doi:xx.xxxx/xxxxx

Post Post-Broadcast Democracy? News Exposure in the Age of Online Intermediaries

SEBASTIAN STIER *GESIS – Leibniz Institute for the Social Sciences*
FRANK MANGOLD *University of Hohenheim*
MICHAEL SCHARKOW *Johannes Gutenberg University Mainz*
JOHANNES BREUER *GESIS – Leibniz Institute for the Social Sciences and Center for Advanced Internet Studies (CAIS)*

Online intermediaries such as social network sites or search engines are playing an increasingly central role in democracy by acting as mediators between information producers and citizens. Academic and public commentators have raised persistent concerns that algorithmic recommender systems would negatively affect the provision of political information by tailoring content to the predispositions and entertainment preferences of users. At the same time, recent research indicates that intermediaries foster exposure to news that people would not use as part of their regular media diets. This study investigates these unresolved questions by combining the web browsing histories and survey responses of more than 7,000 participants from six major democracies. The analysis shows that despite generally low levels of news use, using online intermediaries fosters exposure to non-political and political news across countries and personal characteristics. The findings have implications for scholarly and public debates on the challenges that high-choice digital media environments pose to democracy.

Word Count:

Sebastian Stier, Senior Researcher, GESIS, Department Computational Social Science, Cologne, Germany, sebastian.stier@gesis.org

Frank Mangold, Postdoctoral researcher, Department of Media Research and Media Use, Stuttgart, Germany, frank.mangold@uni-hohenheim.de

Michael Scharkow, Professor, Johannes Gutenberg University, Mainz, Germany, scharkow@uni-mainz.de

Johannes Breuer, Senior Researcher, GESIS, Department Survey Data Curation, Cologne, Germany, and Center for Advanced Internet Studies (CAIS), Team Research Data & Methods, Bochum, Germany, johannes.breuer@gesis.org

This is a manuscript submitted for review.

Liberal democracy is facing threats around the globe. In many academic and public accounts of the current perils of democracy, the high-choice media environment brought about by the Internet is prominently mentioned among the culprits. While mass broadcasters and newspapers traditionally guaranteed common exposure to public affairs information, the ever-expanding variety of digital entertainment and niche content has provided citizens with unprecedented autonomy in their information choices (Van Aelst et al. 2017). The algorithmically and socially driven recommender systems of online intermediaries such as search engines (e.g., Google), social network sites (e.g., Facebook) or online portals (e.g., MSN) potentially even amplify the scope of audience self-selection. Hence, intermediaries seem to be a natural continuation of the "post-broadcast democracy" of the 1990s and 2000s that was characterized by selective avoidance of public affairs coverage in favor of entertainment content (Prior 2007). Yet not least due to the limited accuracy of self-reported media use (Prior 2009), studying the dynamics of news exposure in digital media environments remains a major challenge.

Against this backdrop, we ask: do online intermediaries indeed drive away citizens from news? Or do they actually foster – non-political and political – news exposure? To this end, we combine cross-national data on observed web browsing behavior with the complementary advantages of surveys and content analysis (Stier et al. 2020). The within-person regression models show that even across countries and personal characteristics, intermediaries foster rather than restrict news exposure. These findings are relevant for the fields of political communication and public opinion but also have broader implications for media policy and democracy.

ONLINE INTERMEDIARIES AND EXPOSURE TO NEWS

The emergence of intermediaries as central brokers between information producers and citizens in digital media environments marks a significant evolutionary stage in media development (Van Aelst et al. 2017). Recent scholarship has gravitated between two theoretical poles when it comes to intermediaries' effects on news use.

One school has argued that intermediaries exacerbate the trend towards individualized media diets and selective avoidance of public affairs coverage that has accompanied the expansion of television

broadcasting (Prior 2007). It was expected that users maintain their regular selection patterns when allocating their attention to social media content (Kümpel 2020; Van Aelst et al. 2017). In fact, these predispositions should be amplified due to curated content flows on intermediaries, such that search and recommender systems may show more news to politically interested citizens in the first place (Thorson 2020). Proponents of concepts such as "echo chambers" or "filter bubbles" even anticipate that the most engaged citizens end up being trapped in enclaves of like-minded speech (Sunstein 2009). However, in light of a growing body of evidence showing that ideological self-selection is not widespread (Barberá 2015; Flaxman et al. 2016; Fletcher et al. 2021; Guess 2021; Tucker et al. 2018), the more severe consequence would be that people tune out or get tuned out of news altogether (Fletcher et al. 2021; Skovsgaard and Andersen 2020; Thorson 2020; Van Aelst et al. 2017).

Another stream of research contends that intermediaries counteract selective exposure by guiding people to news which they would not use as part of their regular daily online routines and content choices (Fletcher and Nielsen 2018a; Möller et al. 2020; Scharkow et al. 2020). Although people may go online for using other types of contents or services, they may encounter news on search engines (Fletcher and Nielsen 2018b), while logging into email accounts on online portals or on social network sites, where popularity cues alongside traditional source cues determine the perceived relevance of contents (Anspach 2017; Bode 2016). As online social network ties stem from heterogeneous domains of life (e.g., work, family or childhood friends), they have the potential to foster exposure to political content beyond citizens' own interests and viewpoints (Barberá 2015). While mere encounters of news previews and headlines on intermediary platforms are arguably of restricted democratic value, if citizens are led to news outlets and engage with the full story, this bears potential for gains in political knowledge traditionally attributed to newspaper use (Lee and Kim 2017). Less obvious to users, the inherently social nature of digital media also extends to search engines and portals whose algorithms showcase trending topics from popular news sources that have generated user attention beyond specialized niche audiences.

Using a novel cross-country dataset, this study addresses several gaps in the literature. While some research has advanced that intermediaries amplify inequalities in news use, at least among the politically least engaged (Kümpel 2020; Skovsgaard and Andersen 2020), other studies have identified

reverse patterns (Bode 2016; Fletcher and Nielsen 2018a). Methodologically, survey studies have suffered from the well-known limitations of self-reports of media use (Prior 2009) that are particularly severe when people arrive at online news via intermediaries (Kalogeropoulos et al. 2019); experimental studies simulating online environments have only limited external validity; and much tracking research has rested on highly aggregated behavioral data, making it impossible to determine individual news diets (Stier et al. 2020). Apart from some cross-national research (Barberá 2015; Fletcher and Nielsen 2018a), prior findings were confined to isolated cases, most often the United States. Finally, in light of normative theories of a democratic public sphere, it is crucial to assess the externalities of online intermediaries across different strata of the population and whether positive effects extend to *political* content.

RESEARCH DESIGN

To mitigate the deficits of self-reports of media exposure, this study relies on recordings of individual web browsing histories. The "web tracking" data were collected in six Western democracies: France, Germany, Italy, Spain, the United Kingdom and the United States. By including Democratic Corporatist, Polarized Pluralist and Liberal media systems, the sample captures well-established differences in political parallelism, journalistic professionalism and regulatory models (Hallin and Mancini 2004). The case selection also covers two-party and multi-party systems as well as different civic cultures. In addition, the relevance of online intermediaries for getting news differs across the six countries (see supplementary materials (SM) Table S2). If effects of using online intermediaries are consistent in this country sample, it is likely that the findings apply to developed democracies in general.

We recruited 7,775 study participants from the participant pool of the market research company *Netquest* that maintains online access panels with a continuous web tracking. Participants had given their informed consent and received incentives to install tracking tools and keep them active on their desktop computers.[1] In total, the data comprises 136 million website visits from 15 March to 16 June 2019. Our surveys of the same study participants covered key variables that have been identified as

[1]Participants could pause the tracking tool at any time. The median number of active days is 71 out of a maximum of 94.

sources of inequality in news exposure. Descriptive statistics of the sample composition and all used variables can be found in SM S1 to S4.[2] Despite the non-probabilistic sampling of participants, the data enable a detailed investigation of the effects of intermediary use on news exposure. Furthermore, online and offline news exposure and privacy attitudes of study participants resemble the patterns in external benchmarks (SM S5).

