types of events. In contrast, reliable outlets (bottom row) show weak correlations (Pearson's coefficients: 0.266, −0.274, −0.198, *P*-value < 0.001), suggesting that, for the latter, higher values of selection bias do not necessarily imply higher values of narrative bias.

a convex relationship between narrative and selection biases and engagement(see Table S2). Overall, the plots indicate that unbalanced reporting of facts could potentially boost engagement, particularly in the case of questionable news outlets that prominently promote a negative perspective on vaccines.

Biases and news consumption

The analysis reported in the previous section highlights the relationship between news outlets' biases and online engagement. A natural question is whether news outlets adopting similar publishing strategies are also consumed by the same users. To examine this, we analyze the problem from the perspective of news consumption, using Twitter data on the vaccine debate from January 2020 to December 2021 to study the similarity in the audience of different news outlets. We define a metric based on cosine similarity on retweeters to quantify the connection between news outlets (see section Materials and Methods). Intuitively, outlets sharing a high percentage of retweeters have a higher value of the similarity metric (close to 1), while outlets with only a few shared retweeters will have a low similarity (close to 0). Using this information, we build an undirected network in which nodes represent news outlets and weighted edges indicate the level of similarity. To highlight only the stronger connections, we discard edges with weights lower than the overall mean of the edges (see section Materials and Methods).

