Online Media Boosts Exposure to News but Only for a Small Minority of Hyper-Consumers

Tian Yang[1] and Sandra González-Bailón[1,*]

[1] Annenberg School for Communication, University of Pennsylvania, 19104 Philadelphia, PA, US

* Corresponding author: Sandra González-Bailón

Email: sgonzalezbailon@asc.upenn.edu

Classification

Social Sciences.

Keywords

News exposure; media choice; web tracking; computational social science.

Author Contributions

Research design: SGB. Data analysis: TY and SGB. Writing: TY and SGB.

This PDF file includes:

Main Text.
Figures 1 to 4.

Abstract

We analyze TV, web, and YouTube media consumption data for N = 55,000 panelists over a period of 44 months to identify the incidence and demographic correlates of news consumption across media channels. Less than 10% of the panelists (N ~ 5,300) view and browse news on the three platforms. This small group of news hyper-consumers is formed predominantly by older male users with higher education. We find no evidence of substitution effects in the time these users spend consuming news on each of the three media channels: controlling for demographics and random effects, an increase on news time on one platform has a positive impact on news time on the other two platforms. Our results uncover important demographic divides in how audiences navigate a high-choice media environment, with only an unrepresentative minority of users engaging with news content across the media landscape. These results also highlight the importance of using large panel data to identify behaviors that are statistically rare and, as such, unlikely to be detected with smaller panels.

Significance Statement

The quality of our democracies relies on the quality of the information that citizens consume but we still know very little about how citizens engage with the news "in the wild". We answer this question by analyzing the media choices of a representative panel of the U.S. population as they consume TV, web, and YouTube content. We find that those interested in news spend more time consuming news across channels, but this is a choice that only a small and unrepresentative slice of the population makes. Measuring demographic divides helps us identify groups of people indifferent to current affairs or alienated from politics, and those more vulnerable to misinformation given their lack of systematic exposure to news sources.

Main Text

Digital technologies have decentralized the media landscape, allowing users to opt out of news, or opt into sources that are aligned with their views. The range of options available and the different individual propensities to consume different types of content have translated into a media landscape characterized by long tails. On the supply side, there are many sources attracting small numbers of people, creating the niches that add up to digital audiences (1). On the demand side, only a small fraction of people actively engage with news content, compensating with their rare intensity for the lack of interest shown by the majority (2, 3). However, we still know strikingly little about how people consume news 'in the wild', partly because of the difficulty of piecing together the different channels they use to access news: observational datasets tracking media behavior for the same group of people over time and across platforms (e.g., TV, web, social media) are rare. And yet our determination of whether digital technologies are good or bad for democracy depends on being able to analyze this type of data, contextualizing online behavior within the larger media landscape and the choices people make.

Here, we analyze panel data tracking media consumption across TV, the web, and YouTube to address three questions that have received limited answers in prior research: How many people consume news across sources and media platforms? Do they have a characteristic demographic profile? And does their behavior exhibit evidence of substitution effects in the time spent consuming news? The answers to these questions matter because they help identify subsets of the population more likely to disengage from the news and the rest of the democratic process (4). Prior research provides limited answers because it either relies on self-reported measures (which are known to have accuracy issues, 5); analyzes panels that are too small to capture activity on the tail of viewing/browsing distributions (6); or uses aggregated measures that are not linked to individual-level behavioral and demographic profiles (3, 7). Here we analyze datasets that overcome those limitations to provide evidence that casts both a positive and a negative light on digital technologies. On the positive side, we find that time spent on news consumption correlates positively across platforms, which means that, contrary to the substitution hypothesis, digital media boosts engagement with the news. On the negative side, we find that this is true only for a small slice of the population that is far from being representative and that reveals a clear gender divide.

The data we analyze was provided by a media measurement company that logs viewing activity on TV and browsing activity on the web for hundreds of thousands of consenting panelists (see Materials and Methods and the SI for more details on the data). Figure 1 offers a high-level description of our data. The panels we analyze allow us to map co-exposure to different news sources within and across media platforms (panel A). From the supply side, YouTube offers many more choices of news sources than TV or the web (panel B), but from the demand side it clearly lags substantially behind in terms of reach (or number of unique panelists visiting these sources, panel C). Unsurprisingly, TV is still the most common source of news for most people (in line with what prior research has shown, 7) and its audience is similarly concentrated to the audience consuming news on the web (i.e., most people consuming news on these two platforms gravitate around the same small number of sources); YouTube is, by comparison, much more decentralized (panel D).

Results

Figure 2 offers the first part of our answer to the first question (How common are people that consume news across sources and media platforms?) About half of our panelists consume news from four or less TV channels, the other 30-40% obtain news from five or more channels (note that there is a decreasing trend in this category, though). This suggests that news diets contain some diversity -- at least in terms of sources accessed -- for most TV viewers. The prevalence of TV as a source of news remains stable over the period we consider (panel A). There is also evidence of co-exposure to different sources on the web, but on this platform, less than half of the panelists are consuming news and there is an even sharper decreasing trend for those in the upper tail of the distribution (those consuming 5 sources or more, panel B). This decreasing trend likely results from the fact that our web browsing data does not include mobile activity, which prior research has shown deflates estimates of news consumption (3). The trends for news exposure on YouTube are clearly going in the opposite direction (panel C). This rising trend is consistent with prior research (8, 9). On a monthly basis, however, less than 5% of the panelists view videos classified in the 'news' category.

Figure 3 offers the second part of our answer to the question of hyper-consumers prevalence. There is a small minority of panelists (less than 10%) that consume news from all three channels (panel A). As we detail in table SI1 in the SI, only a very small fraction of all panelists, i.e., 0.04%, accessed news on YouTube only; most news consumers obtain their news from TV first, with the web and, to a much lesser extent, YouTube complementing their news diets. Panels B and C in figure 3 shift attention to our second question, i.e., do hyper-consumers have a characteristic demographic profile? The most important correlates of news consumption across platforms are education and age: older and better educated panelists are more likely to consume news on TV, the web, and YouTube. Employment status, income, and gender are negatively correlated with the hyper-consumption of news: fully employed, high income individuals are less likely to access news on the three platforms, and so are women. (Note that women are more likely than men to consume news on TV, but less likely on the web and on YouTube; see Materials and Methods for details on the regression models used and table SI2 in the SI for the full regression outputs).

To determine if media choice leads to substitution effects, we run additional models using the time spent consuming news as the dependent/independent variable for each pair of media channels. Figure 4A shows the distribution of this variable (time spent) for TV, the web, and YouTube. The expectation under the hypothesis of substitution effects is that time spent consuming news on one platform will have a negative impact on time spent consuming news on the other two platforms. Contrary to this expectation, we find that all the effects are positive and significant (except for time on YouTube, which is not significantly increased by TV news time). Figure 4B shows the estimated coefficients controlling for all the other covariates (i.e., fixed and random effects, see table SI5 in the SI for full regression outputs and table SI6 for additional robustness tests).

Discussion

News offer one of the main mechanisms that allow citizens to monitor those in power and stay abreast of political events. Digital technologies have allowed audiences to be more difficult to capture by multiplying the media choices individuals can make. Understanding how people make those choices is therefore central to our assessment of democratic governance and its reliance on an informed citizenry. Our findings suggest that online media help achieve this ideal by boosting news consumption: rather than triggering substitution effects, we find that online media amplify time spent consuming news. This amplification effect, however, only benefits those who are already interested in the news – and this is a small fraction of the population (less than 10%). This finding reveals important demographic divides. Women, in particular, are clearly less likely to use online sources to consume news, even after controlling for education. The existence of gender gaps in the consumption of news has long been documented by survey research (e.g., 10). That these divides remain visible more than twenty years later and in a high-choice media environment is suggestive of the constraints that many individuals still face to meet the demands of the democratic ideal.

One important limitation of our analyses is that web browsing (including YouTube activity) relies on desktop-only browsing behavior, which prior research has shown underestimates exposure to news (3). However, our data allows us to connect individual behavior across media platforms and therefore draw a more complete picture of media choices and the demographic characteristics associated to those choices. Likewise, our panel allows us to cast light on relatively rare behavior: less than 1 in 10 panelists consume news across platforms, and this requires having large-enough datasets to tap into this tail of the distribution. One more advantage of our panel data is that it allows us to control for temporal variability and other confounders at the individual level not captured by the demographics. Our findings, in other words, provide important new evidence that no prior research has provided yet. However, one priority for future research is to test if the patterns of media choice we identify hold once mobile access is considered (data that, to the best of our knowledge, does not exist yet).

More generally, our results help us characterize the digital equivalent of the 'opinion leaders' figure first proposed to understand the effects of mass media (11, 12). These are the small group of individuals that consume media content and then pass it on through their networks of interpersonal contacts, triggering (potentially) a chain of social influence in opinion formation. In this digital age, those networks are largely mediated by internet technologies like social media. The hyper-consumers we identify in our analyses create the elite of opinion leaders that are then likely to propagate their own curated news through their networks. These leaders have a disproportionate influence in how news content is selected, circulated, and (ultimately) algorithmically amplified. The fact that this small group of opinion leaders is far from representing the population at large unveils one of the ways in which the information circulating online may perpetuate important biases in the salience of some topics over others. Future research should aim to cast light on the impact that hyper-consumers have on information bias and other types of information inequities propagating online.

Materials and Methods

Data. Our data is time-stamped and spans the period January 2016 to August 2019. The TV data is disaggregated at the program level, and each program is classified using a number of categories (including 'news') provided by the media measurement company. The web log data contains the specific URLs visited by the panelists, which allowed us to identify activity within YouTube and obtain additional information via the platform's API, like the channel publishing the videos, or the category under which those videos are classified (i.e., 'news'). For the rest of the web data, we identified the top-level (registered) domain of the URLs visited and we matched those domains with a list of N = 813 domains identified in prior research as being 'news' sources (these five prior studies are 3, 13-16). We found visits to N = 795 of these domains. The TV and web panels each contain information on media activity for N ~ 300,000 unique panelists. About 55,000 of the panelists overlap across the two data sets; this overlapping set is the data we analyze here. See the SI for more details on data collection, data representativeness, longevity of panelists, and robustness checks using the full set of panelists in each data set.

Methods. To identify the demographic profile of hyper-consumers (i.e., panelists consuming news on TV, the web, and YouTube), we fit linear mixed-effects models using panelist ID and month as random effects. We used two dependent variables: a binary that records whether the panelists consumed news in each platform/all three (models reported in figure 3 and in table SI2, models 1-4, in the SI), and an alternative operationalization of the DV that counts the number of channels/websites visited (results reported in table SI2, models 5-7 in the SI). Results are consistent across these alternative operationalizations. To test the substitution effect hypothesis, we also fit linear mixed-effects models (again using panelist ID and month as random effects) with time spent consuming news as the dependent/independent variables for each pair of media platforms (see tables SI5 and SI6 in the SI for full regression outputs for mixed effects and group fixed effects models).

