RESEARCH ARTICLE

SOCIAL MEDIA

Asymmetric ideological segregation in exposure to political news on Facebook

Sandra González-Bailón[1]*, David Lazer[2], Pablo Barberá[3], Meiqing Zhang[3], Hunt Allcott[4], Taylor Brown[3], Adriana Crespo-Tenorio[3], Deen Freelon[1], Matthew Gentzkow[5], Andrew M. Guess[6], Shanto Iyengar[7], Young Mie Kim[8], Neil Malhotra[9], Devra Moehler[3], Brendan Nyhan[10], Jennifer Pan[11], Carlos Velasco Rivera[3], Jaime Settle[12], Emily Thorson[13], Rebekah Tromble[14], Arjun Wilkins[3], Magdalena Wojcieszak[15,16], Chad Kiewiet de Jonge[3], Annie Franco[3], Winter Mason[3], Natalie Jomini Stroud[17,18], Joshua A. Tucker[19,20]

Does Facebook enable ideological segregation in political news consumption? We analyzed exposure to news during the US 2020 election using aggregated data for 208 million US Facebook users. We compared the inventory of all political news that users could have seen in their feeds with the information that they saw (after algorithmic curation) and the information with which they engaged. We show that (i) ideological segregation is high and increases as we shift from potential exposure to actual exposure to engagement; (ii) there is an asymmetry between conservative and liberal audiences, with a substantial corner of the news ecosystem consumed exclusively by conservatives; and (iii) most misinformation, as identified by Meta's Third-Party Fact-Checking Program, exists within this homogeneously conservative corner, which has no equivalent on the liberal side. Sources favored by conservative audiences were more prevalent on Facebook's news ecosystem than those favored by liberals.

Social media platforms have a large and growing role in shaping access to information, including information about politics (*1*). However, few studies systematically map this part of the information ecosystem, with minimal examination of within-platform attention and information consumption (*2*). There is virtually no research comparing what people could potentially see within platforms and what they actually see [with the exception of (*3*)]. In this work, we address these questions with data on users' access and engagement to political news on Facebook during the US 2020 election.

Herein, we examine the supply of and demand for political news by tracking (i) the inventory of all news content that users could have seen, (ii) the subset of content users actually saw on their feeds, and (iii) the smaller subset of content with which users engaged (via clicks, reactions, likes, reshares, or comments). The analysis of this "funnel of engagement"—from potential exposure to actual exposure to engagement—gives a comprehensive picture of the information environment on Facebook during the 2020 election. Our analyses show that both algorithmic and social amplification play a part in increasing ideological segregation. Algorithmic amplification refers to data-driven automated processes that result in some content being more visible in users' feeds; social amplification refers to choices made by users that also grant more visibility to specific content through sharing and reposting. We show that these processes operate asymmetrically across the US political "right" (conservatives or Republican Party) and the political "left" (liberals or Democratic Party), with the presence of much more homogeneous news consumption on the right—a pattern that has no parallel on the left.

There has been a vigorous debate about the role of the internet in shaping the information that people see. Generally, individuals align their attention with their interests, but this tendency does potentially pose a challenge for a democracy if different people see fundamentally incompatible political information. Early research (*4*, *5*) argued that personalization algorithms (filter bubbles) and social curation processes (echo chambers) increase the probability that people will surround themselves with ideologically compatible information. Subsequent research has found that online news diets are still fairly diverse ideologically and not more segregated than offline news consumption (*6*–*13*). Much of this research, however, relies on web-browsing data, and there is still very little research on the news that people see within platforms such as Facebook and Twitter. The literature also suggests asymmetric news consumption: There is a substantial right-leaning bubble that attracts a much more homogeneously conservative audience that does not have an equivalent on the left (*7*, *14*), and there is a higher prevalence of unreliable content on the right relative to the left (*15*, *16*).

Key limitations in past work include a reliance on domain-level aggregations of exposure, thus missing curation effects at the news-story level. For example, both liberals and conservatives may read content from the *Wall Street Journal* domain, but if liberals only read news stories and conservatives only read opinion content, it would be misleading to consider the *Wall Street Journal* a meeting place for liberals and conservatives. Domain-level aggregations of exposure may thus understate the level of segregation in news consumption.

Another important limitation of past work is a lack of data about individual exposure to content within platforms. One highly cited exception is a paper that examined exposure to news content on Facebook throughout the funnel of engagement (*3*). This paper found that there are social and algorithmic effects directing ideologically compatible content to users at each stage of the funnel, but the algorithmic effects are modest in size compared with social curation and individual choice. This study is now quite dated (the data are from 2014), it has problematic generalizability because of the peculiar subsample of Facebook users it examined (i.e., the small fraction of people who volunteered their partisan identity in their profile), and it only analyzed content posted by friends, whereas today Pages and Groups are also important providers of content within the platform (these are public and private profiles created by organizations and collectives). In section 4.2 of the supplementary materials (SM), we offer more details about these limitations and how we overcame them.