The resulting network is visualized in Figure 6 and shows that reliable news outlets dominate the debate on Twitter. Moreover, reliable outlets form the core of the network, while questionable ones have a more peripheral role, as highlighted by the percolation analysis reported in Figure S7. However, it is worth noting that there is no clear separation between questionable and reliable news outlets. This suggests that some users tend to retweet a set of only reliable or questionable news outlets, while others have a mixed news diet, sharing both types of outlets. This interplay between questionable and reliable outlets in the similarity network is further clarified by Figure 6B–D, where the percentage of questionable ones for each cluster detected using Louvain algorithm (32) is color-coded. Most of the clusters are primarily populated by reliable news outlets, with only one (Cluster 1) having more questionable (65%) than reliable outlets. Furthermore, Figure 6 also reports the average value of narrative bias for adverse events (panel B) positive events (panel C), and selection bias (panel D), highlighting the differences between questionable and reliable clusters. Indeed, the most questionable cluster (Cluster 1) has the lowest average narrative bias on adverse events, indicating that news outlets in this cluster tend to emphasize the magnitude of adverse events. At the same time, Cluster 1 also has the lowest narrative bias on positive events, implying that its news outlets are likely to minimize the impact of positive events. Notably, this cluster has the highest value for selection
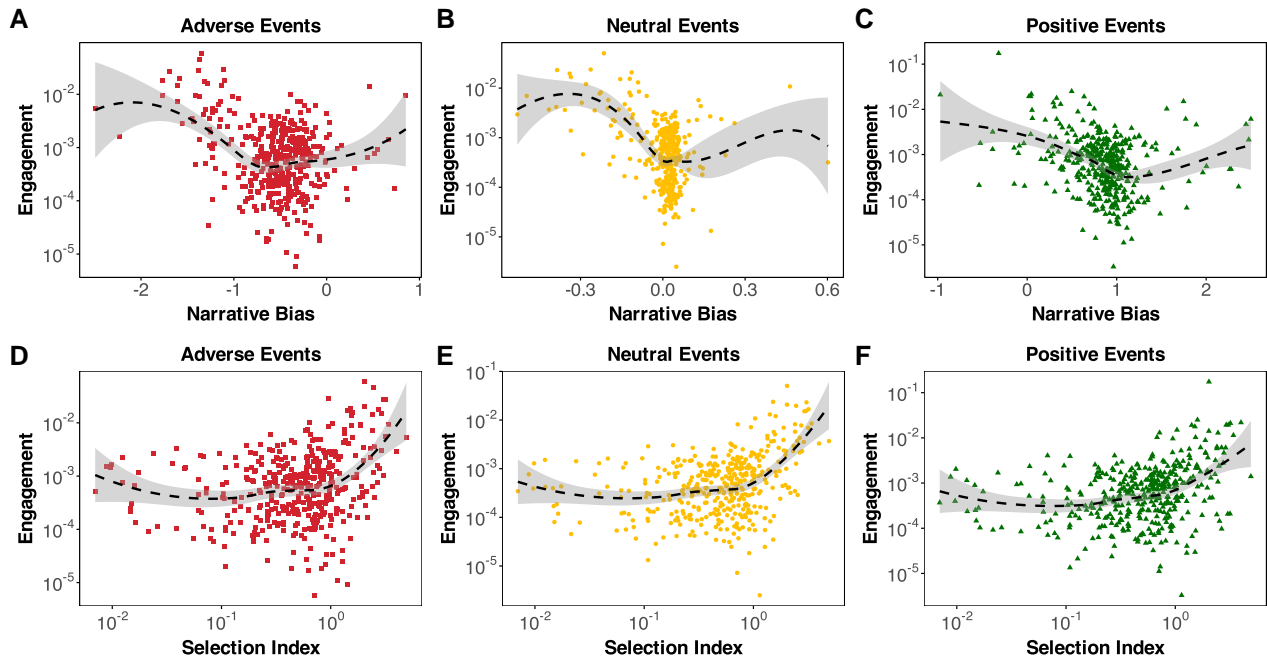
Fig. 5. Engagement vs narrative bias (top panels) and selection index (bottom panels) for adverse, neutral, and positive events. Engagement is measured by considering all interactions with content (reactions, shares, and comments) and adjusting for the size of each news outlet's account.
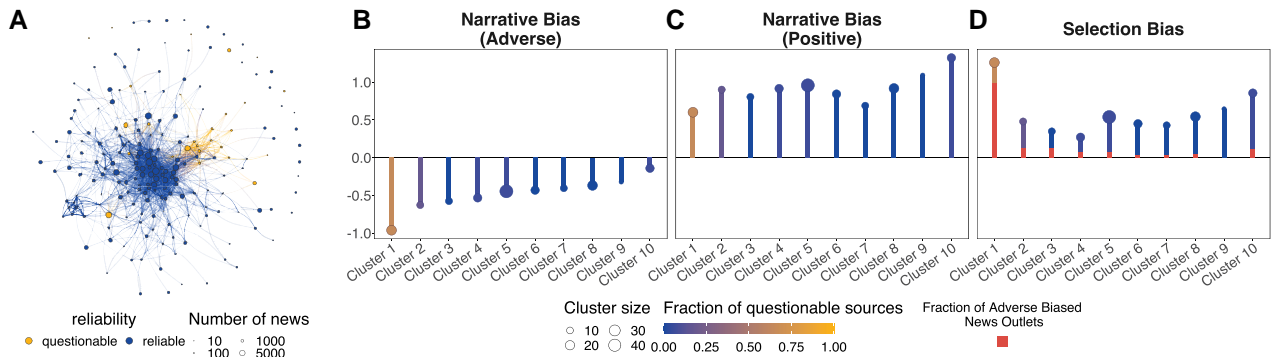
Fig. 6. Panel A displays the network of news outlets built on the retweeters' cosine similarity. The reliability of outlets is color-coded in the network, with questionable outlets marked in yellow and reliable ones in blue. Panels B and C show the average Narrative bias for adverse and positive events, respectively, while Panel D presents the average Selection bias across clusters identified using the Louvain algorithm on the news outlets network. Each cluster is color-coded to indicate the proportion of questionable outlets, with the size of each dot proportional to the cluster's size.

bias, indicating that its news outlets do not cover adverse and positive events equally. On the other hand, Cluster 10 exhibits the opposite behavior. It has the highest narrative bias values for both adverse and positive events, indicating that its news outlets tend to minimize the importance of adverse events and exaggerate the importance of positive events. We also notice that this cluster has the second highest value of selection bias, implying that these news outlets do not cover both types of news equally.

In summary, the information presented suggests that news outlets in the most questionable and reliable clusters (Cluster 1 and Cluster 10, respectively) present events from opposing perspectives: the first one strongly endorses an antivax narrative, while the second one firmly promotes vaccination. Moreover, both exhibit a high selection bias, meaning that they tend to select only the type of events that align with their narrative, indicating the presence of echo chambers. To further verify this, we computed the fraction of news outlets that have a selection bias

toward adverse events for each cluster, and showed it as a darker bar in panel D of Figure 6. Intuitively, this is equivalent to the fraction of points that lie upon the 45-degree line in the left panel of Figure 2 for each cluster. Cluster 1 has the highest fraction of news outlets biased towards adverse events by far(78%), while Cluster 10 has less than 13%. Finally, a manual inspection of the clusters' members revealed that Cluster 10 is composed of news outlets widely recognized as proscience, while Cluster 1 is populated by well-known antiscience and conspiracy-theory outlets.