Acknowledgments

We are grateful to Harmony Labs for facilitating access to the data and for crucial research support. Work on this paper was partly funded by NSF grants 1729412 and 2017655.

References

1. N. J. Stroud, *Niche News: The Politics of News Choice* (Oxford University Press, New York, 2011).
2. A. Guess, J. Nagler, J. Tucker, Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances* **5**, eaau4586 (2019).
3. T. Yang, S. Majó-Vázquez, R. K. Nielsen, S. González-Bailón, Exposure to news grows less fragmented with an increase in mobile access. *Proceedings of the National Academy of Sciences* **117**, 28678-28683 (2020).
4. M. X. Delli Carpini, S. Keeter, *What Americans Know about Politics and Why it Matters* (Yale University Press, New Haven, 1996).
5. D. A. Parry *et al.*, A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour* 10.1038/s41562-021-01117-5 (2021).
6. A. Guess, B. Nyhan, J. Reifler, Exposure to Untrustworthy Websites in the 2016 U.S. Election. *Nature Human Behaviour* **4**, 472-480 (2020).
7. J. Allen, B. Howland, M. Mobius, D. Rothschild, D. J. Watts, Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances* **6**, eaay3539 (2020).
8. H. Hosseinmardi *et al.*, Evaluating the scale, growth, and origins of right-wing echo chambers on YouTube. *arXiv:2011.12843* (2020).
9. K. Munger, J. Phillips, Right-wing YouTube: A supply and demand perspective. *International Journal of Press-Politics* 10.1177/1940161220964767 (2020).
10. S. Verba, N. Burns, K. L. Schlozman, Knowing and Caring about Politics: Gender and Political Engagement. *The Journal of Politics* **59**, 1051-1072 (1997).
11. E. Katz, P. Lazarsfeld, *Personal Influence. The Part Played by People in the Flow of Mass Communications* (Free Press, New York, NY, 1955).
12. P. Lazarsfeld, B. Berelson, H. Gaudet, *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign* (Columbia University Press, New York, NY, 1948).
13. E. Bakshy, S. Messing, L. A. Adamic, Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130-1132 (2015).
14. C. Budak, S. Goel, J. M. Rao, Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis. *Public Opinion Quarterly* **80**, 250-271 (2016).
15. N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, Fake news on Twitter during the 2016 U.S. presidential election. *Science* **363**, 374-378 (2019).
16. E. Peterson, S. Goel, S. Iyengar, Partisan selective exposure in online news consumption: evidence from the 2016 presidential campaign. *Political Science Research and Methods* 10.1017/psrm.2019.55, 1-17 (2019).

Figure Captions

Figure 1. Description of the Data. The panel data we analyze tracks co-exposure to news sources within and across media platforms (TV, web, and YouTube) for the same set of unique panelists (N ~ 55,000). The web data is provided at the URL level, which allowed us to identify YouTube activity. Using the platform's API, we obtained meta-data for the videos associated with each URL (e.g., publishing channel). This resulted in time-stamped, individual-level exposure measures to news sources within and across media channels (panel A). YouTube offers many more choices of news sources than TV or the web (panel B), but it lags substantially behind in terms of reach, i.e., number of unique panelists visiting these sources (panel C, vertical lines mark the means of the distributions). Audiences on TV and the web exhibit similar levels of concentration around a handful of news sources (panel D); YouTube is much more decentralized. (See materials and methods and the SI for more details on the data and approach to identify news content).

Figure 2. Exposure and Co-Exposure to News Sources. About 80% of all panelists are exposed to at least 1 news channel on TV (panel A). Less than half of the panelists access at least 1 news domain (panel B). A substantially smaller fraction of panelists access news channels on YouTube (less than 5%) but this percentage increases during the observation period (panel C), in contrast with the declining trends observed on TV and the web. (Vertical lines stand for five missing months in the data).

Figure 3. Correlates of News Exposure. The number of unique panelists exposed to news sources over the full period varies significantly across media channels, with only about N ~ 5300 panelists accessing news sources on the three platforms (panel A). The models suggest that age is the most important predictor of news exposure: compared to the base category ('Age: 18-25'), older panelists are more likely to consume news on TV and the web, and less likely to access news on YouTube (panel B). Education is positively correlated with online news consumption, and women are less likely to access news online; they are also clearly less likely to consume news from all three platforms (panel C). (See table SI1 in the SI for the full set of coefficients and alternative specifications, and figure SI8 - table SI2 for analyses with all panelists, including those that do not overlap on the TV and Web datasets).

Figure 4. Correlations on Time Spent on News across Platforms. Panel A shows the distribution of average time spent consuming news across platforms (in minutes, log-transformed). Panel B shows the coefficients that correspond to the effects of time spent consuming news on one platform on time spent on another platform (controlling for all other demographic variables and random effects at the month and panelist levels; confidence intervals are so narrow they are not visible). Time spent consuming news on one platform increases time spent consuming news on another platform (except for YouTube, which does not have a significant impact on TV news time). What these results suggest is that there are no substitution effects; rather, online media amplify already high levels of interest in the news. (Full regression outputs are displayed in Table SI5 in the SI).

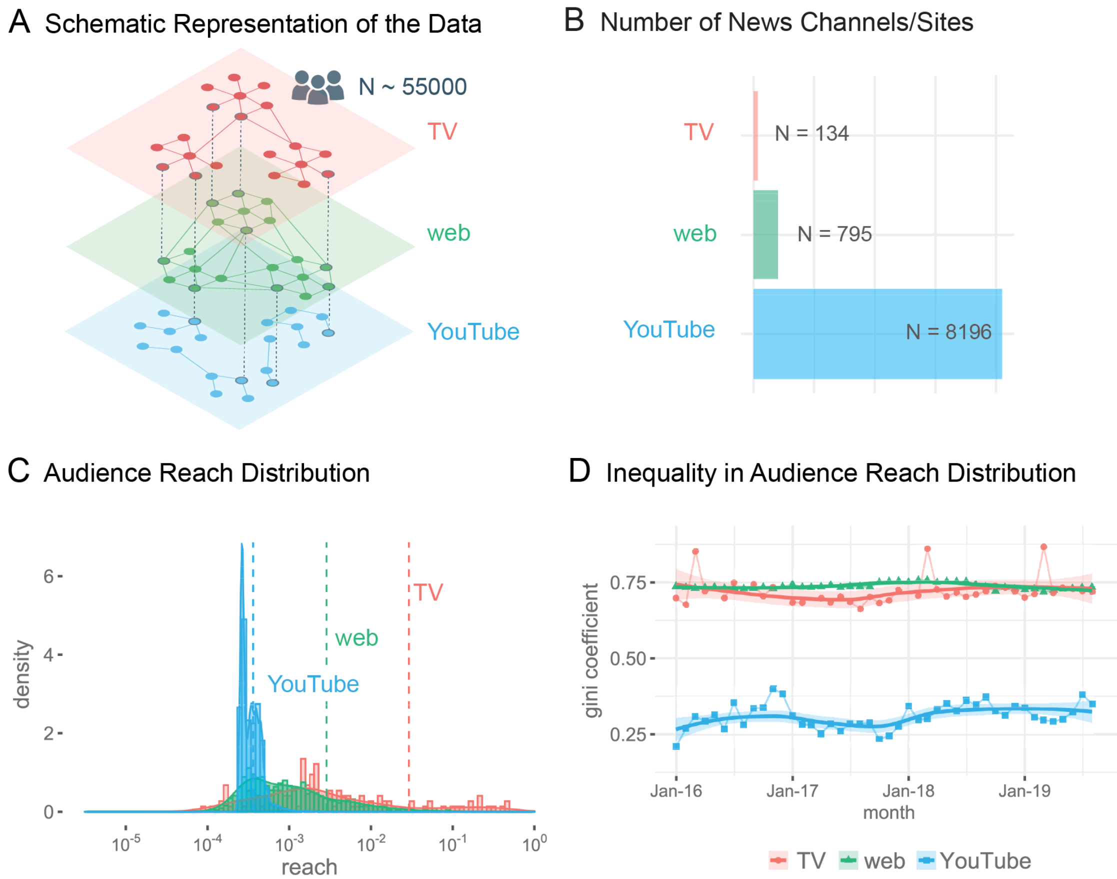

Figure 1. Description of the Data. The panel data we analyze tracks co-exposure to news sources within and across media platforms (TV, web, and YouTube) for the same set of unique panelists (N ~ 55,000). The web data is provided at the URL level, which allowed us to identify YouTube activity. Using the platform's API, we obtained meta-data for the videos associated with each URL (e.g., publishing channel). This resulted in time-stamped, individual-level exposure measures to news sources within and across media channels (panel A). YouTube offers many more choices of news sources than TV or the web (panel B), but it lags substantially behind in terms of reach, i.e., number of unique panelists visiting these sources (panel C, vertical lines mark the means of the distributions). Audiences on TV and the web exhibit similar levels of concentration around a handful of news sources (panel D); YouTube is much more decentralized. (See materials and methods and the SI for more details on the data and approach to identify news content).

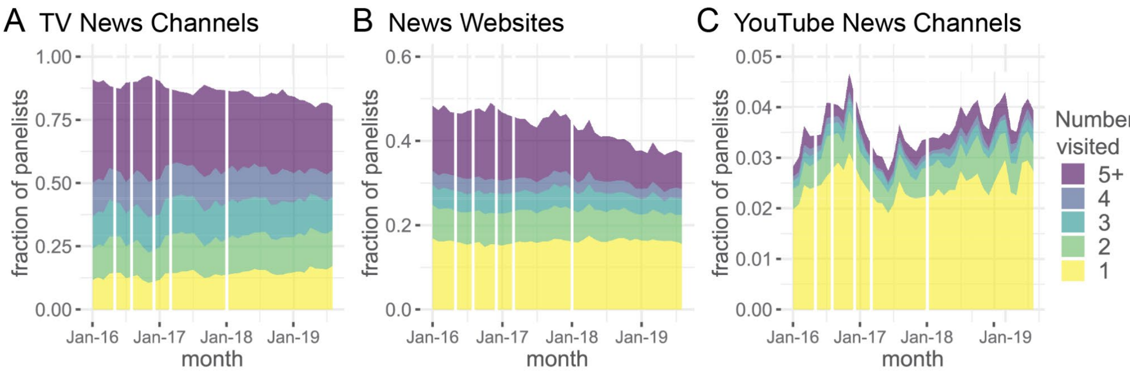

Figure 2. Exposure and Co-Exposure to News Sources. About 80% of all panelists are exposed to at least 1 news channel on TV (panel A). Less than half of the panelists access at least 1 news domain (panel B). A substantially smaller fraction of panelists access news channels on YouTube (less than 5%) but this percentage increases during the observation period (panel C), in contrast with the declining trends observed on TV and the web. (Vertical lines stand for five missing months in the data).

