

Exposure to news grows less fragmented with an increase in mobile access

Tian Yang^a, Sílvia Majó-Vázquez^b, Rasmus K. Nielsen^b, and Sandra González-Bailón^{a,1}

^aAnnenberg School for Communication, University of Pennsylvania, Philadelphia, PA 19104; and ^bReuters Institute for the Study of Journalism, University of Oxford, Oxford OX2 6PS, United Kingdom

Edited by Matthew O. Jackson, Stanford University, Stanford, CA, and approved September 25, 2020 (received for review April 1, 2020)

The abundance of media options is a central feature of today's information environment. Many accounts, often based on analysis of desktop-only news use, suggest that this increased choice leads to audience fragmentation, ideological segregation, and echo chambers with no cross-cutting exposure. Contrary to many of those claims, this paper uses observational multiplatform data capturing both desktop and mobile use to demonstrate that coexposure to diverse news is on the rise, and that ideological self-selection does not explain most of that coexposure. We show that mainstream media outlets offer the common ground where ideologically diverse audiences converge online, though our analysis also reveals that more than half of the US online population consumes no online news, underlining the risk of increased information inequality driven by self-selection along lines of interest. For this study, we use an unprecedented combination of observed data from the United States comprising a 5-y time window and involving tens of thousands of panelists. Our dataset traces news consumption across different devices and unveils important differences in news diets when multiplatform or desktop-only access is used. We discuss the implications of our findings for how we think about the current communication environment, exposure to news, and ongoing attempts to limit the effects of misinformation.

audience networks | news consumption | selective exposure | web tracking | computational social science

Digital technologies have drastically changed the way in which people access news and political content. Audiences today have more choices and flexibility when it comes to consuming content, but what is still unclear is the impact that those choices have on overall patterns of news consumption. The answer to this question is important because democracies rely on informed citizens. Past research has often interpreted the abundance of media options as leading to narrower, more selective news diets. This would result in increasing audience fragmentation (i.e., audiences disperse among the higher number of choices) and decreasing coexposure (i.e., audiences self-select into news outlets that are aligned with their preconceived views) (1). There is evidence that Internet penetration boosted self-selection (2) and that social media platforms allow users to tailor news diets aligned with their preconceived ideological positions (3). However, there is also evidence that offline interpersonal relations exhibit the highest level of ideological segregation (4, 5), which casts doubts on the polarizing effects of digital technologies. Research has documented an increase in partisan news consumption over time, with surges ahead of salient political events (6) and fluctuations that depend on the nature of the story reported (7, 8). However, with the exception of niche outlets with low reach, partisan differences across outlets are small even when taking into account opinionated news content. Many of the mainstream media that account for the largest shares of attention online generally have politically diverse audiences and are ideologically centrist, and most seek to be impartial in their news coverage (9, 10). This results in a less partisan media environment than commonly accepted (11). Numerous studies depict news diets that are more varied than often assumed, at least for those who consume news (9, 12). Some provide evidence that

suggests a significant overlap between Democrats' and Republicans' news diets (13). Others find coexposure to news sources more consistent with diverse diets than segregated ones (14–16). There is a small portion of highly interested and partisan audiences that self-select into ideologically consistent news sites (13), and some of them may inhabit echo chambers, but these audiences do not represent the vast majority of news users who generally use diverse, and cross-cutting news sources—as we show in the analyses below.

This paper offers unprecedented evidence that advances our understanding of the online news ecosystem. In particular, we use observational data from the United States that combines web tracking behavior and self-declared ideological alignment (*Materials and Methods*). We use these data to determine if audiences are becoming more fragmented in their exposure to news sources. For this, we assess patterns of coexposure to news sources and test if the selective exposure hypothesis (the idea that people purposefully select sources that match their political beliefs) (17), receives consistent empirical support. We analyze mobile and desktop tracking data for a period of 5 y and we characterize the ideological alignment of news audiences for a period of 16 mo. We document the differences in exposure estimates that arise when only desktop access is analyzed (as most prior research does, despite the fact that more than half of digital media use is now mobile; see *SI Appendix*, Fig. SI4). Desktop estimates exclude exposure through mobile phones and tablets, which our data include in the form of multiplatform statistics. Our results suggest that multiplatform estimates yield a very different picture

Significance

Access to diverse news strengthens democratic citizenship. Whether digital technologies have narrowed or widened news diets fosters contentious debates. Previous research shows the abundance of digital news sources might be leading to more fragmented audiences, ideological segregation, and echo chambers. Our study resorts to an unprecedented combination of data to show that the increase in mobile access to news actually leads to higher exposure to diverse content and that ideological self-selection explains only a small percentage of co-exposure to news. We also find that more than half of Internet users in the United States do not use online news. Future research should avoid generalizations from desktop-only data and pay attention to the increasing divide between informed citizens and news avoiders.