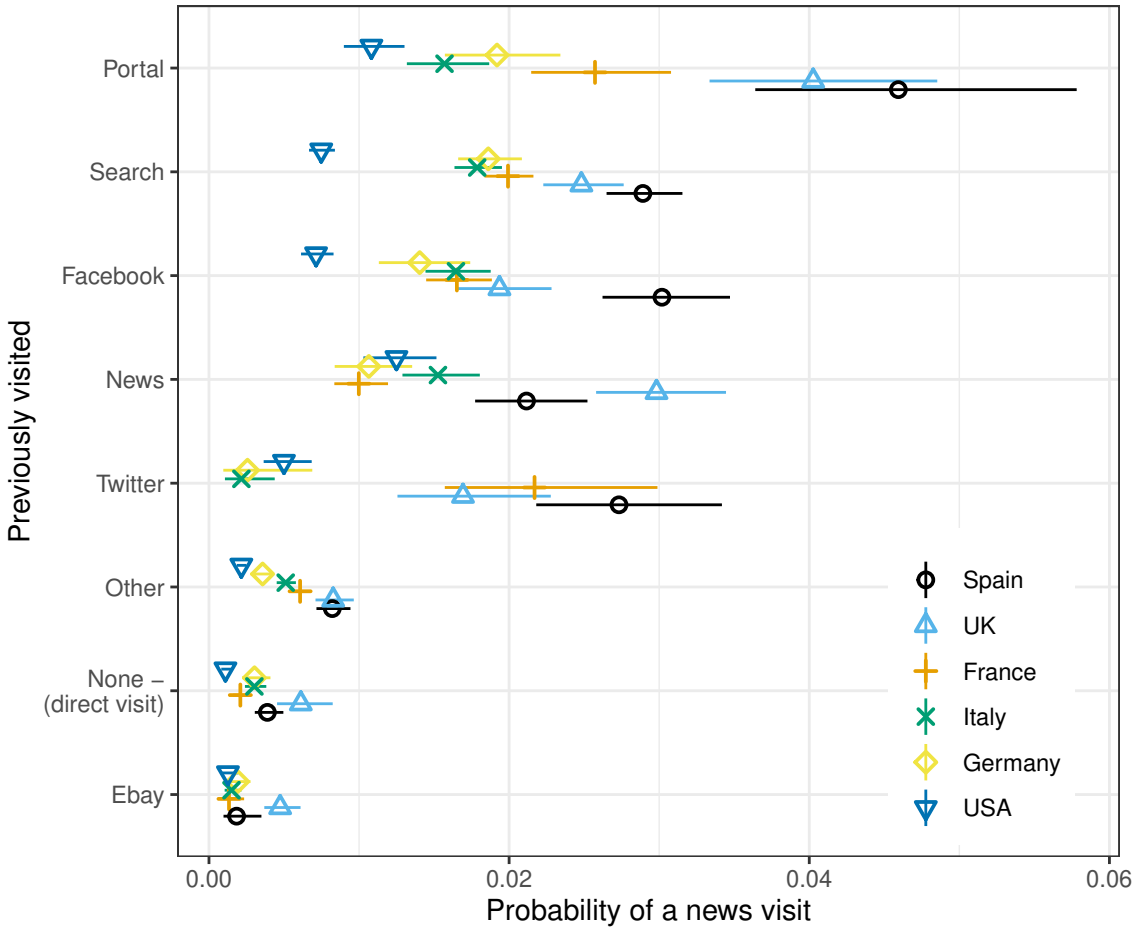
For identifying relevant website visits, the top 5,000 visited domains and most used intermediary platforms in each country were hand coded, covering 89% of all website visits (see SM S3 for a description of the coding and summary statistics). The identified 556 news domains were further grouped into seven news types like public broadcasting or hyperpartisan news with different styles of coverage. To construct a measure of political news exposure, we crawled the texts of all visited news URLs and trained a machine learning classifier (see SM S6 for a description and evaluation). We applied a visit threshold of 10 seconds to establish that respondents had at least somewhat engaged with a website (Lee and Kim 2017; Scharkow et al. 2020), after merging subsequent visits of the same URL to account for automatically reloading browser tabs.

RESULTS

To establish an empirical baseline, we ran a logistic regression model with person-level random intercepts that account for between-person differences in overall online activity. The model yielded that news accounted for only a small proportion of online activity, with estimated baseline probabilities of a news visit ranging from 0.005 for U.S. participants to 0.0157 for Spanish participants (Figure S6, descriptive statistics are in SM S4). To test a central tenet of our theory, Figure 1 breaks down the probability of a news visit conditional on the previously visited website. Despite differences in the base levels of news exposure, there were substantively large effects of using intermediaries across all countries. Compared to visits of other websites or direct visits (the first visit of a browsing session), being on portals, Facebook, search engines or Twitter increased the likelihood that the next URL is a

[2]Replication materials are available on the APSR Dataverse: https://doi.org/10.7910/DVN/P3XVF. The raw web tracking data and textual content of website URLs cannot be shared due to proprietary restrictions and to protect the privacy of study participants.

FIGURE 1. Probability of news exposure, conditional on the previous visit

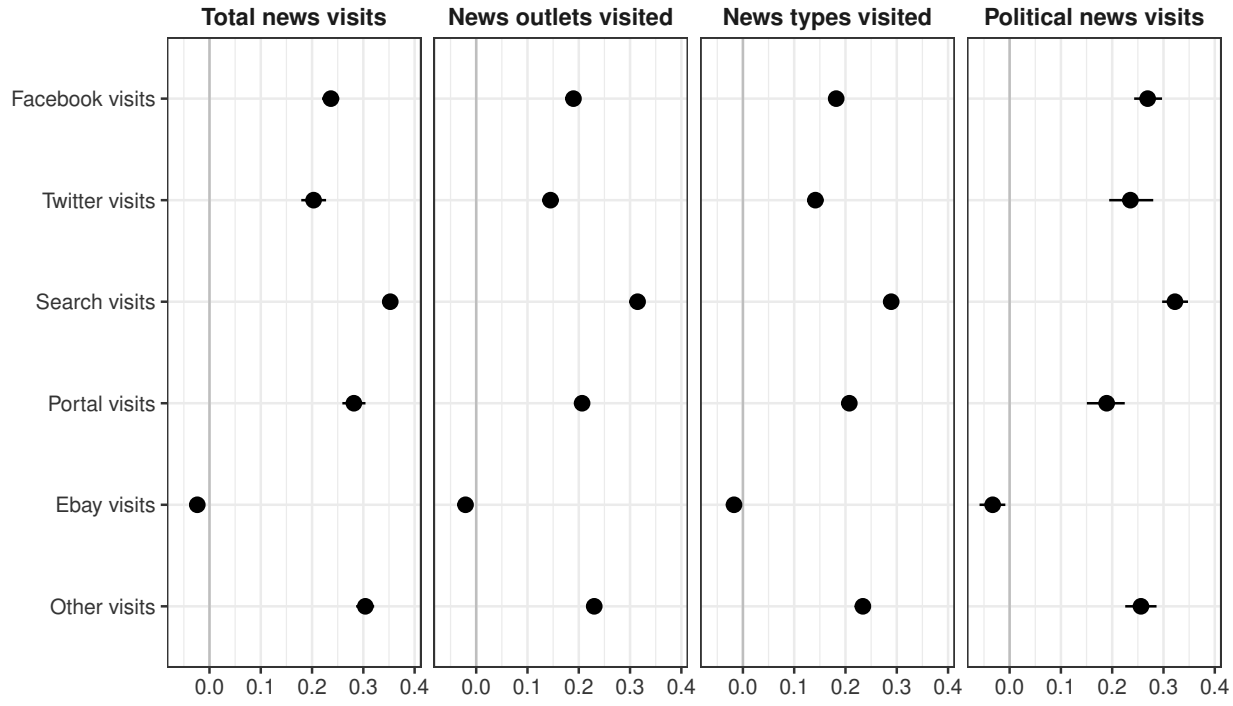


Note: Estimated marginal probabilities and 99% confidence intervals from a logistic regression model with person-level random intercepts. N = 27,028,342 domain visits. Subsequent URLs of the same domain were merged. Figure S7 shows the model results for political news visits.

news domain. Visits of news domains were also more likely to be followed by a different news domain, while the e-commerce platform eBay – that was included as a placebo test – had the weakest effect on subsequent news exposure. However, this clickstream analysis ignores indirect pathways from intermediaries to news (e.g., being primed by news-related content but visiting news sites later) and also does not encapsulate whether intermediaries foster news exposure beyond online users' regular media diets.

To provide a more nuanced view on news exposure and increase the identification power, we conducted a longitudinal analysis of the variability in participants' web browsing behavior across the three-month research period. For this, we aggregated visits on the daily level and used the random-

FIGURE 2. Within-person effects of daily intermediary use on daily news exposure



Note: Within-person Poisson regression coefficients and 99% confidence intervals from REWB models. N = 7,754 persons; 486,789 person-days. Figure S8 displays the estimated variability of the effects.

effects within-between (REWB) model (Bell et al. 2019) to separate the effects of regular online habits (between-person differences) from the effects of daily fluctuations in the use of intermediaries on non-regular news exposure (within-person effects), controlling for overall daily online activity (`Other visits`) and using a Poisson estimation to account for the dependent count variables (see SM S7 for technical details). In order to capture possible effect heterogeneity, all within-person effects were allowed to vary across participants (person-level random slopes).

We constructed two dependent variables similar to previous studies (Fletcher and Nielsen 2018a; Scharkow et al. 2020): the total number of news visits and the number of visited news outlets. Figure 2 shows that using more intermediaries on a given day positively predicts both outcomes. For instance, a person who was using search engines twice as often than on a typical day is estimated to consume 28% more news articles published by 24% more news outlets. The effects were positively significant for all types of intermediaries and general online activity, yet negative for the placebo eBay. Moreover, the estimated slopes were positive for almost all participants and all intermediaries (SM Figure S8). In

line with cross-sectional survey research (Fletcher and Nielsen 2018a), between-person effects were also consistently positive (SM Figure S9). The model results for two additional dependent variables speak against concerns about a shrinking diversity of news sources and avoidance of political news in digital high-choice media environments: Figure 2 also demonstrates that higher usage of any type of intermediary on a given day was also positively related to more distinct news types among the daily news visits and more *political* news exposure.[3]