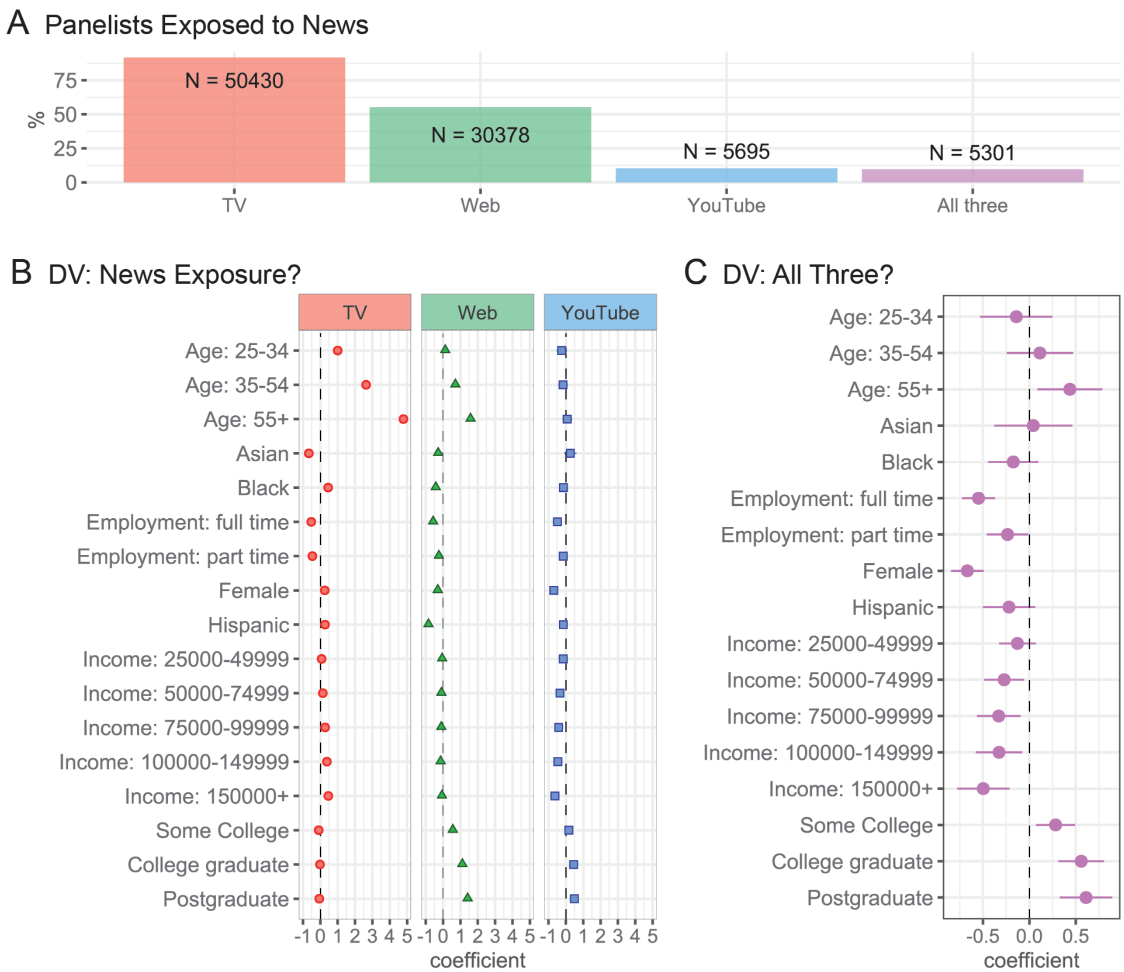

Figure 3. Correlates of News Exposure. The number of unique panelists exposed to news sources over the full period varies significantly across media channels, with only about N ~ 5300 panelists accessing news sources on the three platforms (panel A). The models suggest that age is the most important predictor of news exposure: compared to the base category ('Age: 18-25'), older panelists are more likely to consume news on TV and the web, and less likely to access news on YouTube (panel B). Education is positively correlated with online news consumption, and women are less likely to access news online; they are also clearly less likely to consume news from all three platforms (panel C). (See table SI1 in the SI for the full set of coefficients and alternative specifications, and figure SI8 - table SI2 for analyses with all panelists, including those that do not overlap on the TV and Web datasets).

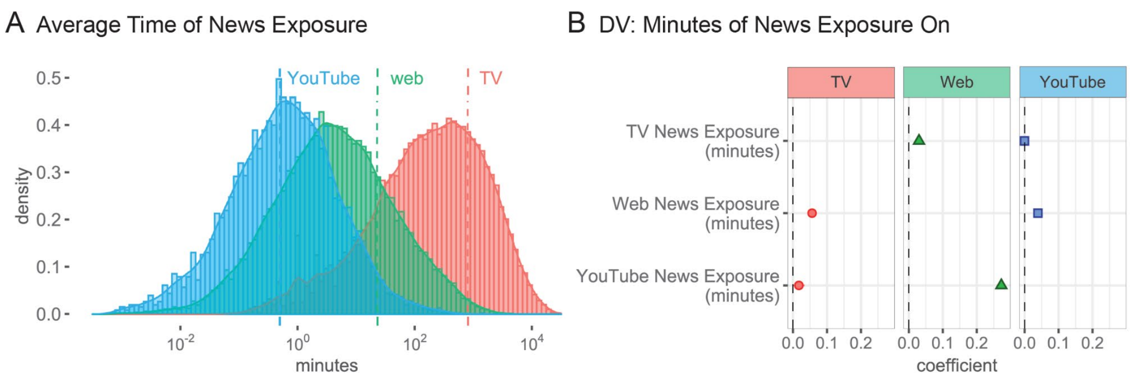

Figure 4. Correlations on Time Spent on News across Platforms. Panel A shows the distribution of average time spent consuming news across platforms (in minutes, log-transformed). Panel B shows the coefficients that correspond to the effects of time spent consuming news on one platform on time spent on another platform (controlling for all other demographic variables and random effects at the month and panelist levels; confidence intervals are so narrow they are not visible). Time spent consuming news on one platform increases time spent consuming news on another platform (except for YouTube, which does not have a significant impact on TV news time). What these results suggest is that there are no substitution effects; rather, online media amplify already high levels of interest in the news. (Full regression outputs are displayed in Table SI5 in the SI).

Supplementary Information

Online Media Boosts Exposure to News but Only for a Small Minority of Hyper-Consumers

Tian Yang[1] and Sandra González-Bailón[1,*]

[1] Annenberg School for Communication, University of Pennsylvania, 19104 Philadelphia, PA, US

* Corresponding author: Sandra González-Bailón, **Email:** sgonzalezbailon@asc.upenn.edu

This PDF file includes:

1. Data

1.1. Nielsen

We analyze TV and web media consumption using two panels provided by Nielsen, both spanning the period January 2016 to August 2019. The TV panel tracks viewing behavior for both live and recorded national TV on a minute-by-minute basis for N ~ 305,000 unique panelists, including information of the name of the program and station being watched. In multi-person households, panelists manually record who is watching at a given time. The web panel tracks desktop web browsing behavior for N ~ 365,000 unique panelists, containing information on the URLs visited and duration of the visit. In multi-person households, panelists manually record who is browsing.

There is an overlap of N ~ 55,000 unique panelists between the TV and web panels, which means we can track their exposure to news across media channels. The analyses reported in the main text are based on this subset of intersecting panelists. This appendix contains additional results based on the two separate panels.

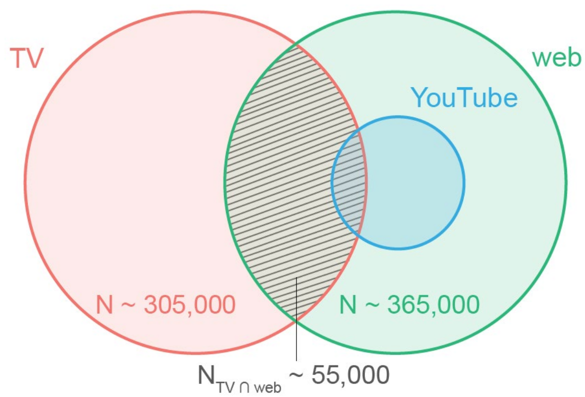

Figure SI1. Schematic representation of the data. We analyzed data contained in two panels: Nielsen's TV panel (*N* ~ 305,000 unique panelists for the period January 2016-August 2019), and Nielsen's Web panel (*N* ~ 365,000 unique panelists same period). A subset of these panelists appears in both data sets (*N* ~ 55,000). The results discussed in the main paper are based on this intersection of panelists. Additional analyses for the full set of TV and web panelists are presented in the sections below.

1.2. YouTube

The web panel contains information on the URLs visited. This allowed us to identify viewing activity within YouTube's domain. Using the platform's API, we queried those URLs to obtain additional information about the videos watched by the panelists, including metadata like the publishing date, channel, category, numbers of likes, views, and comments. The API could provide information for 79% (7,337,679) of 9,288,962 videos extracted from the URLs. The fact that ~20% of video IDs had no return from YouTube's API falls in line with what prior research has reported (1).

1.3. Identification of News Content

TV, web, and YouTube are three media channels with a different set of content providers. We identified news providers on TV using Nielsen's classification of programs, which are nested within stations. About 25% of programs watched by the panelists consuming news on the three platforms are classified as "News" (the percentage for the full set of participants in the TV panel goes down to ~17%). We identified news videos on YouTube using the platform's meta-data and classification tags. In this case, only 5% of all the videos watched by panelists consuming news on the three platforms are in the "News" category (the percentage is ~3% for the full set of participants in the web panel).

To identify news content on the web, we merged the lists of news domains used in five previous studies (2-6). This resulted in a list of N = 813 news domains, of which we could match N = 795 in the Nielsen URL data. Figure 1 in the main text shows the count of news sources for each media channel (panel B) and the reach distribution and gini coefficient of that distribution (panels C and D) for the subset of overlapping panelists. Figure SI2 below shows the same information for the full set of panelists in the two datasets (TV and web). In this larger dataset of web activity, there is a fourfold increase in the number of news YouTube videos watched (even though this amounts to a smaller fraction of all videos watched on the platform, as the treemap in figure SI2 shows).