Author contributions: S.G.B. designed research; T.Y., S.M.V., R.K.N., and S.G.B. performed research; T.Y. and S.G.B. analyzed data; and T.Y., S.M.V., R.K.N., and S.G.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: sgonzalezbailon@asc.upenn.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2006089117/-DCSupplemental>.

First published October 30, 2020.

of how the online media environment has evolved during the past few years.

Results

We define exposure to news content as users accessing the web domain of news organizations. Fig. 1 shows changes in the percentage reach of news outlets and changes in user engagement with those outlets (measured as the average time spent on their domains). The figure shows that there is a significant difference in the estimates extracted from desktop-only and multiplatform data. As the time series and yearly distributions show, desktop measures underestimate exposure to news and the average time that users spend consuming content from news sites. We assess if these differences in measurement change the conclusions we can draw regarding fragmentation by first looking at coexposure and

then determining how much of that coexposure responds to ideological self-selection.

Using statistics of audience overlap, we built monthly coexposure networks where the nodes are news outlets and the ties measure the number of unique users that visited two news sources in a given month. We interpret these networks as measuring similarity in the audience base of news outlets. As Fig. 2A shows, the networks can look very different, depending on how much overlap exists across news domains. We capture these differences in the coexposure networks with the measure of network density (see *Materials and Methods* for definitions). Media landscapes where audiences are fragmented would have density values approximating 0. Density scores are based on the presence of ties, but in the context of our data coexposure ties are very heterogeneous in the magnitude of the overlap (Fig. 2B). To eliminate the weakest ties (e.g., those that may result from random browsing