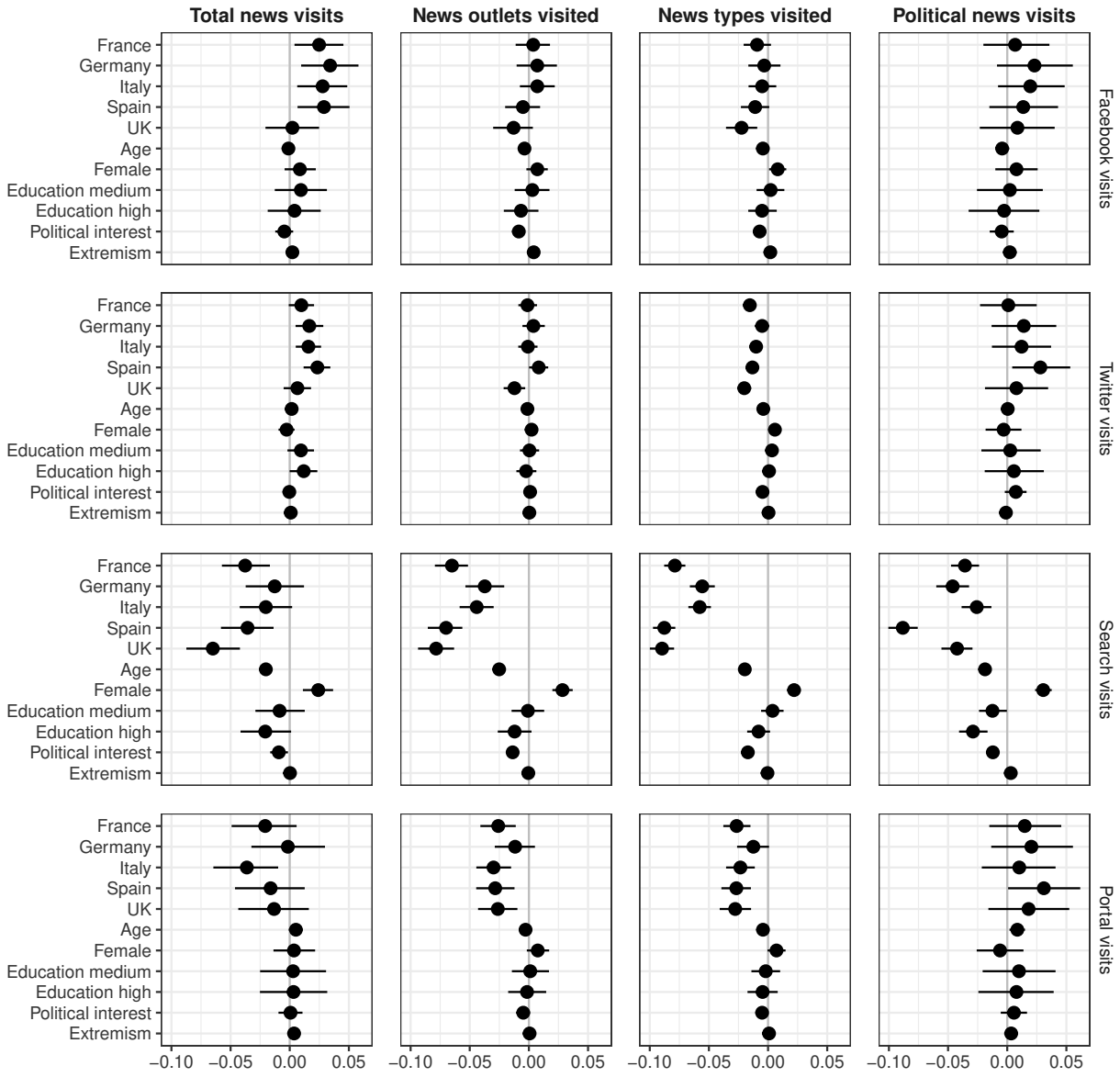
While the fixed effects reported in Figure 2 establish that social media, search and portal visits foster news exposure on average, there could be cross-country and person-level differences such that not all individuals profit equally from using intermediaries. We therefore investigated the moderating effects of country, age, gender, education, political interest and political extremism.[4] The corresponding results are reported in Figure 3. Turning to macro-level differences first, we see some cross-national variation. Most consistently, intermediaries had stronger effects on the number of news outlets and news types in the U.S. (the reference country), especially when it comes to using search engines and portals. On the individual level, deviations from the fixed effects were most pronounced for search engines, which increased news exposure more strongly for less politically interested, younger and female online users. Notwithstanding some statistically significant effects, the most important finding from Figure 3 is the overarching homogeneity of the within-person intermediary effects across all personal characteristics and countries. Even the most pronounced deviations from the average effects in Figure 2 were so small that the resulting conditional effects were still positive and significantly different from zero in all subgroups. For example, even the seemingly strong positive deviation for females when using search engines twice as often than on a regular day merely translates into 30% more news visits per day compared with the average effect of 28%.

With its cross-country design, the paper goes beyond previous web tracking studies that have investigated polarization in the US online information environment in isolation (Bakshy et al. 2015;

[3]The correlations between the dependent variables were highest for the number of news sources and unique news types (Kendall's $\tau = 0.84$). For the other pairs, correlations ranged from $\tau = 0.29$ to $\tau = 0.53$, indicating that these measures tapped into related, yet different behaviors.

[4]The direct effects of each covariate on news exposure were in line with existing research: older, male, highly educated and politically interested citizens consumed more news (SM Figure S10).

FIGURE 3. Variability of within-person effects across countries and personal characteristics



Note: Regression coefficients and 99% confidence intervals from moderation analyses of the random within-person slopes of the REWB model. Coefficients describe how, for any level of the moderating variable, the within-person effects of using intermediaries on news exposure deviated from the fixed effects displayed in Figure 2. Reference categories are "US" and "Education low". Age was divided by 10 before the estimation to improve interpretation. N = 7,622 persons; 478,647 person-days.

Flaxman et al. 2016; Guess 2021). We conducted an additional analysis to establish comparability with these studies and test whether using intermediaries biases news exposure in line with political predispositions. In models for US study participants relying on domain ideology data by Bakshy et al. (2015), media diet slant was not reinforced through intermediaries, neither according to party identification nor ideological self-placement (SM S8.4). Finally, robustness tests including self-reported measures of active news engagement, political discussion behavior and social network characteristics (Lee and Kim 2017; Thorson 2020; Figure S11) and using data from smartphones that were available for 36% of study participants (SM S8.3) confirmed the main results.

CONCLUSION

Online intermediaries such as Facebook or Google are often accused of contributing to the perils of democracy by exacerbating (political) news avoidance and one-sided content exposure. Yet our large-scale observation of web browsing behavior showed that intermediaries foster exposure to political and non-political news and its breadth in terms of the news sources and news types used. These effects were broadly consistent across types of intermediaries (social network sites, search engines, portals), diverse political and media systems and not strongly moderated by individual-level differences. Besides directly moderating media diets ideologically (Barberá 2015; Guess 2021), online intermediaries indirectly expose users to news content they would otherwise not see or click on. As such, the big online platforms counteract the fragmentation tendencies of the "post-broadcast democracy" that has been characterized by a proliferation of niche content (Prior 2007).

While individual-level web browsing histories coupled with surveys provide an unprecedented granularity, more superficial news exposure within platforms themselves is still not captured by tracking tools. Reassuringly, first academic evidence relying on data from the Facebook News Feed shows that the bulk of news exposure concentrates on URLs from credible news sources (Guess et al. 2021), suggesting that the platform creates opportunity structures for "incidental" and more substantive encounters with public affairs coverage. Combining such insights with repeated in-situ surveys would greatly enhance our capability to explore users' varied motivations and reactions to contents while navigating the web. Importantly, more research should be devoted to the downstream political effects

of socially mediated information exposure (Carlson 2019).

Some widely held assumptions about the negative consequences of intermediaries did not withstand scrutiny in our analysis, as there were no indications that intermediaries exacerbate inequalities in news exposure. At the same time, the evidence yielded only weak support that otherwise less avid news users benefit relatively more from intermediary use. The generally limited effect heterogeneity we identified corresponds with recent research showing that most online users devote only a marginal fraction of their online activities to news and that ideology has a minor impact (Fletcher et al. 2021; Guess 2021; Scharkow et al. 2020). Since concerns about apathetic citizens remain critical in an increasingly digital media environment (Prior 2007; Van Aelst et al. 2017), it is hard to overstate intermediaries' role in establishing a minimum level of exposure to news and political information. And yet Facebook's changes to its news feed aiming to prioritize content of friends at the expense of (news) organizations still illustrated that the big online platforms can change algorithms at their own discretion, with potentially profound democratic consequences.[5]

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit [Link to article on APSR website]

DATA AVAILABILITY STATEMENT

Research documentation and data that support the findings of this study are openly available on the APSR Dataverse: https://doi.org/10.7910/DVN/P3XVF. Limitations on data availability are discussed in the text.

ACKNOWLEDGEMENTS

We thank three reviewers, the editors and panel participants at the ECPR General Conference 2020 for helpful comments. We are grateful to Caterina Froio, Justin Ho, Nora Kirkizh, Ralph Schroeder and

[5]https://www.nytimes.com/2018/01/11/technology/facebook-news-feed.html.

Pu Yan for their contributions to the data collection and the coding of news articles. We thank Ina Böckmann and Céline Widera for excellent research assistance.

FUNDING STATEMENT

Sebastian Stier is grateful to the Volkswagen Foundation for funding the data collection (grant number 94 758).

CONFLICT OF INTEREST

The authors declare no ethical issues or conflicts of interest in this research.

ETHICAL STANDARDS

The authors declare that the data collection involving human subjects in this article was reviewed and approved by the Oxford Internet Institute's Departmental Research Ethics Committee at the University of Oxford. The certificate is provided on the APSR Dataverse. The authors affirm that this article adheres to the APSA's Principles and Guidance on Human Subjects Research.

REFERENCES

Anspach, Nicolas M. 2017. The new personal influence: How our Facebook friends influence the news we read. *Political Communication 34*(4), 590–606.

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science 348*(6239), 1130–1132.

Barberá, Pablo. 2015. *How social media reduces mass political polarization. Evidence from Germany, Spain, and the U.S.*

Bell, Andrew, Malcolm Fairbrother, and Kelvyn Jones. 2019. Fixed and random effects models: making an informed choice. *Quality & Quantity 53*(2), 1051–1074.

Bode, Leticia. 2016. Pruning the news feed: Unfriending and unfollowing political content on social media. *Research & Politics 3*(3).

Carlson, Taylor N. 2019. Through the grapevine: Informational consequences of interpersonal political communication. *American Political Science Review 113*(2), 325–339.

Flaxman, Seth, Sharad Goel, and Justin M. Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly 80*(S1), 298–320.

Fletcher, Richard and Rasmus Kleis Nielsen. 2018a. Are people incidentally exposed to news on social media? A comparative analysis. *New Media & Society 20*(7), 2450–2468.

Fletcher, Richard and Rasmus Kleis Nielsen. 2018b. Automated Serendipity: The effect of using search engines on news repertoire balance and diversity. *Digital Journalism 6*(8), 976–989.

Fletcher, Richard, Craig T. Robertson, and Rasmus Kleis Nielsen. 2021. How many people live in politically partisan online news echo chambers in different countries? *Journal of Quantitative Description: Digital Media 1*.

Guess, Andrew M. 2021. (Almost) everything in moderation: New evidence on Americans' online media diets. *American Journal of Political Science*.

Guess, Andrew M., Kevin Aslett, Joshua Tucker, Richard Bonneau, and Jonathan Nagler. 2021. Cracking open the News Feed: Exploring what U.S. Facebook users see and share with large-scale platform data. *Journal of Quantitative Description: Digital Media 1*.