It is plausible that ideological concordance is a much stronger driver in following Pages and/or joining Groups than in selecting friends. That is, although it is likely that people choose politically oriented Pages and/or Groups based in part on ideological concordance, existing research suggests that ideological agreement plays

[1]Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA, USA. [2]Network Science Institute, Northeastern University, Boston, MA, USA. [3]Meta, Menlo Park, CA, USA. [4]Stanford Doerr School of Sustainability, Stanford University, Stanford, CA, USA. [5]Department of Economics, Stanford University, Stanford, CA, USA. [6]Department of Politics and School of Public and International Affairs, Princeton University, Princeton, NJ, USA. [7]Department of Political Science, Stanford University, Stanford, CA, USA. [8]School of Journalism and Mass Communication, University of Wisconsin–Madison, Madison, WI, USA. [9]Graduate School of Business, Stanford University, Stanford, CA, USA. [10]Department of Government, Dartmouth College, Hanover, NH, USA. [11]Department of Communication, Stanford University, Stanford, CA, USA. [12]Department of Government, William & Mary, Williamsburg, VA, USA. [13]Department of Political Science, Syracuse University, Syracuse, NY, USA. [14]School of Media and Public Affairs and Institute for Data, Democracy, and Politics, The George Washington University, Washington, DC, USA. [15]Department of Communication, University of California, Davis, Davis, CA, USA. [16]Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands. [17]Moody College of Communication, University of Texas at Austin, Austin, TX, USA. [18]Center for Media Engagement, University of Texas at Austin, Austin, TX, USA. [19]Wilf Family Department of Politics, New York University, New York, NY, USA. [20]Center for Social Media and Politics, New York University, New York, NY, USA.
***Corresponding author. Email: sandra.gonzalez.bailon@asc.upenn.edu**

only a small role in how people choose friends (and presumably many Facebook friends are friends, in a sociological sense) (*17*, *18*). Finally, little of this past research compares findings regarding news consumption across internet-based modalities (e.g., exposure through browsing the web versus seeing content on Facebook's Feed).

To address these limitations, this paper answers the following six research questions (RQs): RQ1–How ideologically segregated is news consumption on Facebook, and are those patterns of segregation symmetric on the right and left; RQ2–How does segregation vary with potential news consumption versus actual exposure versus engagement; RQ3–How does segregation vary if the level of analysis is URLs rather than domains (thus capturing curation of content within domains); RQ4–How segregated is exposure on Facebook relative to the benchmark of browsing behavior (the predominant source of data in past research); RQ5–How segregated are the streams of content from the major pathways to exposure on Facebook (friends, Pages, and Groups); and RQ6–How prevalent is exposure to unreliable content on the right relative to the left.

Study overview

This research is part of the US 2020 Facebook and Instagram Election Study, a collaborative effort between Meta and a team of external researchers that was initiated in early 2020 to design and produce transparent and reproducible research on the political impact of Facebook and Instagram (SM section S1). The data in this paper draw from the set of $N \sim 208$ million US-based adult active users whose political ideology can be measured and track all URLs classified as political news that were posted on the platform from 1 September 2020 to 1 February 2021. In other words, our results are based on the activity generated by the nearly full set of US-based users that are active on the platform, which was 231 million users during our study period (SM section S2).

To identify political news URLs, we used Facebook's internal civic and news classifiers (SM section S3.4). We used URLs and domains as our units of observation to calculate segregation metrics at the story and source levels. We also analyzed the overall news ecosystem through the analysis of clustering in coexposure networks (where nodes are news stories or domains and edges encode the number of unique users exposed to a given pair; Fig. 1A). Our analyses focus on political news posted by friends, Pages, and Groups. These posts create the supply of content we reference using the term "inventory" (Fig. 1B). This inventory is determined by the underlying network that users build on the platform (i.e., choices to follow Pages, join Groups, and "friend" individuals). This network is itself the result of another curation process that is determined by users' preexisting networks and interests, algorithmic recommendations on whom to follow, and platform affordances (i.e., the existence of Pages and Groups; SM section S3.2).

Facebook's algorithm ranks content from the inventory and presents a selection in list form in users' Feeds (SM section S3.3). Users can then choose to engage with that content by clicking, reacting, liking, commenting, or resharing. Engagement on the platform generates data that are then fed back into the algorithmic curation of content for users' feeds. This is one of the key characteristics of social media: Algorithmic and social curation processes are in constant feedback. In our analyses, we only examined posts classified as political news that contain a URL (SM section S3.1), which amount to about 3% of all posts shared by US adult users and 3.9% of all content that US adult users saw on the platform during our study period. In other words, the algorithmic curation process that we analyzed here relies on many more signals than those generated by the content we analyzed (i.e., political domains and URLs).

For each URL (and corresponding domain), we have measures of the potential, exposed, and engaged audience. The potential audience of a URL is the set of unique users that could have been exposed to that content, the exposed audience is the set of unique users that saw a post containing that URL on their Feed and the engaged audience is the set of unique users that clicked, reacted, liked, reshared, or commented on the post with the URL [note that