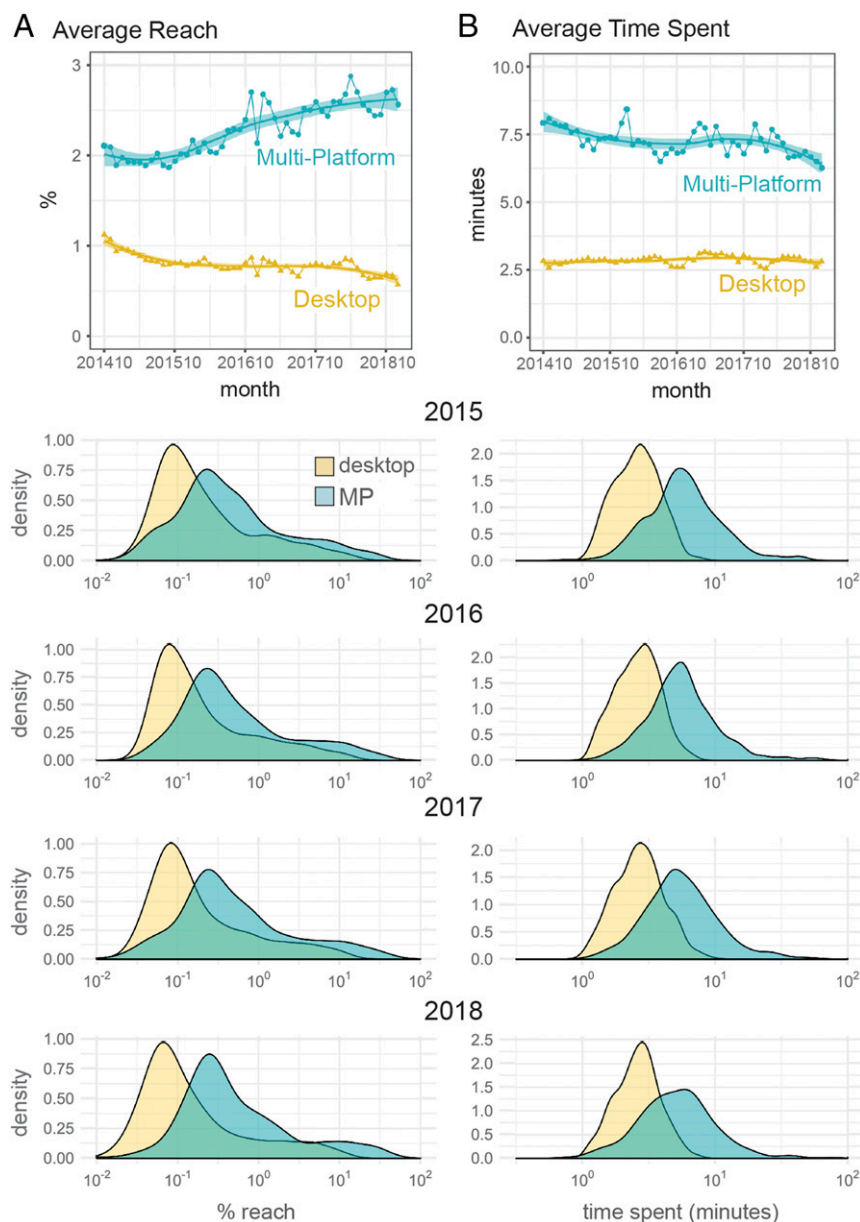


Fig. 1. Audience reach (A) and user engagement (B) for news outlets according to desktop-only and multiplatform estimates. Past research has mostly analyzed web log data reflecting desktop usage. Our data show that desktop-only measures greatly underestimate the number of users exposed to news sources and the time they spend on those sites.

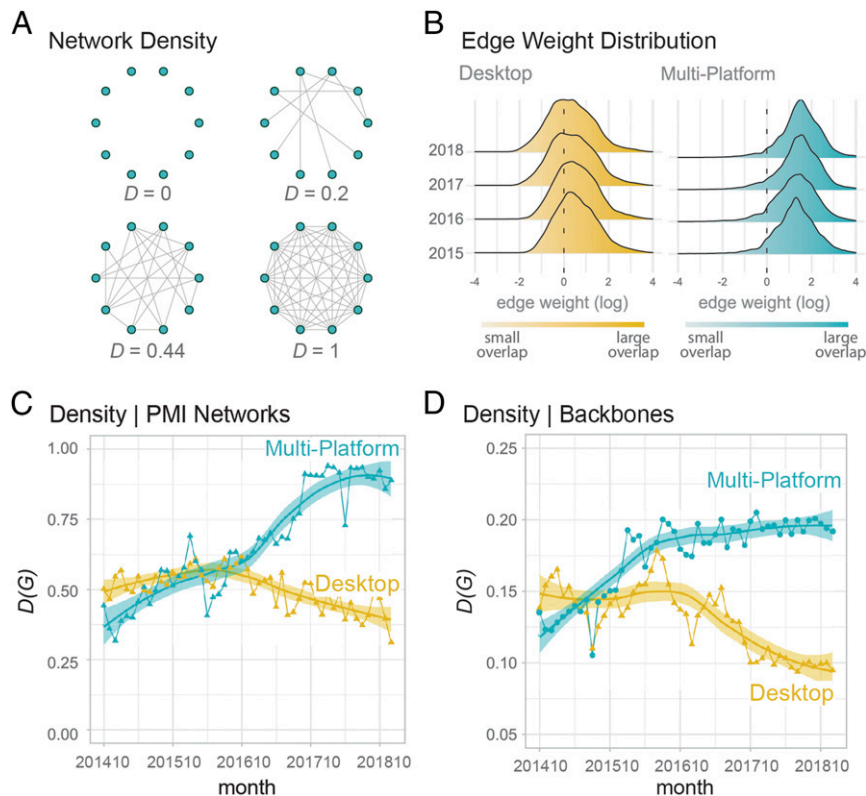


Fig. 2. Networks of audience overlap across news sources. The nodes in these networks are news outlets and the ties measure the number of users coexposed to those outlets. (A) Network density measures the number of existing ties out of all possible ties that could exist in the network. In an empty network with no audience overlap the density is 0; in a complete network (where all outlets share some audience) the density is 1. (B) Edge weight distribution for desktop and multiplatform data (expressed in thousands). Density scores are based on the presence of ties, but coexposure ties are very heterogeneous in terms of audience overlap. To eliminate the weakest ties, we apply two filtering techniques to our data: PMI and backbone extraction. (C and D) Networks of audience overlap grow denser over time, regardless of the filtering technique used: The pattern remains visible even when only the most significant ties are preserved (see *SI Appendix* for trends in the raw, unfiltered data). This trend suggests that exposure to news is becoming more diversified and less fragmented. Data that use desktop-only estimates lead to the opposite conclusion.

behavior) we apply two filtering techniques to the networks (see *Materials and Methods* and *SI Appendix* for results for unfiltered, raw data). Fig. 2 C and D summarize changes in the density of these filtered networks. Although the overall density levels change (e.g., the backbone approach is more stringent in eliminating ties as insignificant), the patterns are consistent: The networks built with multiplatform estimates grow substantially denser over time, suggesting increasing coexposure (and a widening of news diets). The analyses with desktop data lead to the opposite conclusion.