2. Panel Representativeness

To assess the representativeness of the Nielsen TV and Web panel data, we compared the panelists' demographics with the demographics of a representative sample of the U.S. population (7). In both instances, we used the weights provided by Pew and Nielsen, respectively. Figure SI4 shows the demographic composition of the overlapping panelists, in the unweighted and weighted versions of the data. Figure SI5 shows the comparison for the larger TV and web panels. Overall, the distribution of basic demographics in the Nielsen data is aligned with what we expect in a representative sample. Two departures involve low-income individuals, who are underrepresented, and those in the higher tail of the income distribution, who are slightly over-represented.

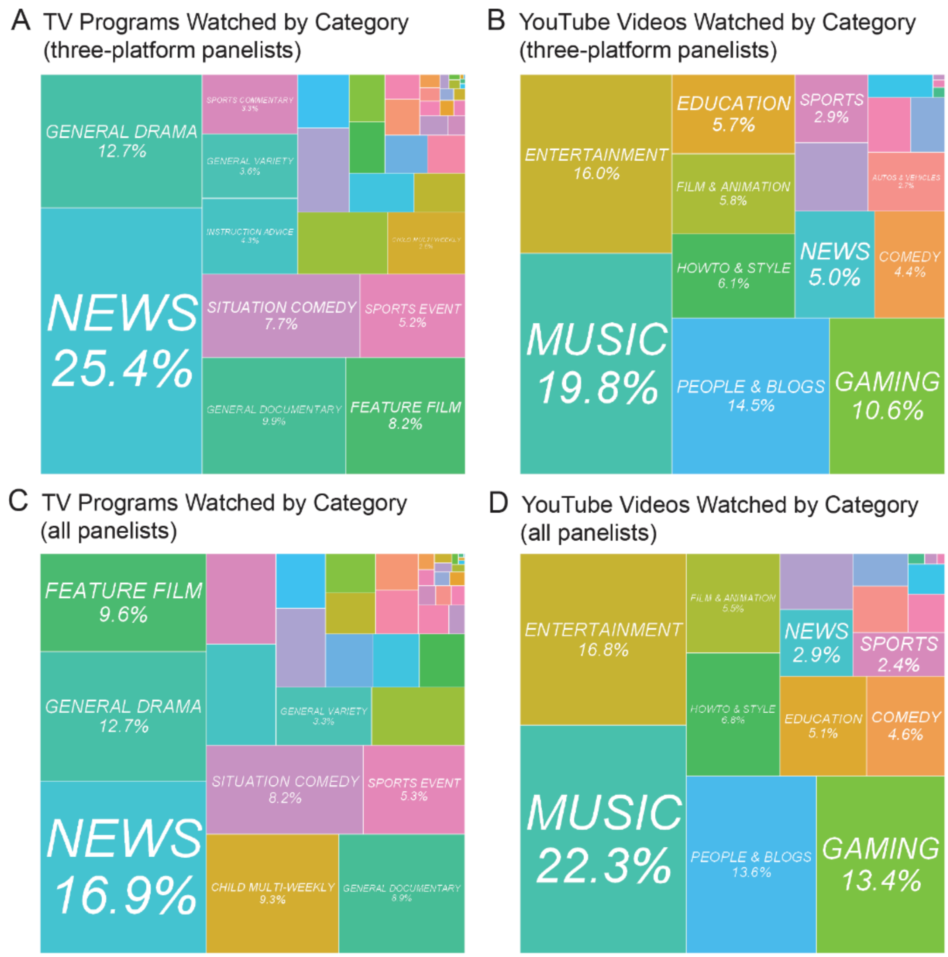

Figure SI2. Relative Prevalence of News Content in Viewing patterns. Panels A and B are based on activity from panelists consuming news on the three platforms (TV, web, and YouTube). Panels C and D are based on activity from all panelists (separate TV and web panels). News content is more common in TV news diets (panels A and C) than on YouTube (panels B and D). The subset of panelists that we can track across platforms (upper row) are more interested in News content than the panelists in the separate TV and Web datasets (lower row).

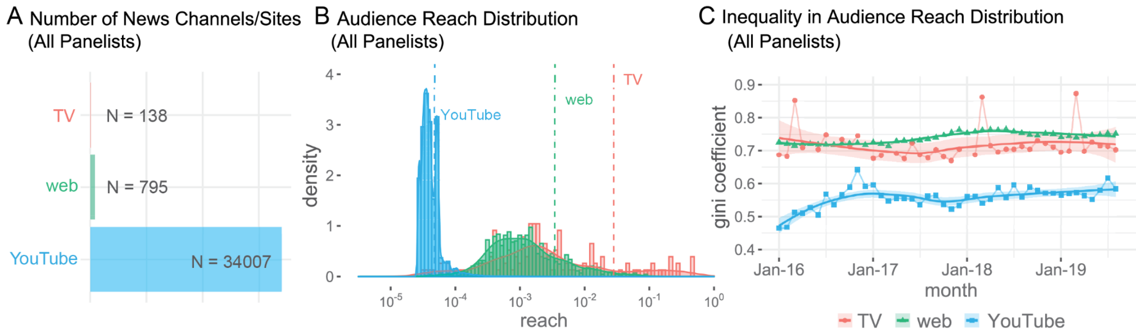

Figure SI3. Reach distribution and inequality index across channels (all panelists). When we consider the activity of all web panelists, the number of YouTube news channels watched increases fourfold (the number of TV news stations/web domains remain similar, panel A). Compared to the set of overlapping panelists, the browsing patterns of news exposure on TV and the web is also now more skewed (B, vertical lines mark the means of the distributions) and concentrated (C). TV is still the most accessed source for news, with YouTube lagging substantially behind.

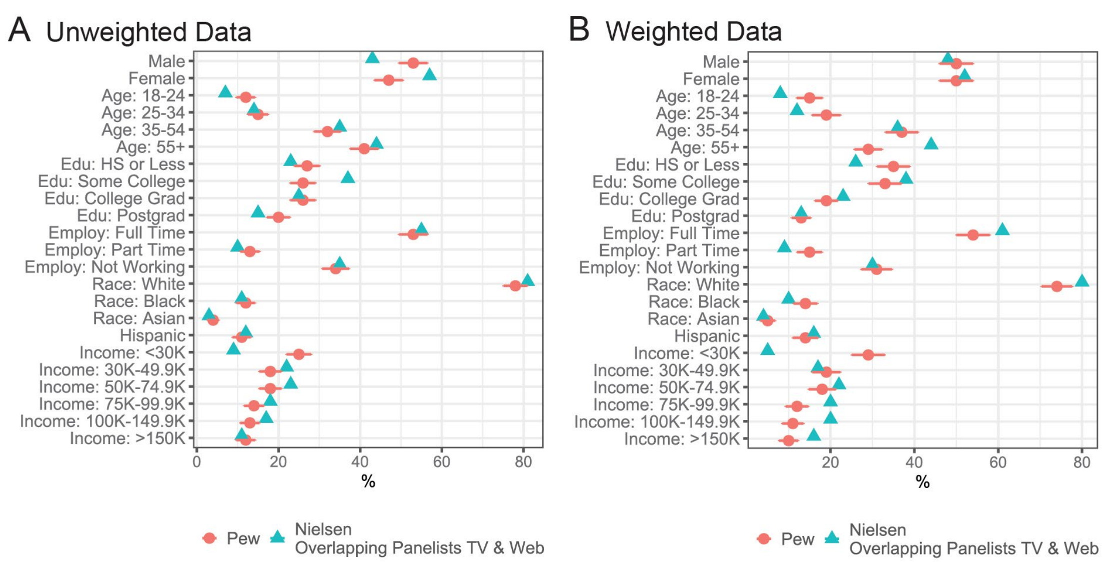

Figure SI4. Comparison of Demographics in the subset of Overlapping Panelists. In this figure, we compare the demographics of the subset of panelists that co-appear in the TV and Web Nielsen panel data. The proportions in panel B are estimated with the weights provided by Nielsen and Pew, respectively. Overall, the distribution of basic demographics in the Nielsen data is not that different, but low-income individuals are underrepresented and those in the higher tail of the income distribution are slightly over-represented.

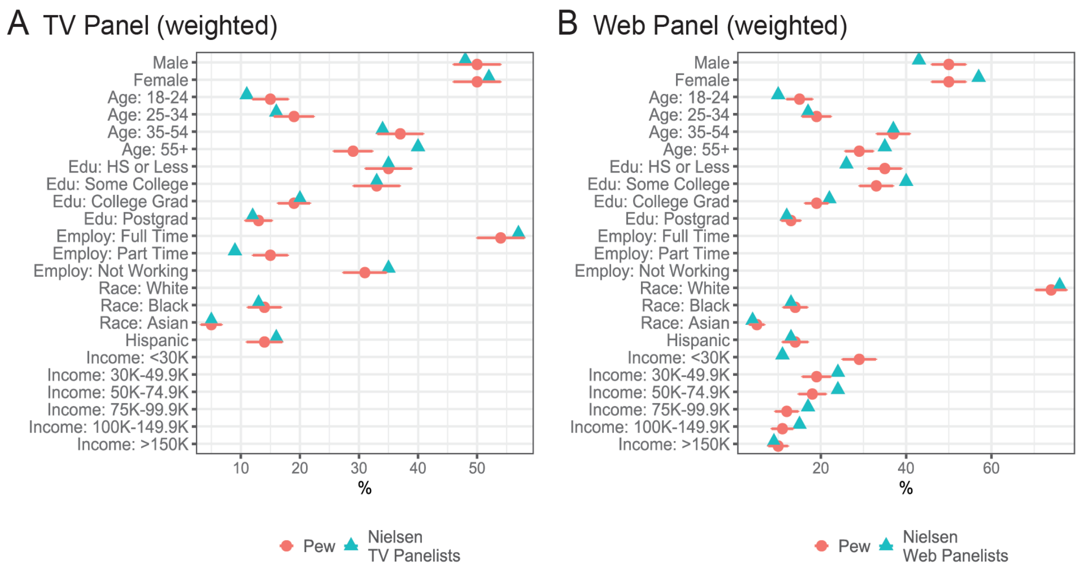

Figure SI5. Comparison of Demographics in TV and Web Panels. We compare the demographics collected for the panelists in the Nielsen data sets (TV and Web) with the demographics of a Pew survey built to be representative of the U.S. population. The proportions estimated with both datasets are weighted (using the weights provided by Nielsen and Pew, respectively).