Hallin, Daniel C. and Paolo Mancini. 2004. *Comparing media systems: Three models of media and politics*. Cambridge, UK: Cambridge University Press.

Kalogeropoulos, Antonis, Richard Fletcher, and Rasmus Kleis Nielsen. 2019. News brand attribution in distributed environments: Do people know where they get their news? *New Media & Society 21*(3), 583–601.

Kümpel, Anna S. 2020. The Matthew Effect in social media news use: Assessing inequalities in news exposure and news engagement on social network sites (SNS). *Journalism 21*(8), 1083–1098.

Lee, Jae Kook and Eunyi Kim. 2017. Incidental exposure to news: Predictors in the social media setting and effects on information gain online. *Computers in Human Behavior 75*, 1008–1015.

Möller, Judith, Robbert N. van de Velde, Lisa Merten, and Cornelius Puschmann. 2020. Explaining online news engagement based on browsing behavior: Creatures of habit? *Social Science Computer Review 38*(5), 616–632.

Prior, Markus. 2007. *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge, UK: Cambridge University Press.

Prior, Markus. 2009. The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly 73*(1), 130–143.

Scharkow, Michael, Frank Mangold, Sebastian Stier, and Johannes Breuer. 2020. How social network sites and other online intermediaries increase exposure to news. *Proceedings of the National Academy of Sciences 117*(6), 2761–2763.

Skovsgaard, Morten and Kim Andersen. 2020. Conceptualizing news avoidance: Towards a shared understanding of different causes and potential solutions. *Journalism Studies 21*(4), 459–476.

Stier, Sebastian, Johannes Breuer, Pascal Siegers, and Kjerstin Thorson. 2020. Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review 38*(5), 503–516.

Sunstein, Cass R. 2009. *Republic.com 2.0*. Princeton, N.J.: Princeton Univ. Press.

Thorson, Kjerstin. 2020. Attracting the news: Algorithms, platforms, and reframing incidental exposure. *Journalism 21*(8), 1067–1082.

Tucker, Joshua A., Andrew M. Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. *Social media, political polarization, and political disinformation: A review of the scientific literature*.

Van Aelst, Peter, Jesper Strömbäck, Toril Aalberg, Frank Esser, Claes de Vreese, Jörg Matthes, David Hopmann, Susana Salgado, Nicolas Hubé, Agnieszka Stępińska, Stylianos Papathanassopoulos, Rosa Berganza, Guido Legnante, Carsten Reinemann, Tamir Sheafer, and James Stanyer. 2017. Political communication in a high-choice media environment: a challenge for democracy? *Annals of the International Communication Association 41*(1), 3–27.

Post Post-Broadcast Democracy? News Exposure in the Age of Online Intermediaries

Online Appendix

Published in *The American Political Science Review*

Sebastian Stier[1], Frank Mangold[2], Michael Scharkow[3], and Johannes Breuer[1,4]

[1]GESIS – Leibniz Institute for the Social Sciences
[2]University of Hohenheim
[3]Johannes Gutenberg University Mainz
[4]Center for Advanced Internet Studies (CAIS)

Contents

S1 Sample composition and country coverage 1
S2 Survey items 2
S3 News domain coding 3
S4 Descriptive statistics 4
S5 Comparison of samples to external benchmarks 7
S6 Classifying political news articles 10
S7 Description of the statistical models 15
S8 Additional results 16
S8.1 Log-level analysis 16
S8.2 Daily within-person analysis 18
S8.3 Mobile use 22
S8.4 Ideological alignment of news domains 25

S1 Sample composition and country coverage

Data was collected from online access panels of the market research company *Netquest* whose participants consented to regularly participate in surveys and install browser plugins tracking their browsing behavior on desktop computers. In countries (France, Spain, US) where the web tracking panels had a sufficient size, approximately 1,500 participants were invited according to population margins, but some quota cells still remained empty. In countries where the panels did not have a sufficient size, all panelists were invited (Germany, Italy, UK).

Compared to national population margins, the samples are skewed towards well-educated female middle-aged people (Table S1). Especially younger and older people as well as the lower educated strata of the population are underrepresented. At the same time, especially elderly and lower educated people generally use the Internet less. Unfortunately, high-quality benchmark data on the demographics of Internet users are not available for each country.

Table S1: Demographics by country (%).

	Gender	Age					Education*		
Country	Female	29/under	30-39	40-49	50-59	60/over	Low	Medium	High
France	55.00	14.84	19.67	22.02	21.88	21.60	4.69	51.48	43.82
Germany	51.27	13.29	19.32	20.55	27.33	19.51	28.28	37.70	34.02
Italy	57.84	14.49	25.24	28.64	19.63	12.00	10.26	46.74	43.00
Spain	51.52	20.97	10.56	15.99	22.53	29.96	24.09	33.46	42.45
UK	52.29	9.72	15.32	20.64	23.85	30.46	4.95	48.53	46.51
US	65.56	13.04	20.75	18.52	23.85	23.85	4.47	60.59	34.94

Note: *Harmonized based on the International Standard Classification of Education (ISCED).

Table S2: Country-level characteristics

Country	Party system	Media system	SM for news (%)	Side door to news (%)
France	Multi-party	Polarized pluralist	42	65
Germany	Multi-party	Democratic corporatist	34	59
Italy	Multi-party	Polarized pluralist	47	67
Spain	Multi-party	Polarized pluralist	53	68
UK	Two-party	Liberal (+ strong public broadcasting)	40	54
US	Two-party	Liberal	46	66

Note: SM for news: came across news stories on social media. Side door: keyword search, social media, aggregator, email, notifications as pathways to news. Data from the Reuters Digital News Report 2019 (Newman et al., 2019).

S2 Survey items

Table S3: Survey items used in the analysis

Variable	Description	Original source
Age	Self-reported age. Divided by 10 before the regression estimations to improve interpretation.	European Social Survey, Round 8 (ESS) (ERIC, 2017)
Education	Country-specific education levels that were recoded into "low", "medium" and "high education" based on the country-comparative ISCED scheme.	ESS
Gender	Self-reported gender. Female was coded as 1, male and the few "other gender" responses as 0.	ESS
Political extremism	Based on a political ideology question ranging from 0 (left) to 10 (right). The end points in the US survey were labeled "very liberal" and "very conservative". Political extremism is calculated as the absolute distance of an individuals' ideology to the country mean (see also Barberá, 2015).	ESS
Political interest	Measured as a 4-point scale ranging from "not at all interested" to "very interested".	ESS
Political talk	Frequency of discussions about national and local political matters with friends and relatives.	Eurobarometer 83.3 (European Commission, 2018)
Social media for news	Importance of social media for keeping up with political news, debates and discussions. 4-point scale from "not at all important" to "very important".	Quello Search Project (Dutton et al., 2017)
Social media disagreement	Disagreement with political opinions or political content contacts post on social media. 5-point scale from "almost never" to "nearly always".	Quello Search Project
Social media engagement	Political and news engagement on social network sites: comment on a news story; share content related to political issues originally posted by someone else; "like" political pages or political posts others have posted; post own thoughts or comments on political issues; post links to political stories or articles.	Reuters Digital News Report 2019 (Newman et al., 2019)

S3 News domain coding

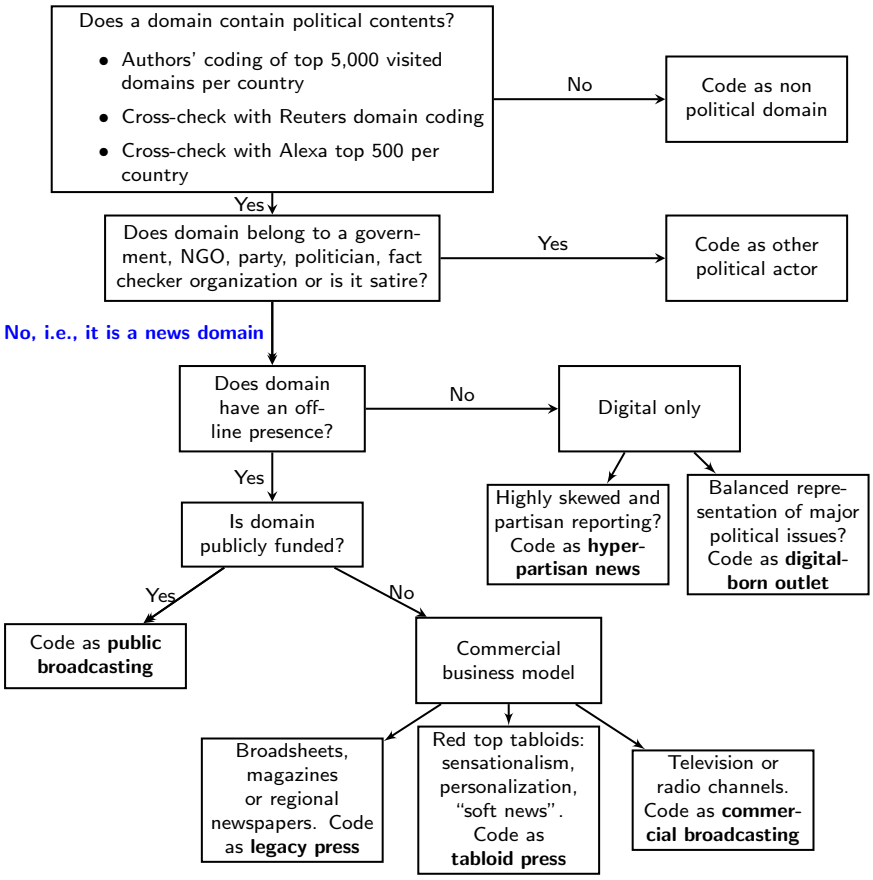


Figure S1: Description of the domain coding.