To determine if this increasing coexposure results from increasing selection along ideological lines, we use two measures of segregation (see *Materials and Methods* for definitions). Fig. 3 illustrates the distribution of these two measures. In Fig. 3 A and B we show changes in the segregation index. As the figures show, segregation levels are, overall, low (for reference, the score reported in prior research using 2009 data from the same source is 0.075 for desktop access, averaged over 12 mo, which suggests that overall levels of online segregation have decreased since then). Consistently, the local average mean for multiplatform data is lower or equal to the desktop data. The differences in the monthly scores are statistically significant according to a Wilcoxon signed-rank test ($P = 0.003$ for party identification and $P < 0.001$ for political outlook; see *SI Appendix*).

In Fig. 3C we summarize the distribution of the second measure: the favorability scores assigned to news outlets. These distributions show that the audiences consuming news in our data tend to lean Democratic but also toward conservative views (see *SI Appendix* for monthly distributions). There is a moderate association

between party identification and political outlook (Fig. 3D). The magnitude of this correlation is consistent with survey data collected through the American National Election Studies (e.g., ref. 18). Fig. 3 E and F zoom into the top 20 sites ranked by percentage reach. These plots show that the sites with a larger audience base converge toward the median of the distributions, but also that their alignment in the continuum has face validity. [Two caveats here are that, as noted, our audience leans Democrat and that ideological differences among mainstream news media is very small, as previous research assigning ideological scores based on the analysis of content has also shown (11)].

The aggregated segregation scores already suggest that ideological selective exposure is not a prominent factor in news consumption: The score is much closer to 0 than to 1, and it has actually decreased compared to what was reported in previous research (4). To further quantify the impact of ideology on coexposure, we calculate the correlation between the coexposure network and matrices measuring pairwise distance for the two ideological scores (*Materials and Methods*). As Fig. 4 shows, the correlations are negative and highly significant: This means that the further two news sites are in the ideological distributions, the lower their audience overlap is. The results of the regression models (Fig. 4C) confirm that distance in both party affiliation and political outlook have a significant and negative effect on overlap. This result, which we interpret as evidence of selective exposure, is robust across the three networks we consider (see *SI Appendix* for additional robustness tests). However, ideology explains at most 16% of the variation in coexposure to news,