Conclusion

In this study, we proposed a new method to analyze the presence of two different types of biases in the selection, production, and dissemination of news. Our approach considers both the selection of events by news outlets (selection bias) and how they are presented (narrative bias) to users. We exploited machine learning

techniques to classify the type (positive, neutral, or adverse) and narrative (provax, neutral, antivax) of Italian vaccine-related events reported by news outlets. We used this information to fit a Bayesian model and quantified the two biases through latent variables. Moreover, using data from fact-checking agencies, we classified news outlets based on their reliability (reliable and questionable). Finally, we analyzed the relationship between news outlets' biases, citizen engagement, and news consumption.

Results show that our method allows us for the quantification and assessment of the relevance of selection bias, whose assessment represents a more challenging task than narrative bias and is often neglected in the quantitative literature. The analysis also verifies the existence (or nonexistence) of an ideological bias in fact-checking organizations that evaluate the quality of the information selected, disseminated, and discussed in public debates. Further, results suggest that questionable news outlets with a prominent negative view of vaccines, both at the selection and narrative stages, attract more engagement, which aligns with previous research (33), and the presence of clusters of users consuming only one type of news, hinting to the presence of echo chambers (10). Ultimately, the article showed that there is a distinct and opposing informational chain between questionable sources and reliable ones, starting with how they select newsworthy events. Indeed, questionable sources, unlike reliable ones, tend to present a false yet consistent view of the world. This view, marked by high and overlapping biases, resonates with a dense and cohesive audience eager to spread it further.

The proposed methodology can be readily adapted to different domains by leveraging newly annotated datasets, ensuring its broad applicability. Moreover, the framework easily extends to topics where contrasting viewpoints (e.g. denialists and supporters of anthropogenic climate change, prochoice and prolife positions on abortion) are debated over extended periods. Crucially, our method depends on the representativeness of the collected news set. Selecting a broad range of sources that reflects the full spectrum of opinions (and events) is essential to accurately measure both selection and narrative biases. This can be effectively achieved by combining data from lists of news outlets, typically sourced from national databases, to cover the vast majority of the information landscape. All of the analyzed aspects—biases in the selection and narrative of newsworthy events, the impact of these on citizen reactions and news consumption patterns, and possible biases in the assignment of quality ratings by fact-checking organizations—are fundamental to the functioning of democratic societies. Undoubtedly, public media plays a pivotal role in shaping public opinion. Firstly, the agenda-setting theory underscores how public media influences the significance of topics within public discourse. Secondly, gatekeeping practices may reflect on the volume and diversity of information users encounter, potentially offering a limited perspective on reality. Lastly, significant bias in reporting factual events can distort reality, sometimes to the extent of completely aligning with the producers' narrative. Understanding the presence and the impact of these biases on public opinion and discourse is crucial for ensuring an informed and engaged citizenry, which is vital for the health of any democracy.

Materials and methods
News corpus data
We collected approximately 350K vaccine-related pieces of content published on Facebook, Instagram, Twitter, and YouTube from almost the entire universe of Italian news outlets in the

Table 1. Breakdown of the dataset.

Category	Outlets	Contents	Interactions
Questionable	161(23.6%)	44,547(12.6%)	10,898,774(11.4%)
Reliable	521(76.4%)	308,983(87.4%)	84,332,137(88.6%)
Total	682(100%)	353,530(100%)	95,230,911(100%)

6-year period from 2016 January 1 to 2021 December 31. This comprehensive picture of the Italian media landscape was retrieved by combining several lists from different third-party organizations, namely NewsGuard, Facta, Pagella Politica, and Butac.

The data collection process was carried out exclusively through the CrowdTangle API of Meta and the official APIs of Twitter and YouTube. The selected news outlets included a wide range of national/local newspapers, radio/TV channels, and online news outlets active in Italy during the aforementioned period, to ensure the most representative picture of both traditional and new media. Specifically, we selected 96 newspapers, 462 online-only news outlets, 89 TV channels, and 35 radio channels. Then, we focus on these 682 outlets to performed a keyword search for content that matched an exhaustive list of vaccine-related keywords, including general terms and vaccine brands/names (see Supplementary Material for the complete list of keywords).