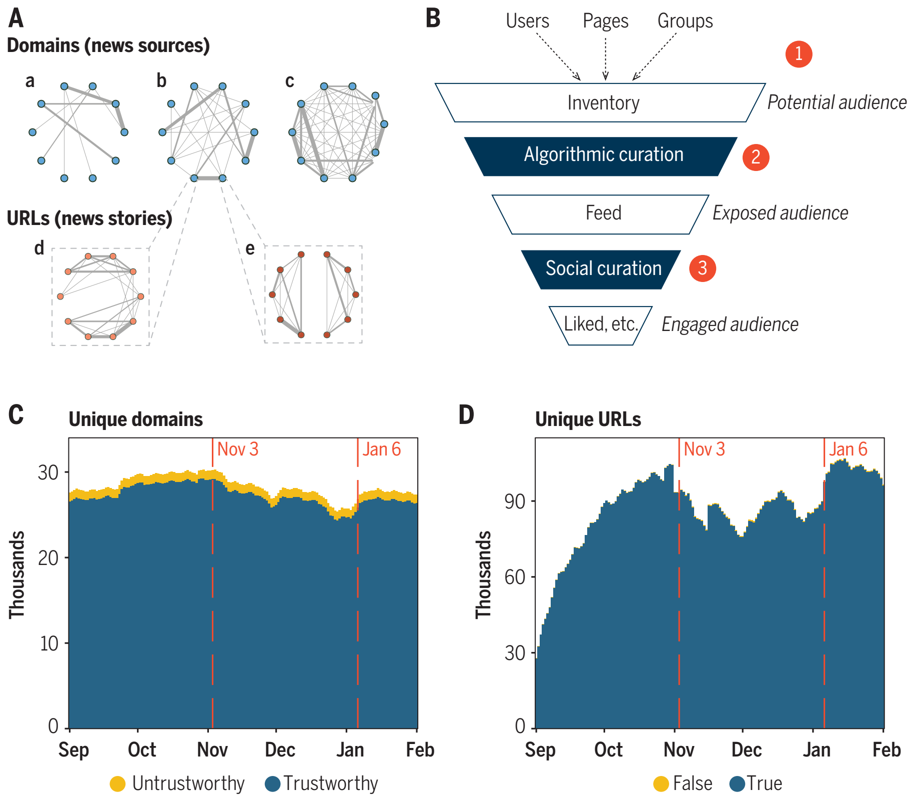

Fig. 1. Summary of data and levels of analysis. (**A**) Our analyses center on three different levels: domains (to calculate segregation metrics at the news source level), URLs (to calculate segregation metrics at the news-story level), and coexposure networks, where nodes are domains or news stories and the edges encode the number of unique users exposed to a given pair. The analysis of coexposure networks allows us to determine whether there is notable clustering in exposure to sources and, within sources, in exposure to stories. The schematic examples in subpanels (a) to (c), for instance, do not show strong evidence of clustering, but the networks in subpanels (d) and (e) do. (**B**) Schematic representation of the funnel of engagement (from potential exposure to actual exposure to engagement). Numbers indicate the level of analysis implied in our RQ2: (i) Are news inventories ideologically segregated, (ii) does segregation increase after algorithmic curation, and (iii) does segregation increase based on what content users choose to engage with. (**C** and **D**) Number of news domains (categorized as untrustworthy or trustworthy) (C) and news URLs (rated false and not rated false by Meta's 3PFC) (D) in our data, presented as daily counts. The increase in the number of unique news stories depicted in (D) between September and November is partly an artifact of the threshold that we impose on the analysis of URLs (for privacy reasons, we only analyze URLs that were shared more than 100 times by US-based users during the observation period). Yellow indicates the volume of news domains or news URLs that were rated untrustworthy or false, and blue indicates the volume of news domains or news URLs that were not rated untrustworthy or false. The red dashed lines highlight November 2 and January 6, which are the dates of the election and the US Capitol riot, respectively.

our measure is agnostic about the affect (as in subjective feeling or emotion) of the engagement]. News stories are classified as "false" if the URL was rated false by Meta's Third-Party Fact-Checking Program (3PFC) as of 15 February 2021 (SM section S3.4). This measure of "misinformation" likely undercounts the total volume of false news circulating on the platform; but whereas specific false news stories may go undetected (i.e., only a tiny amount of inventory ever makes it to the 3PFC), untrustworthy labels at the domain level, which are applied whenever a domain had two or more URLs rated false by the 3PFC, have better coverage and are more reliable (SM section S4.1). Our segmentation of audiences based on ideology relies on Facebook's internal ideology classifier, which predicts the political ideology of US adult active users (SM section S3.4). For every URL (and domain), we have the count of users who viewed the post who are predicted to be conservative and liberal.

Data

Our data are based on activity that includes the N ~ 208 million US-based users who are active on the platform and whose ideology can be predicted using Facebook's classifier. To protect privacy, we do not have access to individual-level data: We only analyzed URLs that were shared (either privately and publicly) more than 100 times, and we only analyzed aggregate exposure and engagement metrics for US adult Facebook users in relation to the URLs that were shared (SM section S2.1).

In total, our data comprise aggregated exposure and engagement metrics for N ~ 208 million US adult active users with an ideology score in relation to N ~ 35,000 unique domains and N ~ 640,000 unique URLs that were classified as political news and were shared more than 100 times during our observation period. The top five most-viewed domains are, in decreasing order, cnn.com, dailywire.com, foxnews.com, nytimes.com, and nbcnews.com. However, the most-viewed news stories (URLs) are not necessarily published by the most-viewed domains. For instance, pjmedia.com published the most-viewed story during the observation period (the story, under the headline "Military ballots found in the trash in Pennsylvania—Most were Trump votes," was viewed 113,272,405