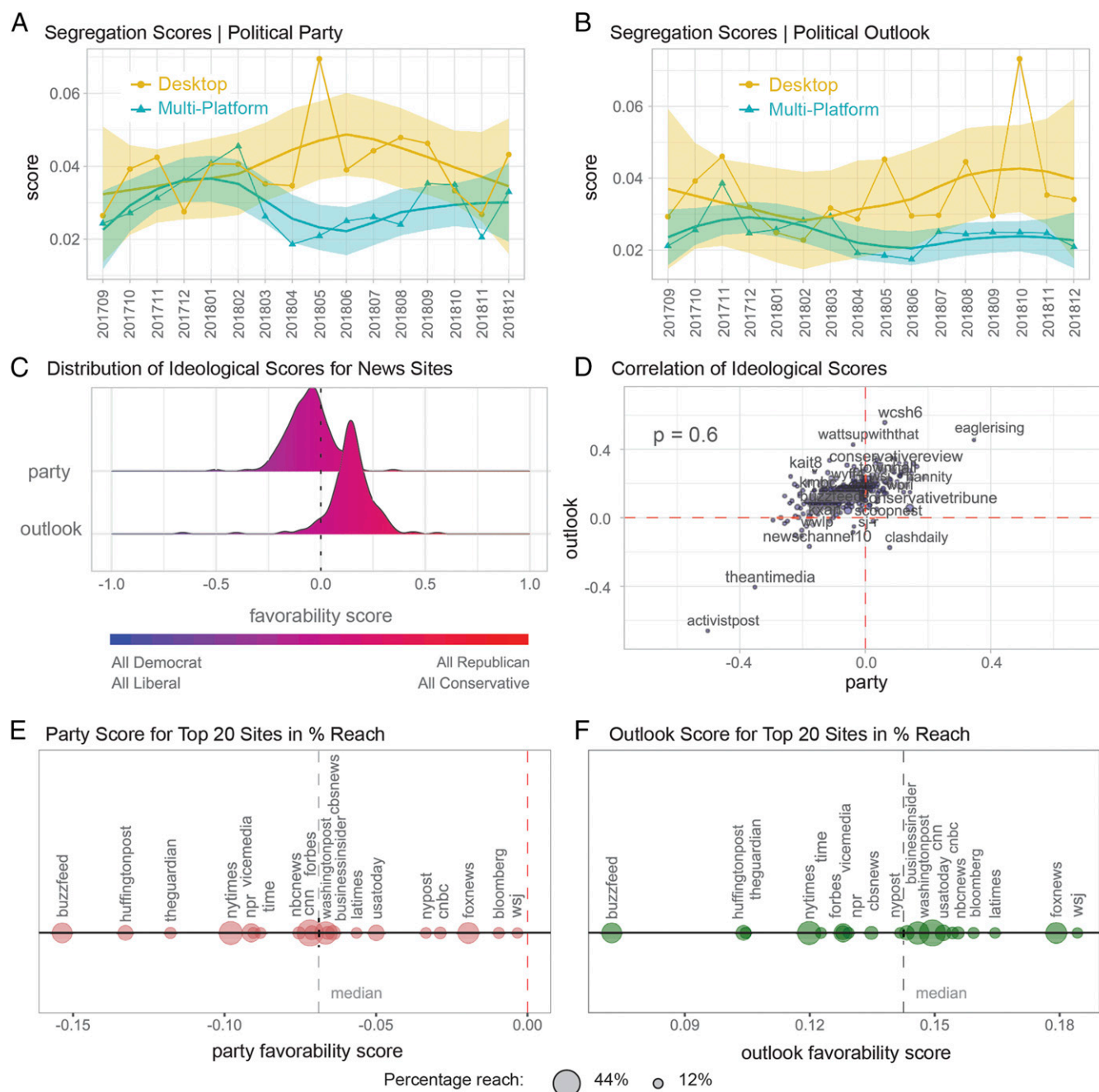


Fig. 3. Ideological scores assigned to news outlets using their audience composition. (A and B) Changes in the segregation scores (*Materials and Methods*) for desktop and multiplatform data. Consistently, the local average means (smooth line) for multiplatform data are lower or equal to the desktop data, suggesting lower segregation levels when mobile access is considered. (C) Distribution of favorability scores for party affiliation and political outlook (*Materials and Methods*). The audiences consuming news in our data tend to lean Democrat and toward conservative views. (D) Scatterplot showing a moderate association of party affiliation and political outlook. (E and F) Alignment of top 20 outlets in the favorability scales. The news sites with higher percentage reach (and, therefore, a larger audience base) converge toward the median of the favorability scores.

which means that selection along ideological lines is not the only mechanism explaining news exposure (especially for sites with the strongest audience overlap: When the backbone network is used as dependent variable, ideology explains less than 4% of the data).

Discussion

Many Americans have strong political beliefs, and some actively dislike and distrust those sympathetic to the other side. This means that, in theory, many people could be engaging in extensive selective

exposure when they consume news, an effect hypothesized to become more pronounced as people have more media choices (19, 20). The selective exposure effects we identify here are significant over time, but despite the growth in media choice these effects do not seem to grow in magnitude. Most importantly, the patterns of ideological segregation that we identify here explain a small percentage of the variance in news diets, while coexposure grows denser and, therefore, less fragmented. This is consistent with previous research analyzing longitudinal data (4, 13) and with