3. Longevity of Panelists

Most of the panelists across datasets generate trace data for longer than a month (figure SI6). More than half of the panelists on the overlapping dataset that we analyze in the main text generate data for at least 5 months.

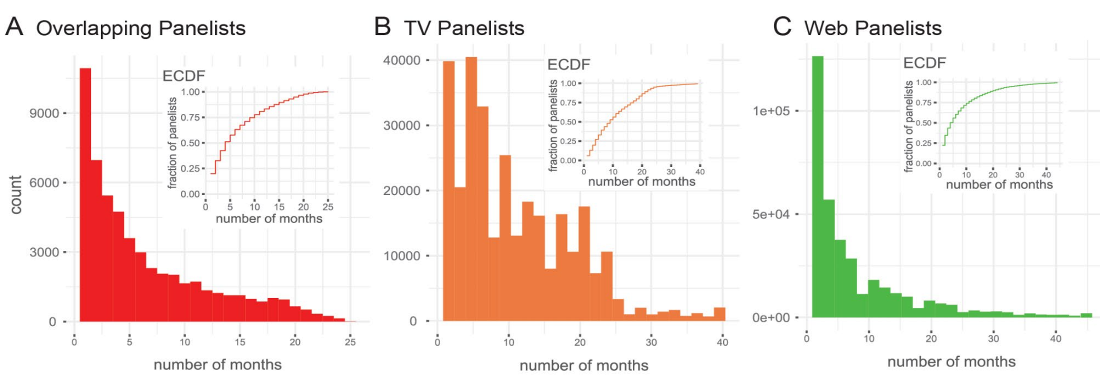

Figure SI6. Longevity of panelists. Half of the panelists in the overlapping dataset (TV and web) generate activity for at least 5 months, about a quarter for longer than 10 months (panel A). The whole set of TV panelists show a longer lifespan (e.g., half of them generate data for at least 10 months, panel B) whereas web panelists have a faster turnover (panel C).

4. Exposure and Co-Exposure to News for All Panelists

The patterns and trends of exposure and co-exposure to news across media channels do not change drastically when analyzing the full TV and web datasets (figure SI7, compared to figure 2 in the main text). As reported in the main text, about 80% of all panelists are exposed to at least 1 news channel on TV (panel A). Slightly more than half of the panelists access at least 1 news domain on the web (panel B), but the decrease over time is sharper (consistent with what prior research has suggested about declining exposure to news from desktop devices, see 2). The fraction of panelists accessing news channels on YouTube is still substantially smaller (less than 6%) but this percentage increases during the observation period (panel C), also in line with what prior research has suggested (1, 8), a trend that stands in contrast with the declining trends observed on the web and TV.

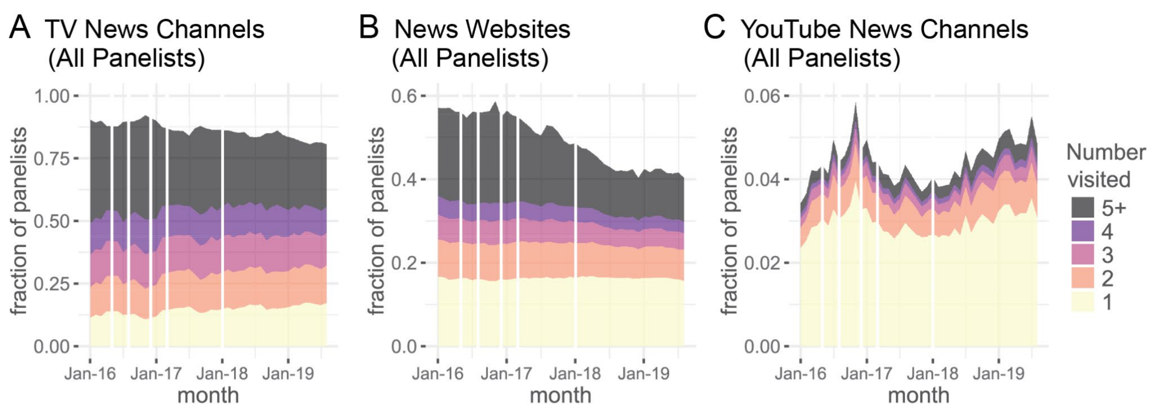

Figure SI7. Exposure and co-exposure to news sources (all panelists). Patterns and trends do not change substantially compared to the trends identified with the overlapping data (figure 2 in the main text). The decrease in web news exposure is now sharper, and the percentage of panelists accessing news videos on YouTube is slightly higher -- but YouTube activity still lags significantly behind web and TV news consumption. (Vertical lines stand for five missing months in the data).

5. Changes in Demographic Composition

5.1. Overlapping Panelists

The demographic attributes for the panelists in the overlapping dataset include age, gender, race, ethnicity (Hispanic), education, income, and employment. Figure SI8 shows changes in the relative prevalence of these different demographic groups for the period of study.

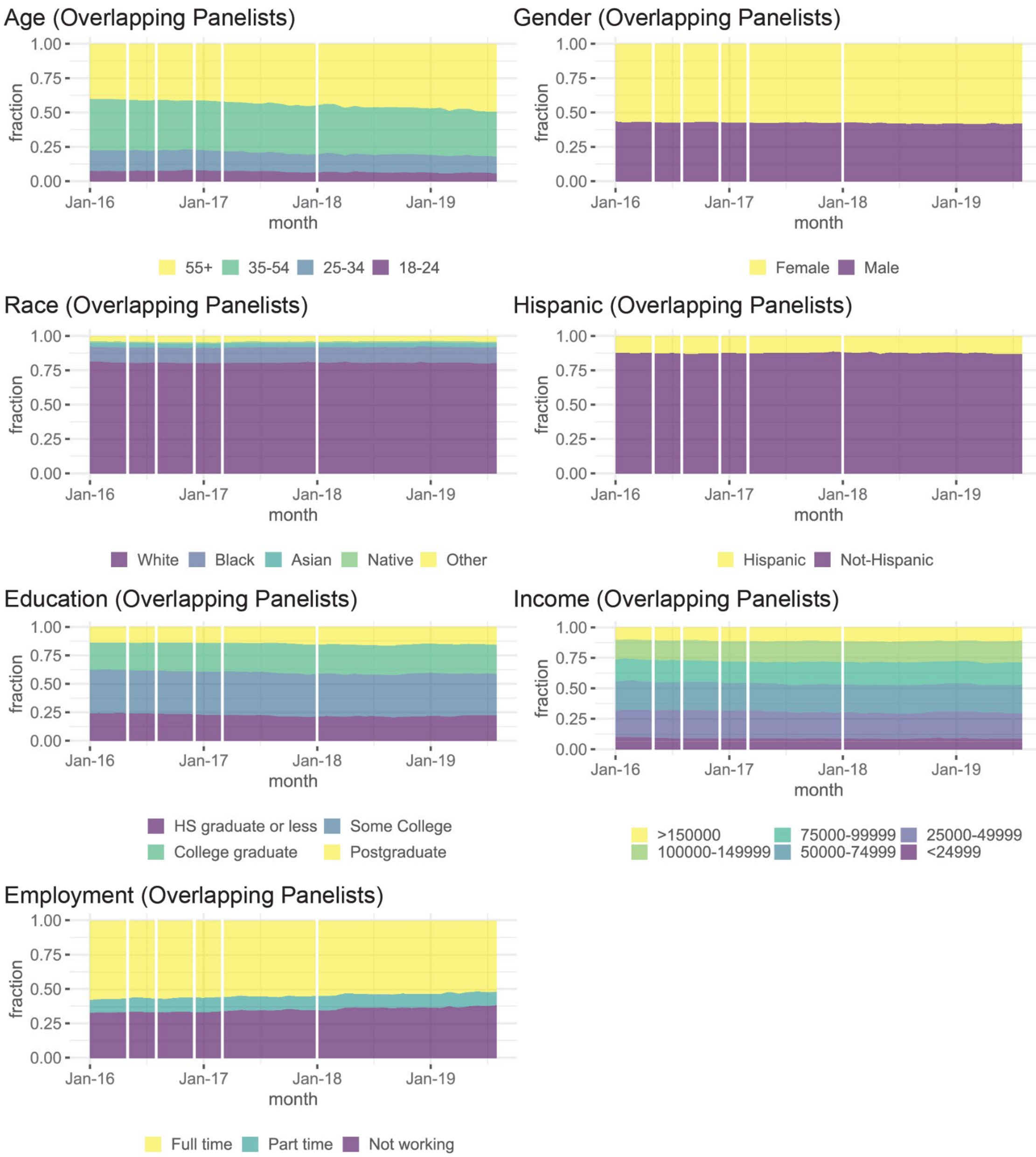

Figure SI8. Demographic Composition of Overlapping Panelists. The relative prevalence of the different demographic groups remains stable for the period under study, although panelists become slightly older towards the end of the temporal window, and the unemployed group increases in size. See also figure SI4 for a comparison of some of these demographics with survey data representative of the U.S. population. (Vertical lines stand for five missing months in the data).

5.2. Non-overlapping Panelists

The demographic attributes available for the panelists in the TV and web datasets vary: the TV dataset does not include information on income, and the web dataset does not include information on employment status. Again, the demographic composition does not change substantially over the period of study. The TV panel is older (Figure SI9, parts 1 and 2).