News outlets were also assigned a binary label to distinguish between two categories: questionable—i.e. a source producing mainly unverified or false content—and reliable. This classification was retrieved from the lists provided by the aforementioned fact-checking organizations. It is worth noting that if a source was listed by more than one organization, there was a $\sim 100\%$ overlap in assigning the label of questionable/reliable (34). Table 1 shows a breakdown of the dataset with the number of news outlets, contents, and corresponding user interactions (understood as the algebraic sum of all possible actions/reactions performed on the four platforms analyzed). Notice that the dataset is the same as that used in (35); for further details, we refer the reader to that study.

Twitter data
To analyze the similarity in news outlet audiences, we used Twitter data on the vaccination and COVID-19 vaccines debate. We collected all tweets made by the accounts we considered in our analysis that contained a keyword related to vaccination (see Supplementary Material for the list of keywords used). We also retrieved all the retweets pointing to these tweets, obtaining a dataset of 23,908 tweets created by 315 news outlets and 254,965 retweets created by 53,074 users. Notice that not all news outlets in our list had an active Twitter account.

Modeling vaccine news narrative and event type
To classify the nature of the event reported (adverse, neutral, positive) and the narrative conveyed (antivax, neutral, or provax) by vaccine-related content, we followed Google's pretrained BERT multilingual cased model (36), which represents the state-of-the-art for semantic text representation in most languages (37), especially when data comes from social media (38, 39). The narrative model is the same one built in (35) by training the BERT model on a manually annotated set of vaccine-related content, representing $\sim 10\%$ of the data gathered. The sample was intentionally selected to contain anti- and provax narratives. Nonetheless, approximately half of the annotated data concerns neutral views. To make the model more balanced between narrative classes and more confident with the local space around

extreme values, augmented pieces of content (40) were added to the sample by inserting words in a selection of data annotated as antivax or provax through the contextual word embedding of BERT. The same sample was here further annotated with the nature of the event reported and then used to fine-tune the BERT model for the corresponding classification task. The data to annotate were split among the authors to get $\sim 20\%$ overlap to compare the annotator agreement results with the model performance. The augmented dataset was split into two parts to produce a dataset for training (80%) and a dataset for evaluating (20%) the model, by ensuring on both sets comparable class distributions with respect to narratives and events. To ensure proper model evaluation, neither the annotated content used as a basis for the augmentation nor the augmented content were included in the evaluation set. The annotation results with respect to the narrative and event for the training and evaluation sets are summarized in Supplementary Material, where we also provide examples of annotations covering all possible combinations of events and narratives. The pretrained BERT multilingual cased model consists of 12 stacked Transformer blocks with 12 attention heads each. We attached a linear layer with a softmax activation function at the output of these layers to serve as the classification layer. As input to the classifier, we took the representation of the special Classify token from the last layer of the language model. Both the narrative and event models were jointly trained end-to-end on the downstream task of three-class identification. We used the Adam optimizer with the learning rate of $5e-5$ and weight decay set to 0.01 for regularization. The models were trained for 4 epochs with batch size 64 through the HuggingFace Transformers library (41). The hyperparameters chosen were among those recommended in (36). In addition, the optimal learning rate was identified by plotting the loss against different learning rates over a few epochs. Statistics of the performances of the models are reported in the Supplementary Material.

A latent space model for news outlets' stance

The latent stance (in adverse, neutral, and positive events) was independently estimated by means of a latent space model (30, 31). We modeled the number of news articles y_{ijk} published by each news outlet $i \in \{1, \ldots, N\}$ within one of three categories ($M = 3$): antivax ($j = 1$), neutral ($j = 2$), provax ($j = 3$), and for the subset of type-k events (with k in {adverse, neutral, positive}) via a Poisson distribution $y_{ijk} \sim \mathcal{P}ois(\lambda_{ijk})$ for which the log-intensity parameter is defined as $\log\lambda_{ijk} = \alpha_{ik} - \|x_{ik} - z_{jk}\|$ where $\|\cdot\|$ denotes the euclidean distance between the stance of news outlet i, x_{ik}, and the ideal stance z_{jk}. The ideal stances are assumed such that $\mu_{z_{jk}} \in \{-1, 0, 1\}$ for all j and k, where $\mu_{z_{1k}} = -1$ is the expected stance associated to antivax, $\mu_{z_{2k}} = 0$ is the expected stance associated to neutral and $\mu_{z_{3k}} = 1$ is the expected stance associated to provax. As an example, consider a news outlet publishing 3 antivax articles, 6 neutral articles, and 9 provax articles related to positive events. The propensity of the news outlet toward positive events is proportional to the total number of articles published by the news outlet on these events (18 in the example). The distribution of articles published across narratives provides information on the narrative of the news outlets, (mostly provax in the example).