Table S4: Website visits covered by coding

Country	Unique domains	Visits	Visits covered	Share covered (%)
France	134,102	30,040,775	26,073,607	87
Germany	95,716	16,392,236	15,482,615	94
Italy	128,110	24,981,435	21,735,489	87
Spain	109,845	15,497,311	13,093,310	84
UK	116,357	20,009,587	18,328,899	92
US	165,751	29,239,470	26,639,489	91

Note: Subsequent visits of the same URL were merged to account for reloading tabs.

S4 Descriptive statistics

Table S5: Descriptive statistics, France

Variable	N	Mean	St. Dev.	Min	Median	Max
Survey variables						
Age	1,444	46.29	14.28	18.0	47.0	85
Education	1,444	2.39	0.58	1.0	2.0	3
Female	1,444	0.55	0.50	0.0	1.0	1
Political interest	1,443	2.73	0.91	1.0	3.0	4
Political extremism	1,440	1.93	1.75	0.1	1.9	5.1
Political talk	1,305	1.98	1.18	0.0	2.0	4
Social media for news	1,341	2.60	0.91	1.0	3.0	4
Social media disagreement	1,341	2.96	0.79	1.0	3.0	5
Social media engagement	1,308	0.95	1.47	0.0	0.0	5
Dependent variables (daily)						
Total news visits	1,444	2.31	8.28	0.0	0.0	289
News outlets visited	1,444	0.75	1.45	0.0	0.0	33
News types visited	1,444	0.61	0.95	0.0	0.0	7
Political news visits	1,444	0.25	1.37	0.0	0.0	154
Intermediaries visited (daily)						
Facebook visits	1,444	13.62	27.07	0.0	2.0	687
Twitter visits	1,444	0.61	5.54	0.0	0.0	245
Search visits	1,444	10.13	17.38	0.0	4.0	320
Portal visits	1,444	7.85	19.37	0.0	0.0	558
Ebay visits	1,444	0.83	6.42	0.0	0.0	211
Total visits	1,444	115.87	109.03	1.0	86.0	1,512

Table S6: Descriptive statistics, Germany

Variable	N	Mean	St. Dev.	Min	Median	Max
Survey variables						
Age	1,055	46.94	14.05	18.00	48.00	84
Education	1,055	2.06	0.79	1.00	2.00	3
Female	1,055	0.51	0.50	0.00	1.00	1
Political interest	1,052	2.87	0.86	1.00	3.00	4
Political extremism	1,055	1.49	1.34	0.36	1.36	5.36
Political talk	901	2.20	1.05	0.00	2.00	4
Social media for news	949	2.70	0.84	1.00	3.00	4
Social media disagreement	949	3.15	1.01	1.00	3.00	5
Social media engagement	903	1.02	1.36	0.00	0.00	5
Dependent variables (daily)						
Total news visits	1,055	2.19	9.00	0.00	0.00	391
News outlets visited	1,055	0.51	1.10	0.00	0.00	17
News types visited	1,055	0.41	0.77	0.00	0.00	6
Political news visits	1,055	0.27	2.08	0.00	0.00	145
Intermediaries visited (daily)						
Facebook visits	1,055	7.36	20.38	0.00	0.00	635
Twitter visits	1,055	0.38	4.09	0.00	0.00	198
Search visits	1,055	7.18	14.64	0.00	2.00	514
Portal visits	1,055	8.40	17.03	0.00	0.00	536
Ebay visits	1,055	2.92	14.80	0.00	0.00	557
Total visits	1,055	94.61	108.92	1.00	61.00	1,604

Table S7: Descriptive statistics, Italy

Variable	N	Mean	St. Dev.	Min	Median	Max
Survey variables						
Age	1,436	43.78	12.82	18.00	43.00	88
Education	1,436	2.33	0.65	1.00	2.00	3
Female	1,436	0.58	0.49	0.00	1.00	1
Political interest	1,434	2.74	0.84	1.00	3.00	4
Political extremism	1,431	2.20	1.68	0.28	2.28	5.28
Political talk	1,306	2.41	1.14	0.00	2.00	4
Social media for news	1,392	2.90	0.77	1.00	3.00	4
Social media disagreement	1,390	2.88	0.73	1.00	3.00	5
Social media engagement	1,309	1.77	1.72	0.00	1.00	5
Dependent variables (daily)						
Total news visits	1,436	2.24	7.12	0.00	0.00	452
News outlets visited	1,436	0.60	1.23	0.00	0.00	27
News types visited	1,436	0.46	0.77	0.00	0.00	6
Political news visits	1,436	0.17	1.07	0.00	0.00	86
Intermediaries visited (daily)						
Facebook visits	1,436	13.68	29.56	0.00	1.00	700
Twitter visits	1,436	0.37	3.82	0.00	0.00	203
Search visits	1,436	10.74	16.94	0.00	5.00	327
Portal visits	1,436	5.94	13.03	0.00	0.00	292
Ebay visits	1,436	1.29	8.58	0.00	0.00	459
Total visits	1,436	99.18	98.77	1.00	69.00	1,045

Table S8: Descriptive statistics, Spain

Variable	N	Mean	St. Dev.	Min	Median	Max
Survey variables						
Age	1,342	48.49	16.75	18.00	50.00	87
Education	1,342	2.18	0.79	1.00	2.00	3
Female	1,342	0.52	0.50	0.00	1.00	1
Political interest	1,341	2.72	0.83	1.00	3.00	4
Political extremism	1,342	2.23	1.47	0.09	1.91	6.09
Political talk	1,191	2.08	1.17	0.00	2.00	4
Social media for news	1,276	2.91	0.91	1.00	3.00	4
Social media disagreement	1,272	2.90	0.71	1.00	3.00	5
Social media engagement	1,193	1.48	1.76	0.00	1.00	5
Dependent variables (daily)						
Total news visits	1,342	3.73	10.70	0.00	0.00	270
News outlets visited	1,342	0.86	1.64	0.00	0.00	27
News types visited	1,342	0.61	0.95	0.00	0.00	7
Political news visits	1,342	0.57	2.55	0.00	0.00	139
Intermediaries visited (daily)						
Facebook visits	1,342	7.49	20.27	0.00	0.00	522
Twitter visits	1,342	1.62	11.34	0.00	0.00	400
Search visits	1,342	11.38	18.96	0.00	5.00	485
Portal visits	1,342	1.34	6.11	0.00	0.00	408
Ebay visits	1,342	0.46	5.59	0.00	0.00	432
Total visits	1,342	79.18	83.97	1.00	53.00	979

Table S9: Descriptive statistics, UK

Variable	N	Mean	St. Dev.	Min	Median	Max
Survey variables						
Age	1,090	50.66	14.90	18.00	51.00	89
Education	1,090	2.42	0.59	1.00	2.00	3
Female	1,090	0.52	0.50	0.00	1.00	1
Political interest	1,089	2.72	0.91	1.00	3.00	4
Political extremism	981	1.34	1.44	0.04	0.96	5.04
Political talk	981	1.90	1.10	0.00	2.00	4
Social media for news	995	2.43	0.97	1.00	3.00	4
Social media disagreement	994	3.10	0.81	1.00	3.00	5
Social media engagement	982	1.10	1.62	0.00	0.00	5
Dependent variables (daily)						
Total news visits	1,090	4.23	10.11	0.00	0.00	208
News outlets visited	1,090	0.76	1.17	0.00	0.00	16
News types visited	1,090	0.67	0.92	0.00	0.00	6
Political news visits	1,090	0.44	2.13	0.00	0.00	170
Intermediaries visited (daily)						
Facebook visits	1,090	10.99	24.96	0.00	1.00	511
Twitter visits	1,090	1.63	12.50	0.00	0.00	459
Search visits	1,090	11.34	23.66	0.00	3.00	1,076
Portal visits	1,090	6.11	15.56	0.00	0.00	430
Ebay visits	1,090	3.98	15.71	0.00	0.00	416
Total visits	1,090	114.18	112.31	1.00	81.00	1,264

Table S10: Descriptive statistics, USA

Variable	N	Mean	St. Dev.	Min	Median	Max
Survey variables						
Age	1,387	47.54	14.80	18.00	48.00	85
Education	1,387	2.30	0.55	1.00	2.00	3
Female	1,387	0.66	0.48	0.00	1.00	1
Political interest	1,386	2.64	0.97	1.00	3.00	4
Political extremism	1,382	2.14	1.69	0.46	1.54	5.46
Political talk	1,173	1.77	1.25	0.00	2.00	4
Social media for news	1,346	2.58	1.00	1.00	3.00	4
Social media disagreement	1,347	3.09	0.93	1.00	3.00	5
Social media engagement	1,177	1.49	1.79	0.00	1.00	5
Dependent variables (daily)						
Total news visits	1,387	1.51	5.84	0.00	0.00	329
News outlets visited	1,387	0.41	1.00	0.00	0.00	32
News types visited	1,387	0.35	0.72	0.00	0.00	7
Political news visits	1,387	0.24	1.50	0.00	0.00	68
Media diet slant	1,132	-0.13	0.24	-0.77	-0.13	0.91
Intermediaries visited (daily)						
Facebook visits	1,387	14.69	29.70	0.00	2.00	815
Twitter visits	1,387	1.05	10.54	0.00	0.00	765
Search visits	1,387	15.22	25.07	0.00	6.00	389
Portal visits	1,387	10.07	21.26	0.00	0.00	627
Ebay visits	1,387	1.48	10.32	0.00	0.00	518
Total visits	1,387	125.11	140.97	1.00	86.00	3,157

S5 Comparison of samples to external benchmarks

To assess the generalizability of the news consumption behavior of study participants, we compare the popularity of news domains in our data to their visit numbers in the top 500 Alexa country rankings for the three months of our data collection.[1] Alexa has the advantage that the data is available across countries, as it tracks the website visits of more than 300 million users who have installed a web browser plugin. Nevertheless, it is still unclear how representative of each countries' online population the data is. Figure S2 shows the correspondence between the number of news website visits in both data sources.