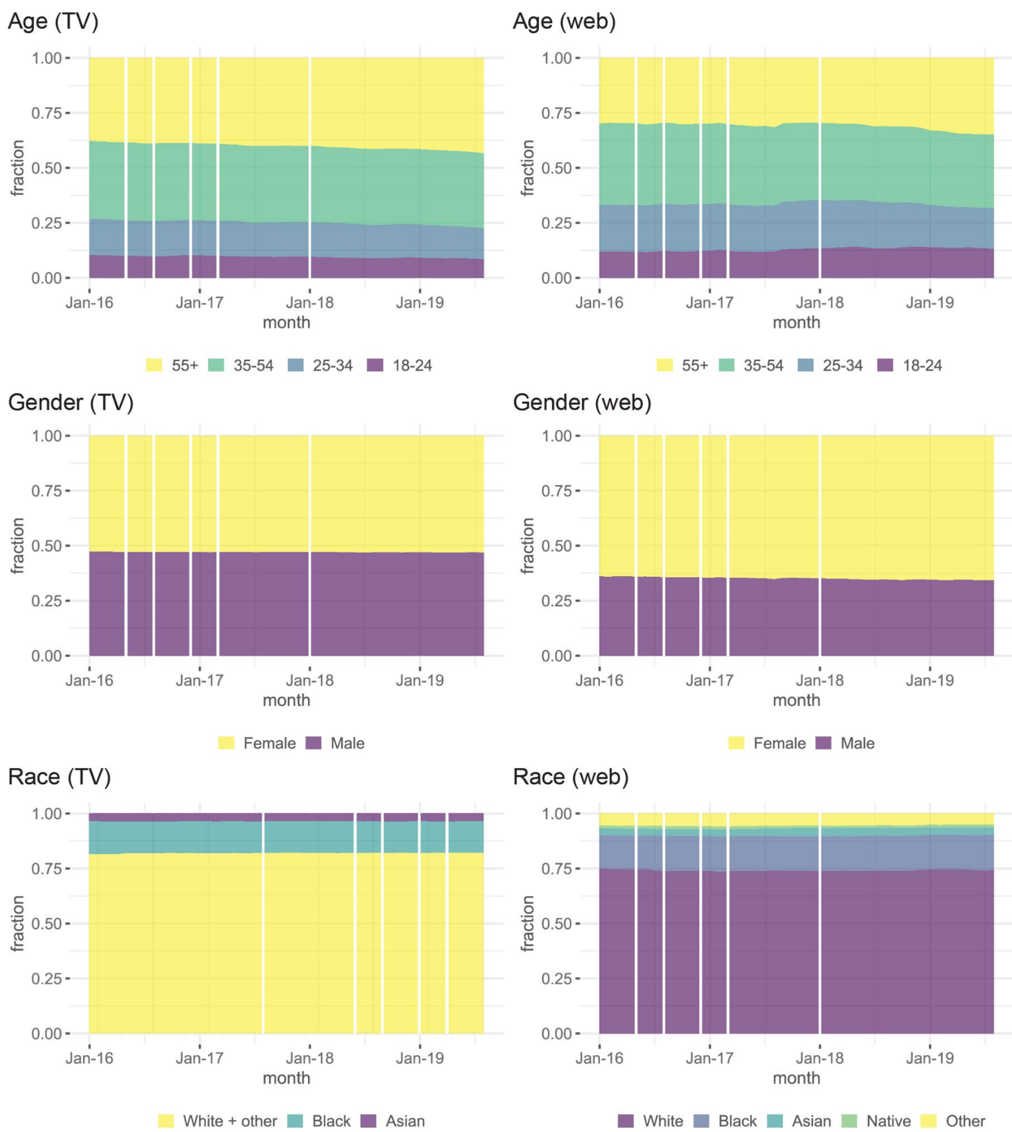

Figure SI9. Demographic Composition of TV and Web Panelists (part 1/2). Demographic groups remain stable for the period. See also figure SI5 for a comparison of these demographics with survey data representative of the U.S. population. (Vertical lines stand for five missing months in the data).

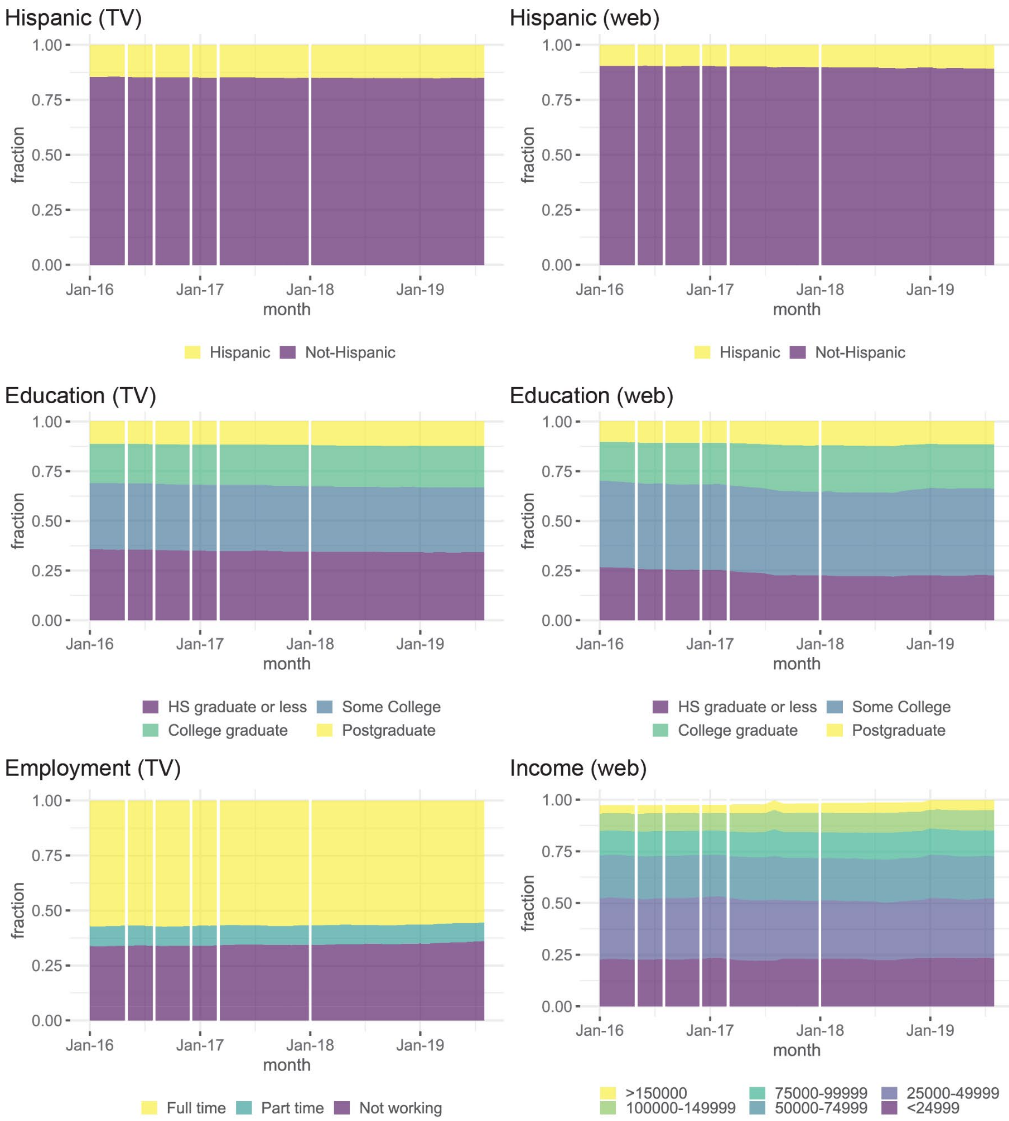

Figure SI9. Demographic Composition of TV and web Panelists (part 2/2). Demographic groups remain stable for the period. See also figure SI5 for a comparison of these demographics with survey data representative of the U.S. population. (Vertical lines stand for five missing months in the data).

6. Co-Exposure across Channels

Table SI1 summarizes the number of panelists in the overlapping dataset creating connections across each pair of media channels or platforms. As the table shows, only a very small fraction of all panelists (0.04%) accessed news on YouTube only. Most consumers of news obtain news from TV first, with the web and YouTube complementing that source of news exposure.

		news on YouTube?	
news on TV?	news on the web?	No	Yes
No	No	3352 (6%)	23 (0.04%)
	Yes	1062 (2%)	147 (0.27%)
Yes	No	21037 (38%)	224 (0.41%)
	Yes	23868 (43%)	5301 (9.6%)

Table SI1. Number of Panelists Co-Exposed Across Channels. Only a very small number of panelists consume news on the Web and YouTube but not on TV (2% and 0.04%, respectively).

7. Regression Models

7.1. Overlapping Panelists

Table SI2 shows the full regression outputs summarized in figure 3 of the main text (models 1-4) and additional models with a different dependent variable (DV), i.e., number of news sources accessed instead of the binary variable measuring exposure to news (models 5-7). These additional models confirm the main effects identified with the models highlighted in the main text.

7.2. Non-overlapping Panelists

Tables SI3 and SI4 show regression outputs for the separate TV and web panels. Figure SI10 summarizes these models.