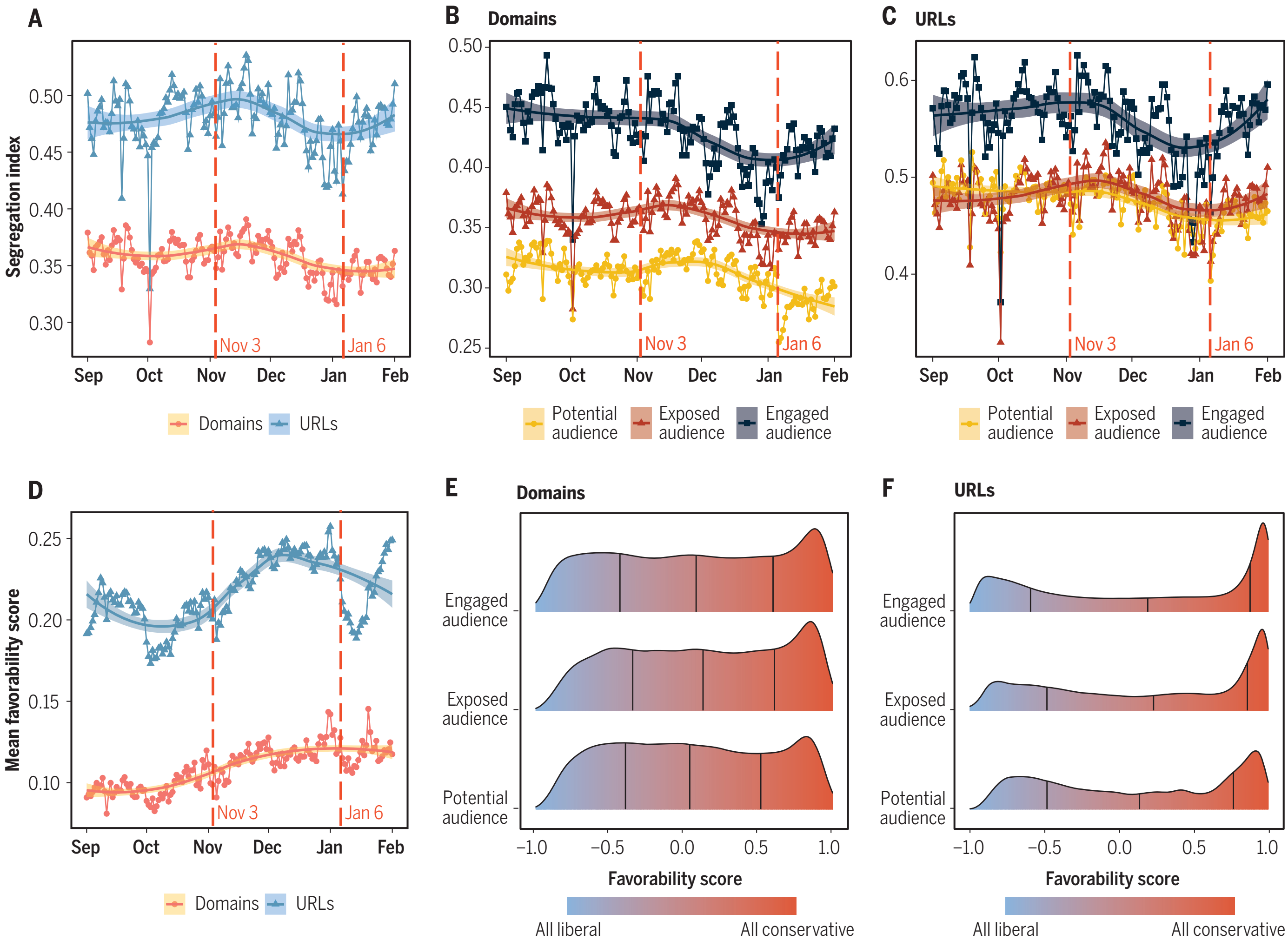

Fig. 2. Segregation and audience polarization at the domain and URL levels. (**A**) The segregation score based on exposed audience and calculated according to Eq. 1 is consistently higher at the URL level, suggesting that there are information curation practices with news stories that get masked when aggregating the data at the domain level. (**B** and **C**) Segregation scores drawn from exposed audiences are higher than those based on potential audiences but lower than the scores from engaged audiences (the difference between potential and engaged audiences is only visible at the domain level). This suggests that algorithmic and social amplification both contribute to segregation (note that we use different *y*-axis ranges to facilitate comparison across curves for domains and URLs). (**D**) The mean favorability scores (calculated according to Eq. 2, with –1 indicating a homogeneously liberal audience and 1 a homogeneously conservative audience) suggest that, overall, audiences that are consuming political news have a right-leaning slant (i.e., all scores are above the zero line). (**E** and **F**) The density plots show that the distribution is substantially skewed toward the right, with more domains and URLs being favored by very conservative audiences. Vertical black lines mark the quantiles of the distributions; scores are calculated for each domain and URL according to Eq. 2. For (A) to (D), shaded regions indicate 95% confidence intervals for the time trend based on a local polynomial regression.

times), but the domain is ranked in the 51st position (SM section S4.1). Figure 1C shows the time series of daily counts for domains, and Fig. 1D shows the daily counts for URLs. Only a small fraction of all domains is classified as "untrustworthy"; the fraction of false news stories is barely perceptible because it is very small compared with the full volume of news stories.

Measures

Our analyses rely on two measures: the segregation index (which offers a summary statistic of the entire information environment) and the favorability scores (which are associated with individual domains and URLs and allow us to infer the ideological composition of their audiences). In particular, we define the segregation index on a given period t as

$$S = \sum_{n\in N}\left(\frac{C_n}{C}\times\frac{C_n}{v_n}\right) - \sum_{n\in N}\left(\frac{L_n}{L}\times\frac{C_n}{v_n}\right) \quad (1)$$

where $n \in N$ indexes the domains and URLs in our data (with daily temporal resolution), C_n is the count of conservatives exposed to a particular domain or URL n, L_n is the count of liberals, v_n is the total number of unique views for domain or URL n, and C and L are the total number of conservatives and liberals, respectively. For the analyses presented here, we dichotomized the ideology scores such that users with a score ≤0.35 are categorized as liberal, and those with a score ≥0.65 are categorized as conservative (see SM section S4.10 for alternative operationalizations and robustness tests).

This index follows (*6*) and is adapted from research on residential segregation. It captures the extent to which conservatives and liberals visit the same neighborhoods in the information ecosystem. The first term is the (visit-weighted) average exposure of conservatives to conservatives, and the second term is the average exposure of liberals to conservatives. As a point of comparison, Gentzkow and Shapiro (*5*) examined the browsing behavior of a representative panel of individuals and found a gap of exposure to conservatives between conservatives and liberals of 7.5 percentage points (60.6% for conservatives and 53.1% for liberals).

The favorability scores are defined as

$$fav_n = \frac{(C_n - L_n)}{(C_n + L_n)} \quad (2)$$

These scores, used also in prior work (*9*, *19*), are assigned daily as well as for the entire period to each domain and URL. The score equals 1 when a given domain or URL n has an audience formed exclusively by conservatives and −1 when it has an audience formed exclusively by liberals (0 means that conservatives and liberals are equally likely to be in the audience of a domain or URL n). We define audience polarization as the extent to which the distribution of favorability scores is bimodal and far away from zero.