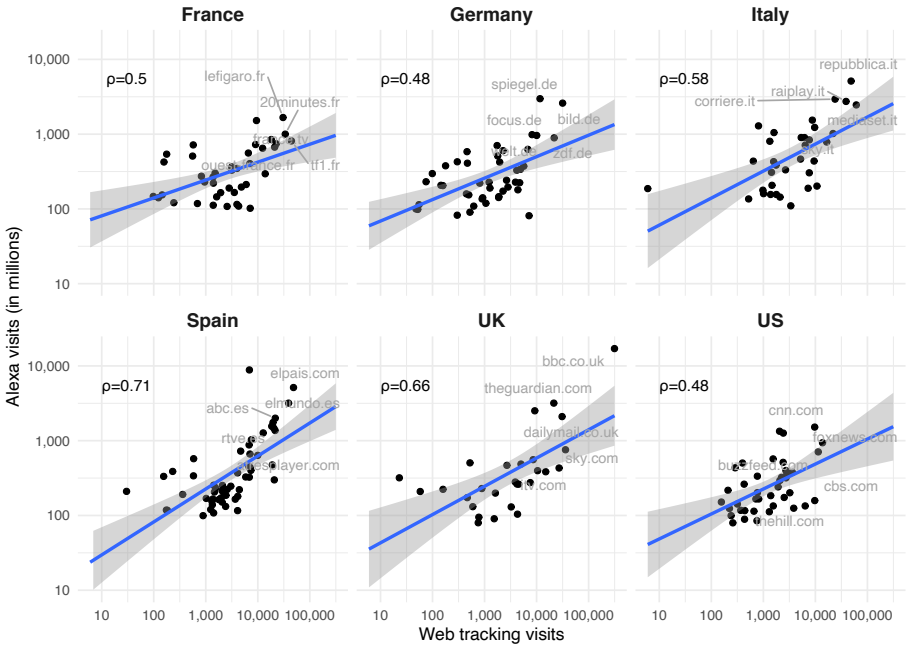


Figure S2: Popularity of news websites in the top 500 Alexa rankings per country and among web tracking panelists. ρ = Spearman's rank correlations.

[1] https://www.alexa.com/siteinfo

Participants in an online web tracking might have a higher propensity to get news from online media instead of newspapers, television and radio. To compare offline news exposure to an external benchmark, we implemented self-report items of media exposure from the Reuters Digital News Report (DNR) 2019 (Newman et al., 2019) in our surveys. The high correlations demonstrate that the study participants were equally likely to get news from newspapers and in particular from the major television news programs in each country.

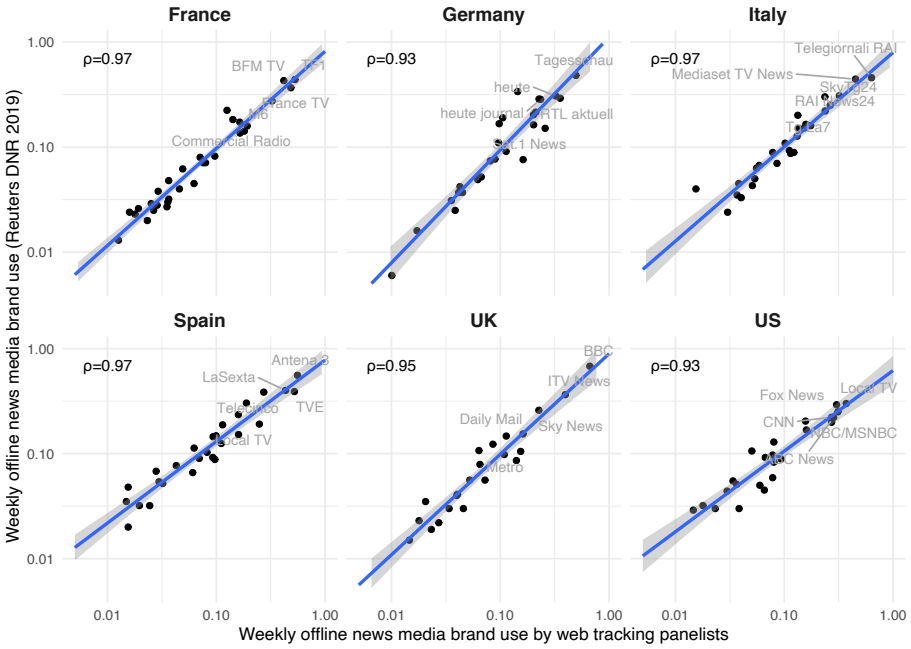


Figure S3: Weekly offline news media brands used, comparison of Reuters Digital News Report 2019 and web tracking panelists. ρ = Spearman's rank correlations.

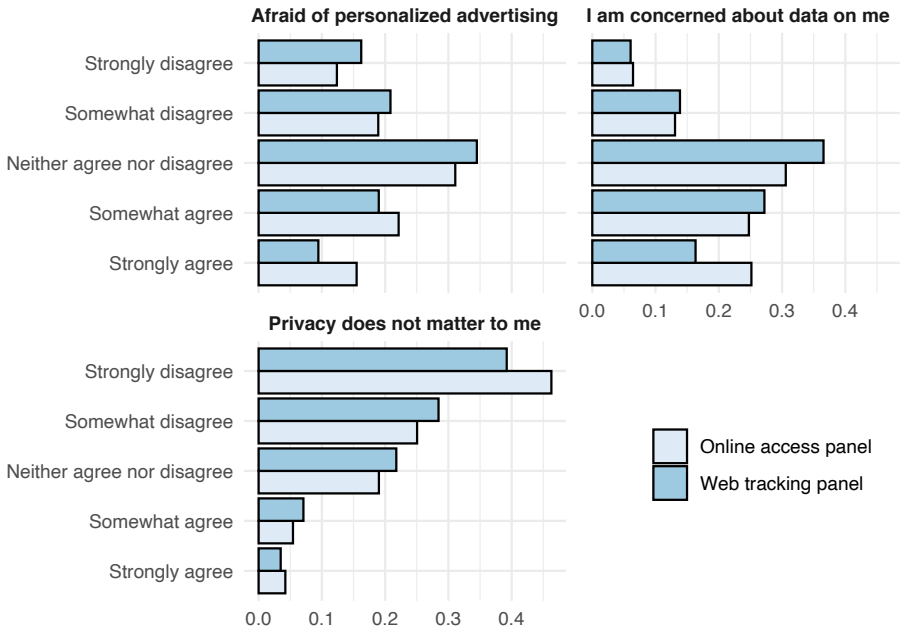


Figure S4: Privacy attitudes among German web tracking participants and a sample of German online access panelists without web tracking.

Another important way in which the study participants could differ from other online news users might be privacy attitudes. We therefore investigated to what extent privacy attitudes of web tracking panelists diverged from panelists who participate in surveys, but do not have tracking tools installed (replicating the approach of Guess, 2021). As a comparison group, we sampled 1,002 German participants based on population margins for gender, age and education from the regular online access panel of the same survey company. Respondents were presented the following statements and asked about their (dis)agreement on a five-point scale.

- Personalized advertising makes me afraid.
- I am concerned about how much data there is about me on the Internet.
- My privacy on the Internet does not matter to me.

Figure S4 shows that there were only marginal differences in the privacy attitudes of online access panelists who participated in the web tracking and those who did not. Yet as outlined in Section S1, we cannot draw inferences to the privacy attitudes of the German general population from these data.

S6 Classifying political news articles

Building on previous work combining web tracking data and article content (Bakshy et al., 2015; Flaxman et al., 2016; Guess, 2021), we constructed a classifier for each country that predicts whether the news articles visited by panelists contain political content. To get the textual contents, all unique URLs were crawled with the R package *rvest* (Wickham, 2020). The article text was parsed from the downloaded html files using the Python library *newspaper* (Ou-Yang, 2013).

For training the text classification model, we first selected five major news websites in each country and identified all of their articles that contain one of the unambiguous political keywords `polit`, `democrac` or `elect` in the URL (see Table S11).[2] The respective five news outlets per country were chosen based on two considerations: (1) they are popular among our panelists and the overall online population according to the Reuters Digital News Report (Newman et al., 2019), and (2) they have a website/URL architecture with a specific `politics` subsection.[3]

Table S11: Selected news domains and political keywords per country

Country	Outlets	Detected keywords in URLs
France	lefigaro.fr, 20minutes.fr, lemonde.fr, francetvinfo.fr, lepoint.fr	polit, democra, elections
Germany	bild.de, welt.de, focus.de, spiegel.de, sueddeutsche.de	polit, demokrat, wahl
Italy	repubblica.it, corriere.it, mediaset.it, leggo.it, ilmessaggero.it	polit, democraz, elezion
Spain	elpais.com, lavanguardia.com, elperiodico.com, eldiario.es, cadenaser.com	polit, democra, elecciones
UK	bbc.co.uk, theguardian.com, telegraph.co.uk, mirror.co.uk, independent.co.uk	polit, policy, democrac, elect
US	cnn.com, foxnews.com, nytimes.com, washingtonpost.com, nbcnews.com	polit, policy, democrac, elect

[2] We defined all articles as political whose content is related to either polity (e.g., political institutions, democracy), politics (e.g., elections, political actors, scandals) or policy (e.g., regulation or legislation with regard to substantive issues, which excludes non-policy aspects like crime reports). A hand-coding of 100 randomly selected articles for each country shows that only 18 out of 600 articles identified by the predefined political keywords were not political.

[3] For instance, the latter criterion disqualified dailymail.co.uk, the third most popular news domain among UK panelists.

We treated the URLs of the five selected news domains that do not include one of the political keywords as the negative set of training articles. This is a restrictive operationalization of political news, as URLs on other sections of these websites also contain political content. Guess (2021) and Flaxman et al. (2016), in contrast, used a more extensive training dataset including URLs published on website subsections such as `business`, `national` or `news`. Our classifiers thus underestimate the share of political content, yet thanks to its parsimony, the approach is comparable across countries. Consequently, if we find effects of intermediary use on political news exposure, the true effects are most likely even stronger.