We estimate the parameters of our model within a Bayesian setup by means of a Markov Chain Monte Carlo (MCMC) algorithm. On the one hand, the Bayesian estimation procedure allows for dealing with a complex model following a straightforward workflow. On the other hand, the Bayesian approach allows us to fully take into account the uncertainty associated with our estimates. The prior specification for our set of parameters is the following: we assume a vague normal prior for the intercept parameter α_{ik}, i.e. $\alpha_{ik} \sim \mathcal{N}(0, 15)$ for each i and k and an informative normal prior for the latent factor x_{ik}, i.e. $x_{ik} \sim \mathcal{N}(0, 1)$ for each i and k and for z_{jk}, i.e. $z_{jk} \sim \mathcal{N}(\mu_{jk}, \sigma^2)$. While we opted for a vague prior for the news-outlet-specific intercept parameter, we decided to set a more informative prior on the latent coordinates. This choice is a soft constraint that helps the identification of the latent coordinates, as suggested in (30). The MCMC technique adopted in this case is a Metropolis-within-Gibbs (42) and is used to approximate the joint posterior of our model. Let $\boldsymbol{\alpha}_k = (\alpha_{1k}, \ldots, \alpha_{Nk})$, $\mathbf{x}_k = (x_{1k}, \ldots, x_{Nk})$ and $\mathbf{z}_k = (z_{1k}, \ldots, z_{Mk})$, the joint posterior can be written as:

$$\pi(\boldsymbol{\theta}_k \mid Y_k) \propto \left(\prod_{i=1}^{N} \prod_{j=1}^{M} f(y_{ijk} \mid \boldsymbol{\theta}_k) \right) \pi(\boldsymbol{\theta}_k),$$

where $\boldsymbol{\theta}_k = \{\boldsymbol{\alpha}_k, \mathbf{x}_k, \mathbf{z}_k\}$ and Y_k is the collection of $\{y_{11k}, \ldots, y_{NMk}\}$.

The algorithm implements the following steps:

1. Set $\boldsymbol{\alpha}_k^{(0)}$, $\mathbf{x}_k^{(0)}$, $\mathbf{z}_k^{(0)}$,
2. for each $h \in \{1, \ldots, H\}$
   (a) Draw $\alpha_{ik}^{(h)}$ from $\pi(\alpha_{ik} \mid \mathbf{x}_k, \mathbf{z}_k \boldsymbol{\alpha}_{-ik})$ via Random-Walk Metropolis-Hastings for each i;
   (b) Draw $x_{ik}^{(h)}$ from $\pi(x_{ik} \mid \mathbf{x}_{-ik}, \mathbf{z}_k, \boldsymbol{\alpha}_k)$ via Random-Walk Metropolis-Hastings for each i;
   (c) Draw $z_{jk}^{(h)}$ from $\pi(z_{jk} \mid \mathbf{x}_k, \mathbf{z}_{-jk}, \boldsymbol{\alpha}_k)$ via Random-Walk Metropolis-Hastings for each j.

We notice that $H = 5{,}000$ is enough to obtain convergence after having discarded the first $1{,}000$ iterations as a burn-in. Since the adopted Bayesian framework allows for the uncertainty quantification of the parameter estimates, we report in Figure S2 the 90% credible ellipses for the Propensity factor and Narrative bias parameters. Although some news outlets exhibit a quite wide range of variation, the difference between questionable and reliable news outlets in the narrative bias space, as well as the presence of a different level of selection bias, remains clear.