Correlates of News Exposure

	Dependent variable:						
	TV news?	Web news?	YouTube news?	All three?	TV news channels	News web sites	YouTube news channels
	generalized linear mixed-effects	*generalized linear mixed-effects*	*generalized linear mixed-effects*	*generalized linear mixed-effects*	*linear mixed-effects*	*linear mixed-effects*	*linear mixed-effects*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age: 25-34	1.001***	0.124**	-0.239**	-0.141	0.102***	0.014***	-0.001
	(0.049)	(0.050)	(0.117)	(0.151)	(0.003)	(0.004)	(0.001)
Age: 35-54	2.637***	0.708***	-0.160	0.113	0.238***	0.054***	-0.001
	(0.049)	(0.047)	(0.108)	(0.139)	(0.003)	(0.004)	(0.001)
Age: 55+	4.789***	1.584***	0.081	0.436***	0.372***	0.115***	0.002**
	(0.055)	(0.047)	(0.106)	(0.136)	(0.003)	(0.004)	(0.001)
Female	0.254***	-0.305***	-0.693***	-0.668***	0.005**	-0.040***	-0.010***
	(0.031)	(0.027)	(0.059)	(0.068)	(0.002)	(0.003)	(0.001)
Black	0.449***	-0.422***	-0.139	-0.173*	0.036***	-0.038***	-0.002**
	(0.045)	(0.038)	(0.088)	(0.104)	(0.003)	(0.003)	(0.001)
Asian	-0.663***	-0.286***	0.261*	0.042	-0.033***	-0.018***	0.005***
	(0.070)	(0.063)	(0.136)	(0.163)	(0.004)	(0.005)	(0.001)
Other race	0.083*	-0.134***	0.106	-0.031	0.010***	-0.009***	0.0004
	(0.049)	(0.040)	(0.091)	(0.110)	(0.002)	(0.003)	(0.001)
Some College	-0.094***	0.561***	0.176**	0.283***	-0.003*	0.038***	0.003***
	(0.035)	(0.030)	(0.069)	(0.082)	(0.002)	(0.003)	(0.001)
College graduate	-0.019	1.111***	0.450***	0.559***	0.002	0.086***	0.006***
	(0.043)	(0.036)	(0.081)	(0.095)	(0.003)	(0.003)	(0.001)
Postgraduate	-0.059	1.413***	0.489***	0.611***	0.003	0.121***	0.006***
	(0.053)	(0.044)	(0.096)	(0.110)	(0.003)	(0.004)	(0.001)
Income: 25000-49999	0.075*	-0.054*	-0.149**	-0.129*	0.002	-0.001	-0.002***
	(0.043)	(0.032)	(0.066)	(0.077)	(0.002)	(0.002)	(0.001)
Income: 50000-74999	0.143***	-0.096***	-0.338***	-0.272***	0.004*	-0.007***	-0.004***
	(0.045)	(0.035)	(0.072)	(0.083)	(0.002)	(0.003)	(0.001)
Income: 75000-99999	0.276***	-0.105***	-0.413***	-0.331***	0.007***	-0.009***	-0.005***
	(0.048)	(0.037)	(0.079)	(0.091)	(0.002)	(0.003)	(0.001)
Income: 100000-149999	0.377***	-0.142***	-0.458***	-0.327***	0.011***	-0.009***	-0.007***
	(0.050)	(0.039)	(0.085)	(0.097)	(0.003)	(0.003)	(0.001)
Income: 150000+	0.459***	-0.064	-0.631***	-0.497***	0.011***	-0.008**	-0.008***
	(0.057)	(0.044)	(0.096)	(0.109)	(0.003)	(0.004)	(0.001)
Hispanic	0.269***	-0.842***	-0.144	-0.220**	0.006**	-0.061***	-0.002**
	(0.044)	(0.040)	(0.090)	(0.109)	(0.003)	(0.003)	(0.001)
Employment: part time	-0.456***	-0.242***	-0.147*	-0.235***	-0.041***	-0.021***	-0.002***
	(0.043)	(0.035)	(0.075)	(0.086)	(0.002)	(0.003)	(0.001)
Employment: full time	-0.524***	-0.573***	-0.486***	-0.548***	-0.052***	-0.049***	-0.006***
	(0.035)	(0.027)	(0.060)	(0.069)	(0.002)	(0.002)	(0.001)
Constant	0.572***	-2.033***	-6.932***	-8.010***	0.298***	0.144***	0.022***
	(0.070)	(0.092)	(0.139)	(0.171)	(0.010)	(0.011)	(0.001)
Random Effects							
Number of Panelists	54984	54984	54984	54984	54984	54984	54984
Panelists Standard Deviation	0.216	0.275	0.049	5.472	2.455	2.433	4.996
Number of Months	39	39	39	39	39	39	39
Months Standard Deviation	0.055	0.065	0.002	0.047	0.168	0.447	0
Observations	354,758	354,758	354,758	354,758	354,758	354,758	354,758
Log Likelihood	-95,945.450	-170,911.400	-40,764.140	-32,942.950	77,298.750	6,451.814	430,202.700
Akaike Inf. Crit.	191,932.900	341,864.900	81,570.280	65,927.900	-154,553.500	-12,859.630	-860,361.400
Bayesian Inf. Crit.	192,159.300	342,091.200	81,796.640	66,154.260	-154,316.400	-12,622.490	-860,124.200
Note:							*p<0.1; **p<0.05; ***p<0.01

Table SI2. Regression Outputs for Exposure to News. Models 1-4 are the basis for figure 3 in the main text. Models 5-7 use a continuous measure as DV (i.e., number of channels or websites).

Correlates of News Exposure (TV)

	Dependent variable:	
	TV news?	TV news channels
	generalized linear mixed-effects	*linear mixed-effects*
	(1)	(2)
Age: 25-34	0.620***	0.083***
	(0.014)	(0.001)
Age: 35-54	1.813***	0.214***
	(0.014)	(0.001)
Age: 55+	3.290***	0.338***
	(0.016)	(0.001)
Female	0.133***	0.005***
	(0.010)	(0.001)
Black	0.091***	0.017***
	(0.014)	(0.001)
Asian	-0.757***	-0.065***
	(0.023)	(0.002)
Some College	-0.018*	0.001
	(0.010)	(0.001)
College graduate	0.086***	0.007***
	(0.013)	(0.001)
Postgraduate	0.118***	0.010***
	(0.017)	(0.001)
Hispanic	0.036***	-0.008***
	(0.013)	(0.001)
Employment: part time	-0.219***	-0.030***
	(0.013)	(0.001)
Employment: full time	-0.312***	-0.044***
	(0.011)	(0.001)
Constant	0.705***	0.323***
	(0.089)	(0.007)
Random Effects		
Number of Panelists	304960	304960
Panelists Standard Deviation	2.14	0.213
Number of Months	39	39
Months Standard Deviation	0.55	0.041
Observations	3,320,294	3,320,294
Log Likelihood	-939,301.900	676,052.000
Akaike Inf. Crit.	1,878,634.000	-1,352,072.000
Bayesian Inf. Crit.	1,878,829.000	-1,351,864.000
Note:		*p<0.1; **p<0.05; ***p<0.01

Table SI3. Regression Outputs for Exposure to News (TV Panel). Results are qualitatively similar to those found with overlapping panelists.

Correlates of News Exposure

	Dependent variable:			
	Web news?	YouTube news?	News web sites	YouTube news channels
	generalized linear mixed-effects	*generalized linear mixed-effects*	*linear mixed-effects*	*linear mixed-effects*
	(1)	(2)	(3)	(4)
Age: 25-34	-0.134***	-0.126***	-0.006***	-0.003***
	(0.012)	(0.018)	(0.001)	(0.0003)
Age: 35-54	0.140***	-0.100***	0.026***	-0.003***
	(0.013)	(0.017)	(0.001)	(0.0003)
Age: 55+	0.674***	0.144***	0.075***	-0.001*
	(0.014)	(0.017)	(0.002)	(0.0004)
Female	0.094***	-0.435***	0.003**	-0.010***
	(0.009)	(0.011)	(0.001)	(0.0002)
Black	-0.314***	-0.023	-0.057***	0.001**
	(0.013)	(0.015)	(0.002)	(0.0003)
Asian	-0.007	0.303***	-0.001	0.009***
	(0.024)	(0.027)	(0.003)	(0.001)
Other race	-0.150***	0.105***	-0.021***	0.002***
	(0.018)	(0.022)	(0.002)	(0.0004)
Some College	0.522***	0.162***	0.073***	0.004***
	(0.011)	(0.014)	(0.001)	(0.0003)
College graduate	0.912***	0.266***	0.134***	0.007***
	(0.014)	(0.017)	(0.002)	(0.0003)
Postgraduate	1.102***	0.311***	0.164***	0.007***
	(0.017)	(0.021)	(0.002)	(0.0004)
Income: 25000-49999	-0.097***	-0.102***	-0.009***	-0.003***
	(0.011)	(0.014)	(0.001)	(0.0003)
Income: 50000-74999	-0.147***	-0.178***	-0.016***	-0.006***
	(0.013)	(0.016)	(0.001)	(0.0003)
Income: 75000-99999	-0.196***	-0.211***	-0.022***	-0.007***
	(0.015)	(0.019)	(0.002)	(0.0004)
Income: 100000-149999	-0.220***	-0.223***	-0.023***	-0.008***
	(0.016)	(0.022)	(0.002)	(0.0004)
Income: 150000+	-0.194***	-0.253***	-0.024***	-0.009***
	(0.021)	(0.029)	(0.002)	(0.001)
Hispanic	-0.390***	0.005	-0.053***	0.0002
	(0.015)	(0.019)	(0.002)	(0.0004)
Constant	-0.952***	-2.344***	0.184***	0.024***
	(0.114)	(0.026)	(0.023)	(0.001)
Random Effects				
Number of Panelists	358672	358672	220298	358672
Panelists Standard Deviation	2.279	0.311	1.446	0.057
Number of Months	44	44	44	44
Months Standard Deviation	0.749	0.155	0.113	0.002
Observations	2,860,072	1,108,946	2,860,072	2,860,072
Log Likelihood	-1,446,357.000	-341,239.500	-299,986.700	3,160,536.000
Akaike Inf. Crit.	2,892,752.000	682,517.000	600,013.500	-6,321,032.000
Bayesian Inf. Crit.	2,892,997.000	682,743.500	600,270.800	-6,320,775.000
Note:				*p<0.1; **p<0.05; ***p<0.01

Table SI4. Regression Outputs for Exposure to News (Web Panel). Results are qualitatively similar to those found with overlapping panelists.

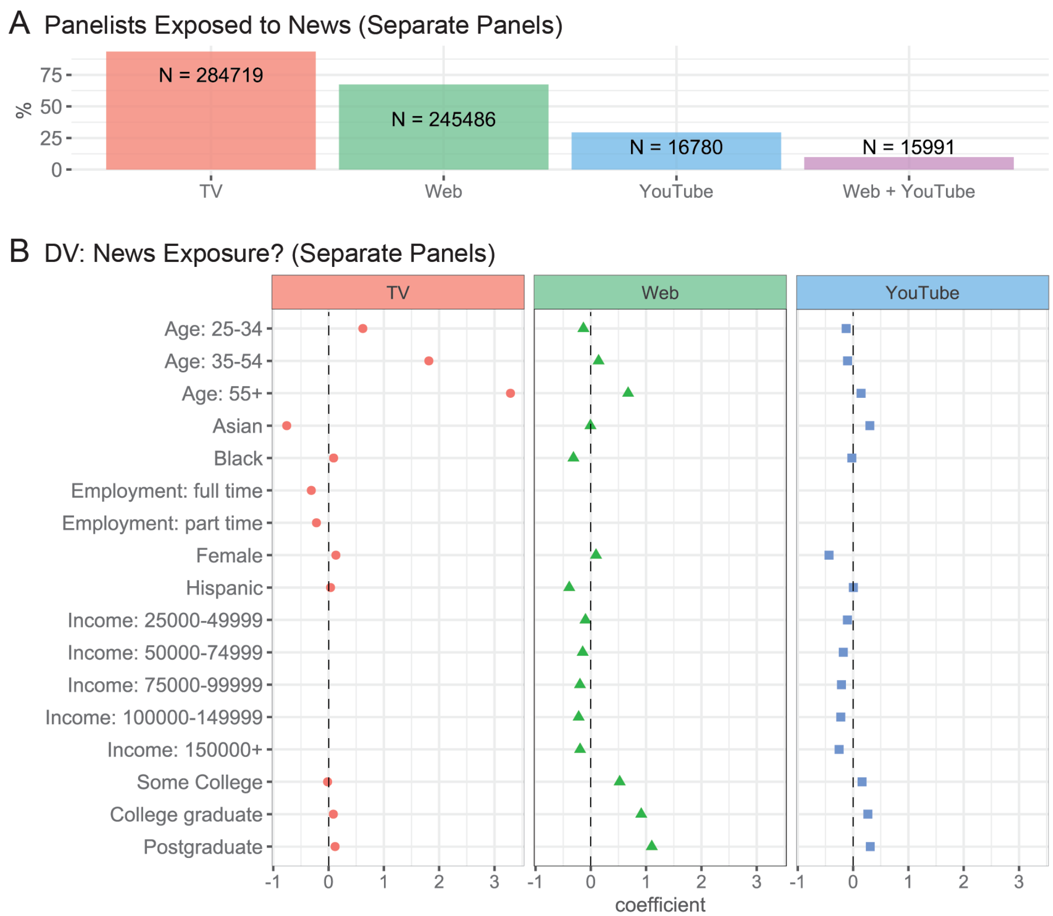

Figure SI10. Correlates of News Exposure for TV and Web Panels. Results are qualitatively similar to those found with overlapping panelists.