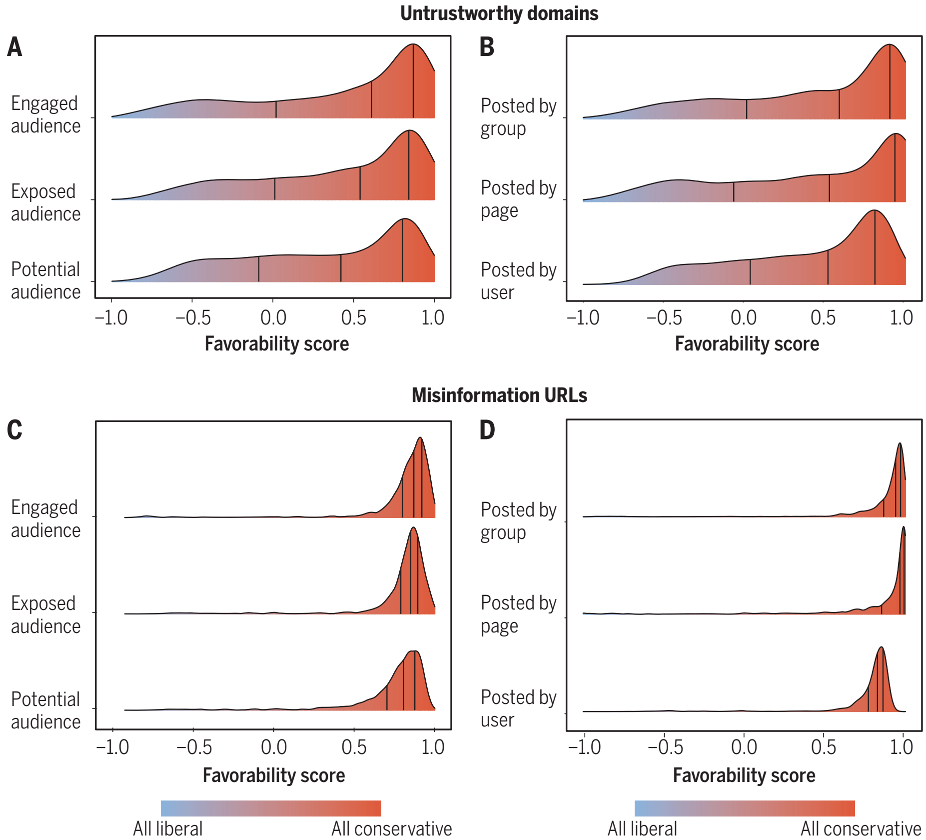

Fig. 3. Audiences of untrustworthy domains and false URLs. (**A**) The density plots show that most domains that are categorized "untrustworthy" are favored by audiences that are predominantly conservative. For potential, exposed, and engaged audiences, 71, 76, and 76% of untrustworthy domains have audiences that are conservative on average (favorability score > 0). (**B**) The audience of domains posted by Pages is the most conservative. (**C**) The data suggest a stark conservative-leaning slant for news stories labeled as "misinformation." For potential, exposed, and engaged audiences, 97% of false URLs in each case have audiences that are conservative on average. (**D**) The sharing of news stories labeled as "false" is much more concentrated among conservative Pages than among Groups and users. Vertical black lines mark the quantiles of the distributions; scores are calculated for each domain and URL according to Eq. 2.

We calculated the segregation index and the favorability scores for potential, exposed, and engaged audiences. We complemented these analyses with network-based measures of clustering in exposed audiences by examining the backbone of coexposure networks (*20*, *21*). This network approach allowed us to detect segregation in news exposure using only behavioral traces that do not take into account the ideology of audiences (SM section S6). As we show below, the clustering that emerges from the behavior of users is aligned with the ideology partition.

Results

Figure 2 provides answers to RQs 1 to 4. The segregation score based on exposed audience for domains fluctuates around 0.35 (i.e., the gap between the intersection of conservatives with conservatives versus liberals with conservatives is 35 percentage points). This is substantially higher than values found in prior research based on web-browsing behavior, which range from 0.02 to 0.1 (*6*, *9*, *11*). The score for URLs is substantially higher still, fluctuating roughly between 0.45 and 0.5, in line with a mechanism of ideological curation of content within domains. As we move down the funnel of engagement, ideological segregation increases. Segregation scores for potential audiences are far higher than what prior research has estimated for news consumption (online and offline), scores for exposed audiences are higher than those based on potential audiences, and, for domain-aggregated data, scores for engaged audiences are higher than those for exposed audiences (Fig. 2, B and C). This suggests that algorithmic and social amplification are both contributing to increased segregation: As we move down the funnel of engagement (i.e., as the footprint of algorithmic and social curation becomes more evident), liberal and conservative audiences become more isolated from each other.

The mean favorability scores indicate that audiences that are consuming political news on Facebook have, overall, a right-leaning slant (all scores are above the zero line; Fig. 2D). However, the mean score is a somewhat misleading summary statistic, given the full distribution (Fig. 2, E and F). The density plots tell us that the favorability scores are substantially skewed toward the right: There are more domains and URLs being favored by very conservative audiences. To address RQ4 in more depth, we ran supplementary analyses in which we compared segregation and favorability scores for the subset of Facebook users that consented to provide data of their web activity (SM section S2.2). With this group of users, we ran intraperson comparisons of exposure to news content on- and off-platform. These comparisons reveal that the segregation score on Facebook is three times as high as a web-tracking benchmark for the same set of individuals (SM section S5).

Figure 3 provides answers to RQs 5 and 6. Most sources of misinformation are favored by conservative audiences. The distribution of favorability scores does not shift substantially as we move down the funnel of engagement (Fig. 3, A and C), suggesting that algorithmic and social amplification do not exacerbate the already existing audience segregation for misinformation content. However, misinformation shared by Pages and Groups has audiences that are more homogeneous and completely concentrated on the right (Fig. 3, B and D).