Selection index

To quantify how unbalanced a news outlet is in reporting on positive and negative events, we define the *SelectionIndex*, which is computed by measuring the distance between the location of any news outlet i on the Propensity Factor plane and the θ-degree line passing through the origin of the plane:

$$SelectionIndex_i = \left| \sin(\theta)\left(PF_i^{(Adv)}\right) - \cos(\theta)\left(PF_i^{(Pos)}\right) \right|,$$

where the propensity factor of news outlet i in reporting an event of type k, is the intercept parameter $\alpha_{i,k}$ of the latent space model estimated on the set of type-k events, i.e. $PF_i^{(k)} = \alpha_{i,k}$. We further assume that $\theta = \frac{\pi}{4}$, i.e. we consider a news outlet to be perfectly balanced if it shows an equal propensity to report on positive and adverse newsworthy events.

News outlets' network

We relied on the Twitter data described above and built an undirected weighted graph G to quantify the similarity of news outlets in terms of audience. We started creating a matrix R with retweeters as rows and news outlets as a column. The entry $r_{i,j}$ of R is the

number of times user i retweeted the news outlet j. We then compute the cosine similarity for each pair of columns to obtain the similarity measure for each pair of news outlets. Thus, the weight $w_{h,k}$ of the edge between node h and k in the graph G is equal to:

$$w_{h,k} = \frac{r_h \cdot r_k}{\|r_h\| \, \|r_k\|},$$

where r_h and r_k are the two column vectors of news outlets i and j, respectively. Notice that $w_h, k \in [0, 1]$, since all the entries of the matrix are nonnegative. Finally, we excluded all the 0-degree nodes and then all the edges with a weight below the mean weight of all edges, obtaining a graph with 206 nodes and 1,555 edges. The weight distribution of the complete network and the weight threshold is shown in Figure S5, while the effect of the cut on the degree distribution is depicted in Figure S6.

Cosine similarity represents just one among several similarity metrics commonly employed to quantify the overlap of news outlets' audiences. Thus, to validate the robustness of our findings, we replicated the analysis using Jaccard similarity, another popular measure for assessing the overlap between two sets. The outcomes, detailed in Figure S3, qualitatively correspond with the findings presented in the main paper.

Acknowledgments

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

Author Contributions

All authors conceptualized the research. A.G. and E.B. collected the data. A.G., A.P., and E.B. conducted the analysis and contributed to the visualization. All authors contributed to the interpretation of the results and to the writing of the manuscript. M.D. and F.Z. supervised the work.

Preprints

A preprint of this article is published at https://arxiv.org/abs/2301.05961.

Data Availability

The data and codes underlying this article are available at the following OSF repository: https://osf.io/nsmk3/, https://osf.io/gyxr2/, https://osf.io/zj5tg/. Data are made available under the platforms' terms of service.

References

1 Habermas J, Lennox S, Lennox F. 1974. The public sphere: an encyclopedia article (1964). *New Ger Crit*. 0(3):49–55.

2 Schäfer MS. 2016. *Digital public sphere*. John Wiley & Sons, Ltd. p. 1–7.

3 Cinelli M, *et al.* 2020. The COVID-19 social media infodemic. *Sci Rep*. 10(1):1–10.

4 Del Vicario M, *et al.* 2016. The spreading of misinformation online. *Proc Natl Acad Sci U S A*. 113(3):554–559.

5 Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D. 2019. Fake news on twitter during the 2016 US presidential election. *Science*. 363(6425):374–378.

6 Shao C, *et al.* 2018. The spread of low-credibility content by social bots. *Nat Commun*. 9(1):1–9.

7 Vosoughi S, Roy D, Aral S. 2018. The spread of true and false news online. *Science*. 359(6380):1146–1151.

8 Wardle C, Derakhshan H. 2017. *Information disorder: toward an interdisciplinary framework for research and policymaking*. Vol. 27. Strasbourg: Council of Europe. p. 1–107.

9 Zarocostas J. 2020. How to fight an infodemic. *Lancet*. 395(10225): 676.

10 Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. 2021. The echo chamber effect on social media. *Proc Natl Acad Sci U S A*. 118(9):e2023301118.

11 Hosseinmardi H, *et al.* 2024. Causally estimating the effect of Youtube's recommender system using counterfactual bots. *Proc Natl Acad Sci U S A*. 121(8):e2313377121.

12 Bakshy E, Messing S, Adamic LA. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science*. 348(6239): 1130–1132.

13 Bourgeois D, Rappaz J, Aberer K. 2018. Selection bias in news coverage: learning it, fighting it. In: Companion Proceedings of the The Web Conference 2018. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE. p. 535–543.