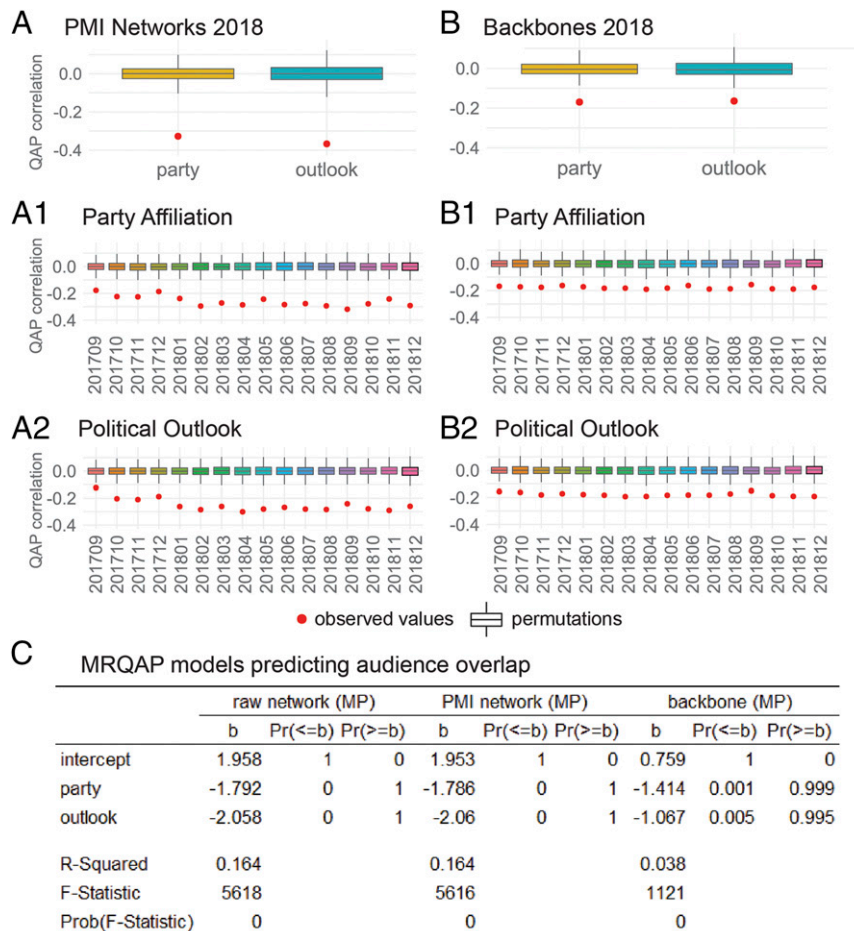


Fig. 4. Correlation of audience overlap and ideological scores. The correlation of the coexposure network and pairwise distance in the ideological distributions is negative and highly significant in both the PMI networks (A) and backbones (B). Boxplots provide nonparametric confidence intervals based on random permutations. The correlation is negative because audience overlap decreases the further news sites are in the ideological distributions shown in Fig. 3C. The magnitude of the correlation does not change substantially over time and there is no apparent trend. (C) Regression results with the raw and filtered networks as dependent variable (data aggregated for 2018). The three models show consistent results: audience overlap is higher between news sites that are ideologically similar, according to the favorability scores. However, these two variables still leave much unexplained structure in the coexposure networks: the most predictive models explain just 16% of the data (much less when only the strongest overlap, or the backbone of the network, is considered).

studies on the mechanisms that shape online news diets, mostly aggregators and social media, which prior research has shown can actually boost coexposure to diverse content (12, 21, 22).

One important caveat is that our analyses rely on traffic data at the top-level domain. A more fine-grained analysis of Uniform Resource Locator-level data (URLs) might yield different estimates of exposure to news: Most people accessing news outlets on the web are not necessarily consuming political content but rather entertainment or sports, as previous research has documented (6, 11). Another important limitation is that our measures of exposure to content rely on visits and time spent, which offer a narrow understanding of actual engagement with the news: These measures do not really tell us how people process the information they read to make it compatible (or not) with their beliefs. However, our analyses do suggest that we should be cautious in interpreting the results of prior work that relies on desktop-only estimates: Such data miss an important and growing segment of online news exposure and, as we show, that gap leads to a substantive difference in findings. Our data also do not include all social media activity. The statistics we analyze only include users that click on links posted on platforms like Facebook or Twitter and visit the corresponding domains. Likewise, messaging applications are outside the scope of our analyses, even though they

have also become an important channel for news consumption (10). There is very little research in this area in part due to the difficulties (logistical and ethical) of accessing and collecting data generated in private, end-to-end encrypted environments. Still, social media and messaging application activity create blind spots in our assessment of the news ecosystem. Selective exposure on those platforms might be on the rise.

More generally, our results point to a much more profound form of selective exposure that goes beyond the traditional focus on ideological segregation and partisan polarization, namely the growing inequality between those who consume news and those who, despite the abundance of offers and ease of access, consume little or no news at all (20). Political ideology may matter less for online news consumption than previous research suggests, with levels of interest in political news being a more significant determinant of online exposure. Less than half of the online population consumes news on the web [with recent research suggesting that news consumption of any sort is, in fact, heavily outweighed by other forms of media consumption, “comprising at most 14.2% of Americans’ daily media diets” (23)]. This creates an even more urgent problem for democracies to the extent that this opt-out population is also more likely to be susceptible to misinformation and other forms of manipulation. Future research

should pay more attention to whether populations with little or no exposure to news are also more vulnerable to misinformation campaigns or echo-chamber effects.

Materials and Methods

Data. We use two main sources of observational data: 1) Comscore's Media Metrix panel, which is a representative panel of the online population that offers estimates for desktop and multiplatform media usage (monthly aggregates at the domain level from October 2014 to December 2018, $N \sim 250,000$ panelists) and 2) Comscore's Plan Metrix panel, which offers data on audience demographics, including partisan affiliation (i.e., Democrat–Republican self-placement) and political outlook (i.e., liberal–conservative self-placement), provided as monthly aggregates at the domain level (from September 2017 to December 2018, $N \sim 20,000$). A full list of the news outlets we analyze ($n = 424$) as well as more information about the characteristics of the data can be found in *SI Appendix*.