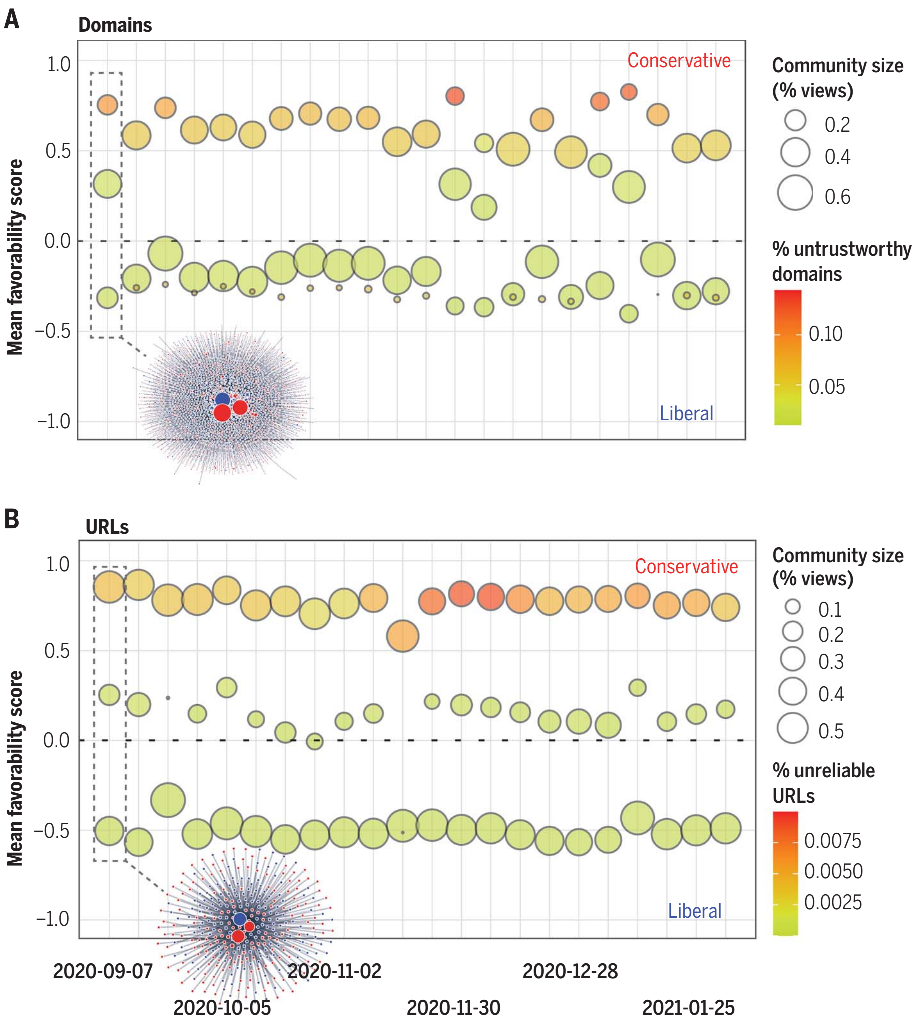

Fig. 4. Top communities in coexposure networks. (**A** and **B**) Mean favorability scores for the top three communities by views in the weekly coexposure networks for domains (A) and URLs (B). The graphic in the lower-left part of each panel shows the full network for week 1 collapsed to the communities (i.e., each node is a group of domains and URLs with higher audience overlap). The presence of clusters in these networks suggests selective exposure: The networks are clearly organized around large clusters formed by mostly liberal or mostly conservative audiences, regardless of the level of aggregation (domain or URL). These clusters also provide additional evidence of asymmetric ideological segregation, which is especially visible in the URL networks: Across the observation window, there is a sizable cluster of news stories consumed predominantly by very conservative audiences who are exposed to a higher percentage of unreliable news stories (color coded).

The SM contains additional analyses disaggregating the data for Pages, Groups, and users; COVID-related content; high-political-interest and low-political-interest users (under different measures of political interest); and weekly aggregations. The picture that these additional analyses draw is that, overall, the patterns that we identify here are consistent under different operationalizations and data aggregations. These additional analyses reveal two further insights: (i) Pages and Groups contribute much more to segregation and audience polarization than users (SM section S4.4), and (ii) users classified as having high political interest are twice as segregated as low-interest users (SM section S4.7).

Figure 4 provides additional evidence that confirms the robustness of the ideological asymmetry in news consumption (RQ1) and the prevalence of exposure to unreliable content on the right relative to the left (RQ6). The analysis of coexposure networks (exposed-audience only) allows us to determine whether we observe evidence of segregation even if we know nothing about the ideology of users: The presence of clusters in the network is evidence of selective exposure (which, in this case, is shaped by users' social networks and news-feed algorithms).

We used a community detection approach to identify the presence of these clusters (*20*). Figure 4 plots the top three communities, sized in terms of views, for each weekly network. Each of these communities is formed by domains and URLs that have a higher overlap of audiences internally than externally (i.e., with domains and URLs in other communities). We then calculated the mean favorability score for each community (i.e., the average score for the domains and URLs classified in the same community). This revealed that the clustering that was identified based only on coexposure behavior maps onto the two sides of the ideological continuum: There is clearly a cluster of domains and URLs that liberal users are coexposed to, and a cluster on the conservative side. However, we see again a major asymmetry between the left and right. News sources and stories consumed by conservative audiences depart more clearly from the zero line of cross-cutting exposure, which means that their audiences are more homogeneously conservative and, therefore, more isolated. These outlets on the right also post a higher fraction of news stories (URLs) rated false by Meta's 3PFC program (color coded in the figure), which means that conservative audiences are more exposed to unreliable news. (See SM section S6 for more details on network construction, statistics, and figures.)