The following text preprocessing steps were taken before training the classifiers.[4]

1. We excluded the top level news domains (e.g., nytimes.com), as the content on these pages changes dynamically and therefore differed at the time of crawling from the time of the actual website visit made by a panelist.
2. As the included news domains publish in five different languages, we removed English, French (keeping the string "eu"), German, Italian and Spanish stopwords.
3. We removed punctuation, numbers, hyphens and symbols.
4. We reduced the corpus for each country to words that occur at least 20 times.

The frequencies of words in these pre-processed corpora already reveal a clear signal: political articles have a distinct vocabulary compared with non-political articles (Figure S5).

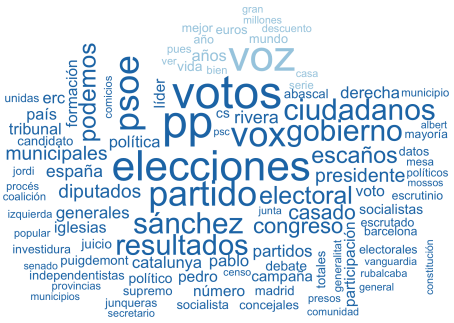
Using the articles including the URL keywords in Table S11 as "gold standard" labels for political news coverage, a Naive Bayes classifier was trained for each country and evaluated against a held-out set of test data using ten-fold cross-validation. The average performance of each Naive Bayes classifier per country across its respective ten folds is listed in Table S12. The classifiers accurately identified political articles, mirroring the performance achieved in similar applications (Flaxman et al., 2016; Guess, 2021). Table S13 shows the most predictive features for classifying articles as political or non-political.

[4]The text analysis was performed using the R package *quanteda* (Benoit et al., 2018).

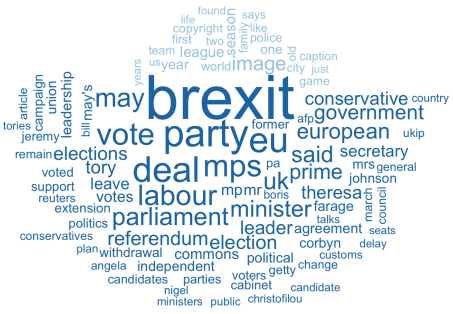
France

Germany

Italy



Spain



UK

US

Figure S5: Wordclouds with most frequent words in political articles (bottom of each figure) vs. non-political articles (top of each figure) on the websites of five major news outlets per country.

Table S12: Results from ten-fold cross-validation

Country	Accuracy	Precision	Recall	F1
France	0.92	0.92	0.99	0.96
Germany	0.91	0.91	0.98	0.95
Italy	0.93	0.93	0.99	0.96
Spain	0.89	0.88	0.98	0.93
UK	0.94	0.95	0.99	0.97
US	0.88	0.89	0.96	0.92

Note: Mean values from ten-fold cross-validation.

Table S13: Most predictive features for classifying political vs. non-political articles

Country	Political	Non-political
France	plus, c'est, macron, liste, france, aussi, président, comme, fait, parti, européennes, qu'il, politique, emmanuel, d'un, tout, d'une, être, faire, deux	plus, c'est, d'un, deux, ans, aussi, comme, d'une, fait, tout, france, après, euros, bien, être, paris, selon, qu'il, prix, faire
Germany	prozent, spd, mehr, sagte, eu, partei, deutschland, cdu, afd, wurde, trump, anzeige, menschen, grünen, seit, europawahl, wahl, zwei, schon, lesen	mehr, wurde, schon, euro, zwei, immer, lesen, ab, gibt, beim, jahren, deutschland, seit, geht, anzeige, mal, menschen, sagte, drei, bild
Italy	salvini, lega, governo, pd, partito, stato, m5s, ministro, maio, poi, presidente, dopo, fa, solo, italia, prima, elezioni, due, c'è, essere	anni, stato, dopo, due, prima, essere, poi, solo, quando, fatto, ancora, stata, sempre, così, casa, fa, via, euro, molto, fare
Spain	pp, partido, gobierno, elecciones, vox, psoe, votos, sánchez, ciudadanos, dos, podemos, españa, país, electoral, ser, madrid, presidente, casado, años, tras	voz, años, dos, ser, puede, así, hace, ahora, españa, según, vez, euros, solo, después, madrid, tres, día, además, cada, año
UK	brexit, said, party, deal, may, eu, vote, mps, uk, labour, people, parliament, minister, government, mr, one, european, prime, getty, new	said, image, one, people, first, year, time, can, two, new, just, caption, getty, us, years, now, says, copyright, like, last
US	trump, said, president, house, report, mueller, trump's, one, news, democrats, new, campaign, people, justice, told, fox, us, white, barr, investigation	said, one, people, new, like, just, time, year, can, told, get, news, first, two, years, now, trump, according, day, even

After the evaluation, the classifier was applied to all news articles that were visited by the study participants to predict whether these are political or not. A validation of the classifier predictions by hand-coding 100 randomly selected articles for each country showed a highly accurate out-of-sample performance.[5] For constructing the final measure identifying political articles, we first applied the political URL keywords listed in Table S11 to all URLs and only used the classifier predictions if there was no positive string match. In addition to the 113,420 news website visits classified as political by the URL keywords, we identified additional 178,969 visits to political articles (e.g., on website sections such as `national` or `society`) thanks to the classifier.

Taken together, the share of political news among all URLs of news websites varied between 7.8% (Italy) and 17.2% (US). While highest among our set of countries, the share for the US was still smaller than reported in Guess (2021) (19% in 2015, 23% in 2016). This can be explained by (1) our less expansive definition of news (defined through only a few political URL keywords) and (2) the ongoing presidential election in his 2016 study that most likely increased exposure to political news.

[5] Accuracy 0.90, Precision 0.90, Recall 0.97, F1 0.94, with only minor variation across countries.

S7 Description of the statistical models

Model specification

Following Bell et al. (2019), we estimated random effects within-between (REWB) models, which are mixed effects regressions that include both person-mean centered (within) predictors and person-level averages (between). For a single predictor variable x, i respondents and t repeated measurements, the REWB model is specified as

$$y_{it} = \mu + \beta_{1W}\left(x_{it} - \overline{x}_i\right) + \beta_{2B}\overline{x}_i + v_{i0} + v_{i1}\left(x_{it} - \overline{x}_i\right) + \epsilon_{it0}$$

with β_{1W} as the within-person effect, β_{2B} as the between-person effect. In addition to random intercepts for respondents v_{i0}, the model also includes random slopes v_{i1} for the within-person effect in order to obtain conservative estimates and allow for subsequent analyses of effect heterogeneity. Since the dependent variables are counts, we used a Poisson GLM and included random intercepts for days and observations in order to control for possible period effects as well as overdispersion (Harrison, 2014). All within-between predictor variables were log(x+1)-transformed to account for days with zero intermediary or news visits and since we expected nonlinear effects.

Model estimation

The REWB models are computationally demanding, especially with very large samples such as ours. As a consequence, estimating the model above using the full sample and predictors including cross-level interactions to investigate effect heterogeneity resulted in convergence problems. In order to get reliable estimates, we therefore split the data into ten equally sized respondent samples, stratified by country. We then followed a three-step approach:

1. For every fold and every outcome, we estimated the above mentioned Poisson REWB model using REML implemented in the R package *lme4* (Bates et al., 2015). We saved all parameter estimates, both for the fixed and random parts of the model.
2. We then ran several mini meta-analyses using the fixed effects estimates and their standard errors as data, using the R package *brms* (Bürkner, 2018). This yielded a meta-analytic (average) effect and credible intervals for every predictor in the model, as shown in Figure 2 in the main paper.
3. In order to investigate the between-person heterogeneity of the effects, we extracted and pooled the random intercepts and slopes from all folds, including their standard errors, and estimated a second set of meta-analyses, this time including person-level characteristics as covariates, again using *brms*. This slopes-as-outcome analysis allowed us to estimate the differences in the intercepts and within-person effects between different (groups of) respondents, as displayed in Figure 3 in the main paper.

S8 Additional results

S8.1 Log-level analysis

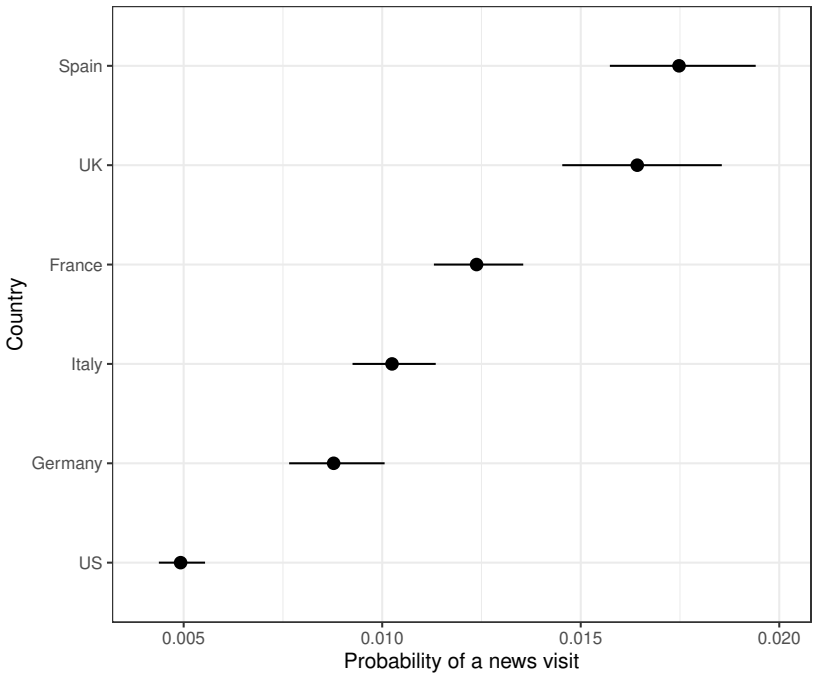


Figure S6: Baseline probability of a news visit. Results from a logistic regression model with person-level random intercepts that take into account the between-person differences in overall online activity. N = 27,028,342 domain visits (subsequent URLs of the same domain were merged).