Methods. We build the networks of audience overlap using the number of unique users coexposed to a pair of news sources for a particular period (e.g., a month or year). We analyze these weighted networks as built using the observational data (what we call “raw networks”; see *SI Appendix*) and after applying two filtering techniques that only retain ties with a statistically significant overlap, that is, more audience overlap than expected by chance. The first version, which we call “PMI networks,” uses the pointwise mutual information measure of association. It is defined on pairs of news sites (i, j) as $PMI(i, j) = \log \frac{p(i, j)}{p(i)p(j)}$, where $p(i, j)$ is the fraction of all online users that access both sites and $p(i)$ and $p(j)$ is the fraction of all online users accessing each site. These filtered networks only retain the edges with a PMI larger than zero. The second version is what we call the “backbones,” which result from a filtering procedure that uses a null model of equal edge weight distribution at the local (focal-node) level to determine which edges depart significantly from the null model (24). These two techniques differ in the threshold of statistical significance: The backbone approach is more stringent; the PMI approach eliminates fewer ties compared to the raw version of the data. We analyze these two types of filtered networks to assess the sensitivity of our results to choices in the construction of the coexposure networks and to exclude the possibility of noise driving the results.

Measures. Graph density is defined as the ratio of the number of observed edges and the number of possible edges, given the size of the network. For undirected, simple networks

$D_t = \frac{2E_t}{V_t(V_t-1)}$, where E_t is the number of edges at a particular time t and V_t is the number of vertices or nodes at a particular time t (25). We calculate the “segregation score” following ref. 4 to create an index of audience isolation in terms of ideology. The score for the news sites in our list is calculated as

$$S_t = \sum_{i \in I} \left(\frac{R_i}{R} \times \frac{R_i}{V_t} \right) - \sum_{i \in I} \left(\frac{D_i}{D} \times \frac{R_i}{V_t} \right), \text{ where } \frac{R_i}{V_t} \text{ is the share of Republican (or conservative) visits for site } i.$$

The segregation score approximates 1 when Republicans only visit Republican outlets and Democrats only visit Democrat outlets, and it approximates 0 when Republicans and Democrats visit the same outlets. We also calculate the “favorability scores” following ref. 6 to create another measure of ideological diversity in the audience of news sites. We assign a favorability score to a news site i in time t according to the formula $fav(i_t) = \frac{(R-D)}{(R+D)}$, where R stands for the audience share of a news site that self-identify as Republican and D stands for the share that self-identify as Democrat. The score equals 1 when the news site domain is visited exclusively by Republicans and -1 when it is visited exclusively by Democrats (0 means that Republicans and Democrats are equally likely to visit the site). In addition to party affiliation, we also calculate segregation and favorability scores for “political outlook,” where the categories are C for conservative and L for liberal. We calculate these scores monthly and, for some analyses, we aggregate them yearly. See *SI Appendix* for more details on these measures and monthly distributions.

Models. Since we analyze relational data, we test the association of coexposure and ideology using the multiple regression quadratic assignment procedure (MRQAP), a permutation test for multiple linear regression developed to analyze data organized in square matrices (26–28). The variables in these matrices indicate audience overlap (according to the raw network data and the filtered networks) and distance between news sources in the distribution of favorability scores for party affiliation and political outlook. See *SI Appendix* for additional details on specifications.

Data Availability. The data and code necessary to replicate the findings reported in this article can be accessed in the Open Science Framework repository (<https://osf.io/hp5fm/>) (29).

ACKNOWLEDGMENTS. Work on this project was funded by NSF grant 1729412 and it was completed while S.G.B. was on sabbatical leave at the Center for Advanced Study in the Behavioral Sciences at Stanford University. The authors thank Subhayan Mukerjee, Ariadna Net, Felix Simon, and Yayoi Teramoto for research assistance as well as Federico Sieder, Queralt Simó, and Isabel Woodford for collaboration on the data collection process.