Discussion

Our analyses highlight that Facebook, as a social and informational setting, is substantially segregated ideologically—far more than previous research on internet news consumption based on browsing behavior has found. But our analyses also show that individual preferences and platform affordances intertwine in a complex fashion. Ideological segregation manifests far more in content posted by Pages and Groups than in content posted by friends. That Pages and Groups are associated with higher levels of ideological segregation suggests that the choice of which Pages to follow and which Groups to join is driven far more by ideological congruence than the choice of with whom to be friends, as discussed earlier. An important direction for further research is to understand how individuals discover and decide to follow Pages and join Groups.

The research we report here has some limitations. First, following the convention of most of the literature, we focus on URLs and domains because this allows us to examine how Facebook enables access to content that is produced off-platform. However, this approach excludes platform content that may generate different segregation patterns. Second, our definition of misinformation at the URL level relies on third-party fact checkers and their limited coverage of URLs. And third, we rely on Facebook's classifiers in part because they perform better than ad hoc alternatives, but testing the robustness of our results with other classifiers is an area that requires more research. Future research would ideally also incorporate non-URL content and expand the coverage of our misinformation metric by relying on human annotation of random samples or automated classification methods.

More generally, future research should analyze how the underlying social graph, and algorithmic recommendations on whom or what to follow, determine the inventory of content available on the platform. The algorithmic promotion of compatible content from this inventory is positively associated with an increase in the observed segregation as we move from potential to exposed audiences. Future research should consider whether segregation of potential audiences increases over time as a long-term effect of this algorithmic promotion. We hypothesize that many of the patterns that characterize Facebook's funnel of engagement also characterize other social media.

Finally, it is essential to replicate these analyses for future elections and in other national contexts to evaluate to what extent these results are contingent on the particular moment and the national and/or cultural setting. This research and, more generally, the US 2020 Facebook and Instagram Election Study have created a set of innovative processes and conventions to improve research transparency and integrity (SM section S1.2) and research ethics (SM section S1.3), which can serve as a blueprint for future collaborations between academic and industry researchers. We offer a proof of concept of the feasibility of productive collaboration between industry and external researchers, which demonstrates the credible empirical insights that can be gained through such collaboration. Present regulatory efforts, such as the European Union's Digital Services Act (DSA), include requirements that would enable academic researchers to access data from the largest internet platforms and thus should facilitate replication of our analyses. Our research has produced a machinery of processes and checks that can provide the foundation for such research infrastructure. Compared with the baseline created by past research (*3*, *22*), we have greatly moved the dial toward greater transparency and accountability.

Our results uncover the influence that two key affordances of Facebook—Pages and Groups—have in shaping the online information environment. Pages and Groups benefit from the easy reuse of content from established producers of political news and provide a curation mechanism by which ideologically consistent content from a wide variety of sources can be redistributed. As a result of social curation, exposure to URLs is systematically more segregated than exposure to domains. In the 20th century, local news media oligopolies slanted news coverage toward the mainstream of local audiences (*23*). In the 21st century, content from a large range of accessible sources may be providing the raw material for ideologically homogeneous feeds—similar to those produced by Pages and Groups. These patterns also have important implications for future research in this area: A focus on domains rather than URLs will likely understate, perhaps substantially, the degree of segregation in news consumption online.

Finally, our results uncover the clearly asymmetric nature of political news segregation on Facebook—the right side of the distributions for potential, actual, and engaged audiences looks robustly different from the left side. Thus, although there are homogeneously liberal and conservative domains and URLs, there are far more homogeneously conservative domains and URLs circulating on Facebook. This asymmetry is consistent with what has been found in other social media platforms (*24*–*26*). We also observe on the right a far larger share of the content labeled as false by Meta's 3PFC. Overall, these patterns are part of a broader set of long-standing changes associated with the fracturing of the national news ecosystem, ranging from Fox News to talk radio, but they are also a manifestation of how Pages and Groups provide a very powerful curation and dissemination machine that is used especially effectively by sources with predominantly conservative audiences (*14*).

REFERENCES AND NOTES

1. E. Shearer, "More than eight-in-ten Americans get news from digital devices" (Technical Report, Pew Research Center, 2021).
2. D. Lazer, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 21–22 (2020).
3. E. Bakshy, S. Messing, L. A. Adamic, *Science* **348**, 1130–1132 (2015).
4. C. R. Sunstein, *#Republic: Divided Democracy in the Age of Social Media* (Princeton University Press, 2017).
5. E. Pariser, *The Filter Bubble: How the New Personalized Web is Changing What We Read and How We Think* (Penguin, 2011).
6. M. Gentzkow, J. M. Shapiro, *Q. J. Econ.* **126**, 1799–1839 (2011).
7. P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, R. Bonneau, *Psychol. Sci.* **26**, 1531–1542 (2015).
8. G. Eady, J. Nagler, A. Guess, J. Zilinsky, J. A. Tucker, *SAGE Open* **9**, 2158244019832705 (2019).
9. T. Yang, S. Majó-Vázquez, R. K. Nielsen, S. González-Bailón, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 28678–28683 (2020).
10. A. M. Guess, *Am. J. Pol. Sci.* **65**, 1007–1022 (2021).
11. S. Flaxman, S. Goel, J. M. Rao, *Public Opin. Q.* **80**, 298–320 (2016).
12. E. Peterson, S. Goel, S. Iyengar, *Political Sci. Res. Methods* **9**, 242–258 (2021).
13. J. G. Webster, T. B. Ksiazek, *J. Commun.* **62**, 39–56 (2012).
14. Y. Benkler, R. Faris, H. Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (Oxford Univ. Press, 2018).
15. A. Guess, J. Nagler, J. Tucker, *Sci. Adv.* **5**, eaau4586 (2019).
16. N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, *Science* **363**, 374–378 (2019).
17. W. Minozzi, H. Song, D. M. Lazer, M. A. Neblo, K. Ognyanova, *Am. J. Pol. Sci.* **64**, 135–151 (2020).
18. D. Lazer, B. Rubineau, C. Chetkovich, N. Katz, M. Neblo, *Polit. Commun.* **27**, 248–274 (2010).
19. M. Tyler, J. Grimmer, S. Iyengar, *J. Polit.* **84**, 716950 (2022).
20. P. Pons, M. Latapy, in *Computer and Information Sciences - ISCIS 2005*, P. Yolum, T. Güngör, F. Gürgen, C. Özturan, Eds. (Lecture Notes in Computer Science, vol. 3733, Springer, 2005s).
21. M. Ángeles Serrano, M. Boguñá, A. Vespignani, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6483 (2009).
22. D. Lazer, *Science* **348**, 1090–1091 (2015).
23. M. Gentzkow, J. M. Shapiro, *Econometrica* **78**, 35–71 (2010).
24. F. Huszár *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2025334119 (2022).
25. S. González-Bailón, V. d'Andrea, D. Freelon, M. De Domenico, *PNAS Nexus* **1**, pgac137 (2022).
26. W. Chen, D. Pacheco, K.-C. Yang, F. Menczer, *Nat. Commun.* **12**, 5580 (2021).