14 Sharot T, Sunstein CR. 2020. How people decide what they want to know. *Nat Hum Behav*. 4(1):14–19.

15 Groeling T. 2013. Media bias by the numbers: challenges and opportunities in the empirical study of partisan news. *Annu Rev Polit Sci (Palo Alto)*. 16(1):129–151.

16 Eberl J-M, Boomgaarden HG, Wagner M. 2017. One bias fits all? three types of media bias and their effects on party preferences. *Commun Res*. 44(8):1125–1148.

17 Puglisi R, Snyder Jr JM. 2015. Empirical studies of media bias. In: Anderson SP, Waldfogel J, Strömberg D, editors. Handbook of media economics, vol. 1. Elsevier. p. 647–667.

18 Shoemaker PJ, Vos T. 2009. *Gatekeeping theory*. Routledge.

19 Della Vigna S, Kaplan E. 2007. The fox news effect: media bias and voting. *Q J Econ*. 122(3):1187–1234.

20 Flaxman S, Goel S, Rao JM. 2016. Filter bubbles, echo chambers, and online news consumption. *Public Opin Q*. 80(S1):298–320.

21 Groseclose T, Milyo J. 2005. A measure of media bias. *Q J Econ*. 120(4):1191–1237.

22 Bhadani S, *et al.* 2022. Political audience diversity and news reliability in algorithmic ranking. *Nat Hum Behav*. 6(4): 495–505.

23 Bovet A, Makse HA. 2019. Influence of fake news in Twitter during the 2016 US presidential election. *Nat Commun*. 10(1):1–14.

24 Media bias fact check. [accessed 2023 Dec 13]. https://mediabiasfactcheck.com/.

25 Allsides — balanced news and media ratings. [accessed 2023 Dec 13]. https://www.allsides.com/.

26 The journalism trust initiative. [accessed 2023 Dec 13]. https://www.journalismtrustinitiative.org/.

27 D'Alessio D, Allen M. 2000. Media bias in presidential elections: a meta-analysis. *J Commun.* 50(4):133–156.

28 Legge 31 luglio 2017, n. 119. [accessed 2022 Dec 19]. https://www.trovanorme.salute.gov.it/norme/dettaglioAtto?id=60201&articolo=2.

29 NewsguardTech.com. Social impact report 2021, 2022. [accessed 2022 Aug 27]. https://www.newsguardtech.com/wp-content/uploads/2022/01/NewsGuard-Social-Impact-Report-1.21.22.pdf.

30 Barberá P. 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Polit Anal.* 23(1):76–91.

31 Hoff PD, Raftery AE, Handcock MS. 2002. Latent space approaches to social network analysis. *J Am Stat Assoc.* 97(460):1090–1098.

32 Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. 2008. Fast unfolding of communities in large networks. *J Stat Mech.* 2008(10):P10008.

33 Robertson CE, *et al.* 2023. Negativity drives online news consumption. *Nat Hum Behav.* 7(5):812–822.

34 Lin H, *et al.* 2023. High level of correspondence across different news domain quality rating sets. PNAS Nexus. 2(9):pgad286.

35 Brugnoli E, Delmastro M. 2024. Dynamics and triggers of misinformation on vaccines, arXiv, arXiv:2207.12264, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.2207.12264

36 Devlin J, Chang M-W, Lee K, Toutanova K. 2018. BERT: pre-training of deep bidirectional transformers for language understanding, arXiv, arXiv:1810.04805, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.1810.04805

37 Abas AR, El-Henawy I, Mohamed H, Abdellatif A. 2020. Deep learning model for fine-grained aspect-based opinion mining. *IEEE Access.* 8:128845–128855.

38 Kokab ST, Asghar S, Naz S. 2022. Transformer-based deep learning models for the sentiment analysis of social media data. *Array.* 14:100157.

39 Vaswani A, *et al.* 2017. Attention is all you need. In: Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. Advances in neural information processing systems. vol. 30, Curran Associates, Inc. p. 6000–6010.

40 Edward Ma: Nlp augmentation. 2019. https://github.com/makcedward/nlpaug.

41 Wolf T, *et al.* 2019. Huggingface's transformers: state-of-the-art natural language processing, arXiv, arXiv:1910.03771, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.1910.03771

42 Robert CP, Casella G, Casella G. 1999. *Monte Carlo statistical methods.* vol. 2. Springer.