8. Substitution vs Amplification Test

To determine if news consumption on one platform has a substitution effect on consumption on other platforms, we calculated the time spent on news content on each of the three platforms and looked at their associations. Given the skewness of the measure (see Figure 4A in the main text), we used the logarithmic transformation of this variable. We used linear mixed-effects models including news time on the other two channels (and the same demographic variables used above as controls and including panelist ID and month as random effects). We used two measures of news exposure: a continuous measure of time spent (models 1-3 in Table SI5) and a dummy variable indicating whether panelists accessed news on each platform (models 4-6 in Table SI5).

A negative coefficient in the association of news consumption across platforms would support the substitution effect: it would mean that a longer time spent on news content on one platform is associated with decreasing news exposure on the other platform; this, in turn, would suggest that the different platforms compete for news audiences – e.g., time spent on TV is replaced by time spent on the web and/or YouTube. Contrary to this expectation, our models return positive coefficients, thus supporting the alternative effect of amplification: longer time spent on news on one platform predicts more time consuming news on the other two platforms. In other words, our findings show that different media platforms complement each other and expand overall news consumption for those interested in the news.

In addition to the mixed-effects models, we run additional robustness checks by fitting group fixed-effects models that use month and panelist ID as the grouping variables. These models offer an alternative estimation approach to address the concern of time shocks and time-invariant confounding factors. As Table SI6 shows, they produce largely similar results.

	Dependent variable:					
	TV news	Web news	YouTube news	TV news	Web news	YouTube news
	(1)	(2)	(3)	(4)	(5)	(6)
Web news (minutes)	0.056^{***}		0.039^{***}			
	(0.002)		(0.001)			
TV news (minutes)		0.030^{***}	-0.0003			
		(0.001)	(0.0004)			
YouTube news (minutes)	0.017^{***}	0.272^{***}				
	(0.006)	(0.005)				
Web news (dummy)				0.045^{***}		0.028^{***}
				(0.003)		(0.001)
TV news (dummy)					0.022^{***}	-0.00003
					(0.003)	(0.001)
YouTube news (dummy)				0.028^{***}	0.285^{***}	
				(0.006)	(0.004)	
Age: 25-34	0.340^{***}	0.024^{***}	-0.002	0.342^{***}	0.033^{***}	-0.002
	(0.013)	(0.008)	(0.002)	(0.013)	(0.008)	(0.002)
Age: 35-54	0.872^{***}	0.108^{***}	-0.007^{***}	0.876^{***}	0.131^{***}	-0.004^{**}
	(0.013)	(0.008)	(0.002)	(0.013)	(0.008)	(0.002)
Age: 55+	1.411^{***}	0.253^{***}	-0.010^{***}	1.420^{***}	0.290^{***}	-0.002
	(0.013)	(0.009)	(0.002)	(0.013)	(0.008)	(0.002)
Female	0.067^{***}	-0.082^{***}	-0.016^{***}	0.064^{***}	-0.081^{***}	-0.019^{***}
	(0.008)	(0.005)	(0.001)	(0.008)	(0.005)	(0.001)
Black	0.127^{***}	-0.065^{***}	0.001	0.125^{***}	-0.061^{***}	-0.0003
	(0.010)	(0.007)	(0.002)	(0.010)	(0.007)	(0.002)
Asian	-0.160^{***}	-0.034^{***}	0.015^{***}	-0.160^{***}	-0.037^{***}	0.014^{***}
	(0.016)	(0.011)	(0.003)	(0.016)	(0.011)	(0.003)
Other race	0.030^{***}	-0.025^{***}	0.004^{**}	0.029^{***}	-0.024^{***}	0.003^{*}
	(0.008)	(0.006)	(0.002)	(0.008)	(0.006)	(0.002)
Some College	-0.006	0.074^{***}	0.004^{***}	-0.004	0.075^{***}	0.006^{***}
	(0.007)	(0.005)	(0.001)	(0.007)	(0.005)	(0.001)
College graduate	0.004	0.159^{***}	0.004^{**}	0.008	0.159^{***}	0.008^{***}
	(0.010)	(0.006)	(0.001)	(0.010)	(0.006)	(0.002)
Postgraduate	0.031^{**}	0.228^{***}	0.0003	0.037^{***}	0.228^{***}	0.007^{***}
	(0.012)	(0.008)	(0.002)	(0.012)	(0.008)	(0.002)
Income: 25000-49999	0.015^{**}	-0.002	-0.009^{***}	0.015^{**}	-0.002	-0.009^{***}
	(0.007)	(0.005)	(0.001)	(0.007)	(0.005)	(0.002)
Income: 50000-74999	0.025^{***}	-0.015^{***}	-0.013^{***}	0.025^{***}	-0.014^{***}	-0.013^{***}
	(0.007)	(0.005)	(0.002)	(0.008)	(0.005)	(0.002)
Income: 75000-99999	0.024^{***}	-0.015^{***}	-0.014^{***}	0.024^{***}	-0.015^{**}	-0.015^{***}
	(0.008)	(0.006)	(0.002)	(0.008)	(0.006)	(0.002)
Income: 100000-149999	0.040^{***}	-0.023^{***}	-0.016^{***}	0.039^{***}	-0.023^{***}	-0.017^{***}
	(0.009)	(0.006)	(0.002)	(0.009)	(0.006)	(0.002)
Income: 150000+	0.040^{***}	-0.015^{**}	-0.020^{***}	0.040^{***}	-0.013^{*}	-0.020^{***}
	(0.010)	(0.007)	(0.002)	(0.010)	(0.007)	(0.002)
Hispanic	-0.025^{**}	-0.114^{***}	0.003^{*}	-0.027^{**}	-0.114^{***}	0.0004
	(0.011)	(0.007)	(0.002)	(0.011)	(0.007)	(0.002)
Employment: part time	-0.141^{***}	-0.040^{***}	-0.002	-0.142^{***}	-0.043^{***}	-0.002
	(0.008)	(0.006)	(0.002)	(0.008)	(0.006)	(0.002)
Employment: full time	-0.191^{***}	-0.094^{***}	-0.008^{***}	-0.194^{***}	-0.099^{***}	-0.011^{***}
	(0.007)	(0.005)	(0.001)	(0.007)	(0.005)	(0.001)
Constant	0.971^{***}	0.219^{***}	0.037^{***}	0.973^{***}	0.232^{***}	0.040^{***}
	(0.030)	(0.020)	(0.002)	(0.031)	(0.020)	(0.003)

Table SI5. Regression outputs to test substitution vs. amplification hypothesis (part 1/2)

Random Effects						
Number of Panelists	54984	54984	54984	54984	54984	54984
Panelists Standard Deviation	0.897	0.549	0.096	0.897	0.547	0.099
Number of Months	39	39	39	39	39	39
Months Standard Deviation	0.165	0.106	0.004	0.168	0.109	0.003
Observations	354,758	354,758	354,758	354,758	354,758	354,758
Log Likelihood	-353,032.200	-242,617.200	143,771.700	-353,198.100	-242,505.000	141,874.900
Akaike Inf. Crit.	706,112.400	485,282.500	-287,495.500	706,444.200	485,058.000	-283,701.800
Bayesian Inf. Crit.	706,371.100	485,541.200	-287,236.800	706,702.900	485,316.700	-283,443.100
Note:						*p<0.1; **p<0.05; ***p<0.01

Table SI5. Regression outputs to test substitution vs. amplification hypothesis (part 2/2). This table summarizes the estimates produced by linear mixed-effects models with time spent on one platform as DV and time spent on the other platforms as IV (models 1-3); models 4-6 use as IV a binary measure of news exposure. Time spent is measured in minutes and is log-transformed.

	Dependent variable:		
	Web news time (logged)	TV news time (logged)	YouTube news time (logged)
	(1)	(2)	(3)
TV news time (logged)	0.029^{***}		0.002^{***}
	(0.002)		(0.001)
Web news time (logged)		0.050^{***}	0.029^{***}
		(0.003)	(0.001)
YouTube news time (logged)	0.212^{***}	0.023^{***}	
	(0.009)	(0.007)	
Observations	354,869	354,869	354,869
R^2	0.773	0.834	0.459
Adjusted R^2	0.732	0.804	0.360
Residual Std. Error	0.404	0.532	0.150
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table SI6. Regression outputs to test substitution vs. amplification hypothesis (group fixed-effects models). This table summarizes the estimates produced by linear fixed-effects models with time spent on a given platform as DV and time spent on the other two platforms as IV. Time spent is measured in minutes and is log-transformed. We use month and panelist ID as the fixed-effects to address the concerns of time-shocks and time-invariant confounding factors.

References

1. H. Hosseinmardi *et al.*, Examining the consumption of radical content on YouTube. *Proceedings of the National Academy of Sciences* **118**, e2101967118 (2021).
2. T. Yang, S. Majó-Vázquez, R. K. Nielsen, S. González-Bailón, Exposure to news grows less fragmented with an increase in mobile access. *Proceedings of the National Academy of Sciences* **117**, 28678-28683 (2020).
3. C. Budak, S. Goel, J. M. Rao, Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis. *Public Opinion Quarterly* **80**, 250-271 (2016).
4. E. Bakshy, S. Messing, L. A. Adamic, Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130-1132 (2015).
5. N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, Fake news on Twitter during the 2016 U.S. presidential election. *Science* **363**, 374-378 (2019).
6. E. Peterson, S. Goel, S. Iyengar, Partisan selective exposure in online news consumption: evidence from the 2016 presidential campaign. *Political Science Research and Methods* 10.1017/psrm.2019.55, 1-17 (2019).
7. Pew Research Center, Social Media Update 2016. (2016).
8. K. Munger, J. Phillips, Right-Wing YouTube: A Supply and Demand Perspective. *The International Journal of Press/Politics* **0**, 1940161220964767 (2021).