ACKNOWLEDGMENTS

The Facebook Open Research and Transparency (FORT) team provided substantial support in executing the overall project. We are grateful for support on various aspects of project management from C. Nayak, S. Zahedi, I. Rosenn, L. Ahmad, A. Bhalla, C. Chan, A. Gruen, B. Hillenbrand, D. Li, P. McLeod, and D. Rice; engineering from Y. Chen, S. Chen, T. Lohman, R. Pyke, and Y. Wan; data engineering from S. Chintha, J. Cronin, D. Desai, Y. Kiraly, T. Li, X. Liu, S. Pellakuru, C. Xie, and B. Xiong; data science and research from H. Connolly-Sporing; academic partnerships from R. Mersey, M. Zoorob, L. Harrison, S. Aisiks, Y. Rubinstein, and C. Qiao; privacy and legal assessment from K. Benzina, F. Fatigato, J. Hassett, S. Iyengar, P. Mohassel, A. Muzaffar, A. Raghunathan and A. Sun; and content design from C. Bernard, J. Breneman, D. Leto, and S. Raj. NORC at the University of Chicago partnered with Meta on this project to conduct the fieldwork with the survey participants and pair the survey data with web tracking data for consented participants in predetermined aggregated forms. We are particularly grateful for the partnership of NORC Principal Investigator J.M. Dennis and NORC Project Director M. Montgomery. **Funding:** The costs associated with the research (such as participant fees, recruitment and data collection) were paid by Meta. Ancillary support (for example, research assistants and course buyouts) was sourced, where applicable, by academics from the Democracy Fund, the Hopewell Fund, the Guggenheim Foundation, the John S. and James L. Knight Foundation, the Charles Koch Foundation, the Hewlett Foundation, the Alfred P. Sloan Foundation, the University of Texas at Austin, New York University, Stanford University, the Stanford Institute for Economic Policy Research, and the University of Wisconsin-Madison. **Author contributions:** S.G.-B., D.L., and P.B. supervised all analyses, analyzed data, and wrote the paper. As the academic lead authors, S.G.-B. and D.L had final control rights. P.B. and M.Z. were the lead authors at Meta. S.G.-B., D.L., and P.B. designed the study. P.B., D.F., Y.M.K., N.M., D.M., B.N., E.T., R.T., C.V.R., A.W., and M.W. contributed study materials (for example, survey questionnaires, classifiers and software). H.A., P.B., A.C.-T., A.F., D.F., M.G., S.G.-B., A.M.G., S.I., C.K.d.J., Y.M.K., D.L., N.M., W.M., D.M., B.N., J.P., C.V.R., J.S., N.J.S., E.T., R.T., J.A.T., A.W., and M.W. contributed to the design of the project. P.B., T.B., A.C.-T., A.F., W.M., C.V.R., and A.W. collected and curated all platform data. S.G.-B. and P.B. contributed the figures and tables. A.F., M.G, A.M.G., B.N., N.M., J.P., and J.S. provided feedback on the manuscript. N.J.S. and J.A.T. were joint principal investigators for the academic involvement on this project, responsible for management and coordination. C.K.d.J., A.F., and W.M. led Meta's involvement on this project and were responsible for management and coordination. **Competing interests:** None of the academic researchers nor their institutions received financial or any other compensation from Meta for their participation in the project. Some

authors are or have been employed by Meta: P.B., T.B., A.C.-T., D.D., D.M., D.R.T., C.V.R., A.W., B.X., A.F., C.K.d.J., and W.M. The following academic authors have had one of more of the following funding or personal financial relationships with Meta (paid consulting work, received direct grant funding, received an honorarium or fee, served as an outside expert, or own Meta stock): M.G., A.M.G., B.N., J.P., J.S., N.J.S., R.T., J.T., and M.W. For additional information about the above disclosures as well as a review of the steps taken to protect the integrity of the research, see SM section 1.2. **Data and materials availability:** The registration of the preanalysis plan is available at https://osf.io/2y84h. Deidentified data and analysis code from this study will be archived in the Social Media Archive (SOMAR) at ICPSR, part of the University of Michigan Institute for Social Research, and made available for university IRB-approved research on elections or to validate the findings of this study (https://socialmediaarchive.org). ICPSR will receive and vet all applications for data access.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade7138
Supplementary Text
Figs. S1 to S45
Tables S1 to S19
References (*27*–*34*)
MDAR Reproducibility Checklist

Submitted 2 September 2022; accepted 31 May 2023
10.1126/science.ade7138