1. S. Iyengar, K. S. Hahn, Red media, blue media: Evidence of ideological selectivity in media use. *J. Commun.* **59**, 19–39 (2009).
2. Y. Lelkes, G. Sood, S. Iyengar, The hostile audience: The effect of access to broadband internet on partisan affect. *Am. J. Pol. Sci.* **61**, 5–20 (2017).
3. E. Bakshy, S. Messing, L. Adamic, Exposure to ideologically diverse news and opinion on facebook. *Science* **348**, 1130–1132 (2015).
4. M. Gentzkow, J. M. Shapiro, Ideological segregation online and offline. *Q. J. Econ.* **126**, 1799–1839 (2011).
5. D. C. Mutz, P. S. Martin, Facilitating communication across lines of political difference: The role of mass media. *Am. Polit. Sci. Rev.* **95**, 97–114 (2001).
6. M. Tyler, J. Grimmer, S. Iyengar, “Partisan enclaves and information bazaars: Mapping selective exposure to online news” (Stanford University, 2019).
7. E. Peterson, S. Goel, S. Iyengar, Partisan selective exposure in online news consumption: Evidence from the 2016 presidential campaign. *Political Sci. Res. Methods*, 10.1017/psrm.2019.55 (2019).
8. A. S. Cardenal, C. Aguilar-Paredes, C. Cristancho, S. Majó-Vázquez, Echo-chambers in online news consumption: Evidence from survey and navigation data in Spain. *Eur. J. Commun.* **34**, 360–376 (2019).
9. S. Flaxman, S. Goel, J. Rao, Filter bubbles, echo chambers, and online news consumption. *Public Opin. Q.* **80**, 298–320 (2016).
10. N. Newman, R. Fletcher, A. Kalogeropoulos, R. K. Nielsen, Reuters Institute digital news report 2019. <http://www.digitalnewsreport.org/uncategorized/2019/explore-the-2019-report/>. Accessed 27 October 2020.
11. C. Budak, S. Goel, J. M. Rao, Fair and balanced? Quantifying media bias through crowdsourced content analysis. *Public Opin. Q.* **80**, 250–271 (2016).
12. M. Scharnow, F. Mangold, S. Stier, J. Breuer, How social network sites and other online intermediaries increase exposure to news. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 2761–2763 (2020).
13. A. Guess, “(Almost) everything in moderation: New evidence in Americans’ online media diets” (2020). *Am. J. Pol. Sci.*, in press.
14. R. Fletcher, R. K. Nielsen, Are news audiences increasingly fragmented? A cross-national comparative analysis of cross-platform news audience fragmentation and duplication. *J. Commun.* **67**, 476–498 (2017).
15. S. Majó-Vázquez, R. K. Nielsen, S. González-Bailón, The backbone structure of audience networks: A new approach to comparing online news consumption across countries. *Polit. Commun.* **36**, 227–240 (2018).
16. J. G. Webster, T. B. Ksiazek, The dynamics of audience fragmentation: Public attention in an age of digital media. *J. Commun.* **62**, 39–56 (2012).
17. N. J. Stroud, “Selective exposure theories” in *The Oxford Handbook of Political Communication*, K. Kenski, K. H. Jamieson, Eds. (Oxford University Press, New York, 2018), pp. 531–548.
18. R. N. Lupton, S. M. Smallpage, A. M. Enders, Values and political predispositions in the age of polarization: Examining the relationship between partisanship and ideology in the United States, 1988–2012. *Br. J. Polit. Sci.* **50**, 241–260 (2017).
19. S. Iyengar, Y. Lelkes, M. Levendusky, N. Malhotra, S. J. Westwood, The origins and consequences of affective polarization in the United States. *Annu. Rev. Polit. Sci.* **22**, 129–146 (2019).
20. M. Prior, *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections* (Cambridge University Press, 2007).
21. P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, R. Bonneau, Tweeting from left to right: Is online political communication more than an echo chamber? *Psychol. Sci.* **26**, 1531–1542 (2015).
22. D. Lazer, Studying human attention on the Internet. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 21–22 (2020).
23. J. Allen, B. Howland, M. Mobius, D. Rothschild, D. J. Watts, Evaluating the fake news problem at the scale of the information ecosystem. *Sci. Adv.* **6**, eaay3539 (2020).
24. M. Á. Serrano, M. Boguñá, A. Vespignani, Extracting the multiscale backbone of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6483–6488 (2009).
25. S. Wasserman, *Social Network Analysis: Methods and Applications* (Cambridge University Press, 1994).
26. D. Krackhardt, Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Soc. Networks* **10**, 359–381 (1988).
27. D. Krackhardt, QAP partialling as a test of spuriousness. *Soc. Networks* **9**, 171–186 (1987).
28. D. Dekker, D. Krackhardt, T. A. B. Snijders, Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika* **72**, 563–581 (2007).
29. T. Yang, S. Majó-Vázquez, S. González-Bailón, Exposure to news grows less fragmented with increase in mobile access. OSF. <https://osf.io/hp5fm/>. Deposited 31 March 2020.