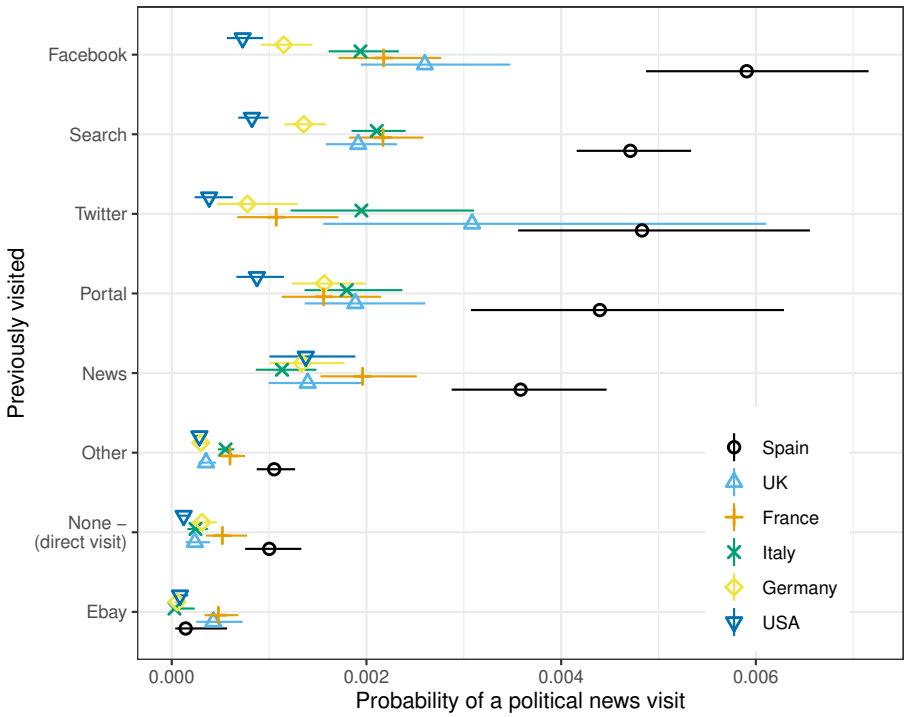


Figure S7: Probability of exposure to **political news**, conditional on the previously visited website. Estimated marginal probabilities and 99% confidence intervals from a logistic regression model with person-level random intercepts. N = 27,028,342 domain visits. Subsequent URLs of the same domain were merged and the visit marked as political when at least one URL was classified as such (see Section S6).

S8.2 Daily within-person analysis

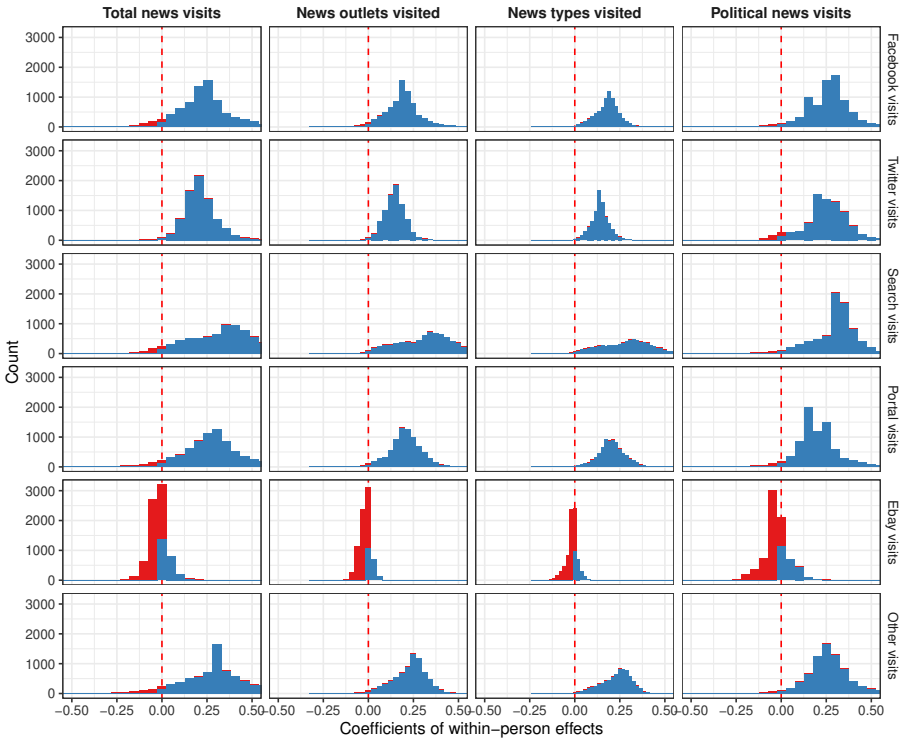


Figure S8: Distribution of varying coefficients of the within-person effects (Figure 2 in the main paper). N = 7,754 persons; 486,789 person-days.

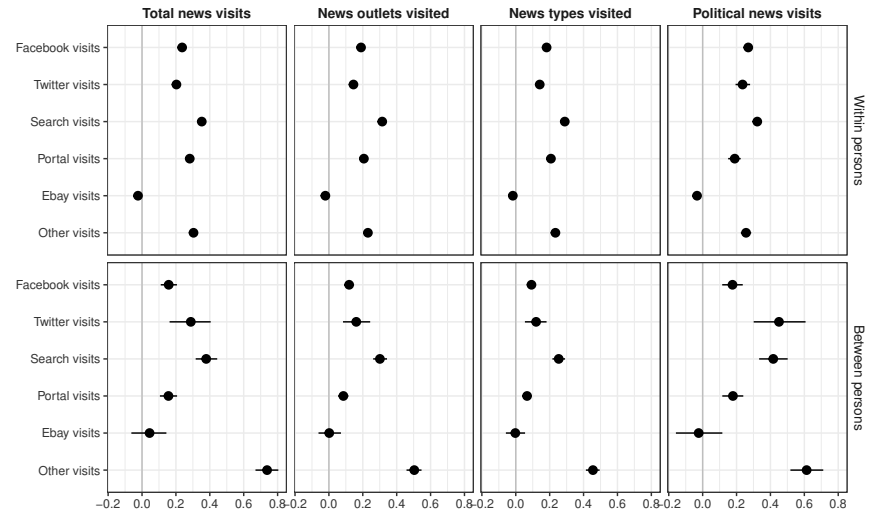


Figure S9: Within-person (see also Figure 2 in the main paper) and between-person Poisson regression coefficients and 99% confidence intervals from REWB models. N = 7,754 persons; 486,789 person-days.

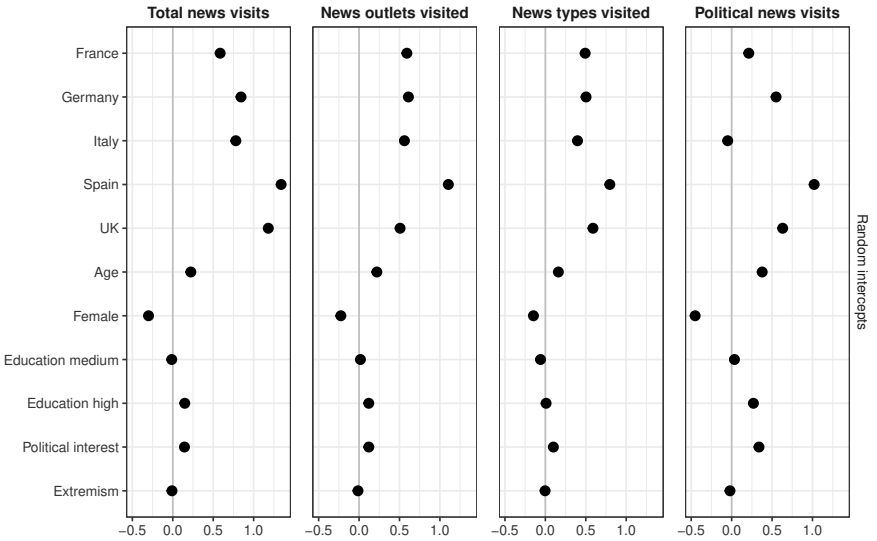


Figure S10: Moderation analyses of the **random intercepts** from the REWB models. Regression coefficients and 99% confidence intervals. Reference categories are "US" and "Education low". Age was divided by 10 before the estimation to improve interpretation. N = 7,622 persons; 478,647 person-days.

A host of studies have identified political discussion behavior and the composition of personal social networks as important correlates of online and offline political news engagement (Barberá, 2015; Boulianne and Koc-Michalska, 2021; Lee and Kim, 2017; Vaccari et al., 2016). The number and political heterogeneity of contact networks, the share of weak ties, and the propensity of discussing news and politics with these contacts affect how often online users will encounter news. The algorithms of online intermediaries are likely to pick up these patterns in user behavior and further feed news content into the information stream of politically engaged citizens (Thorson, 2020).

To incorporate such individual-level correlates of getting exposed to news via online intermediaries, we use additional survey items (see Table S3 for original sources of the survey items and question wording and Section S4 for descriptive statistics):

- Importance of social media for getting news. Only available for persons who reported having an account on at least one social network site.
- Disagreement with the political opinions or political content contacts post on social media. Only available for persons who reported having an account on at least one social network site.
- Political and news engagement on social network sites. Summed index of activities such as commenting on a news story and posting on political issues in the last 12 months (range 0 to 5).
- Political talk frequency with friends and relatives. Summed index of discussions about national and local political matters (range 0 to 4; "occasionally" coded as 1, "frequently" coded as 2).

While more fine-grained measurements have been used in some studies (Lee and Kim, 2017), the available survey items represent the theoretically relevant target concepts well. It is noteworthy that political discussion frequency and the political and news engagement battery were implemented in a later survey wave six weeks after the baseline survey. Due to unit non-response, the number of included respondents is reduced to $N = 6,408$ persons in these models.

Figure S11 shows that even after inclusion of these covariates and despite a reduced number of included persons, the main results from Figure 3 in the main paper hold. The most likely explanation is that these additional variables require a high degree of personal involvement that is already captured by political interest – an important predictor of news exposure (see Figure S10).

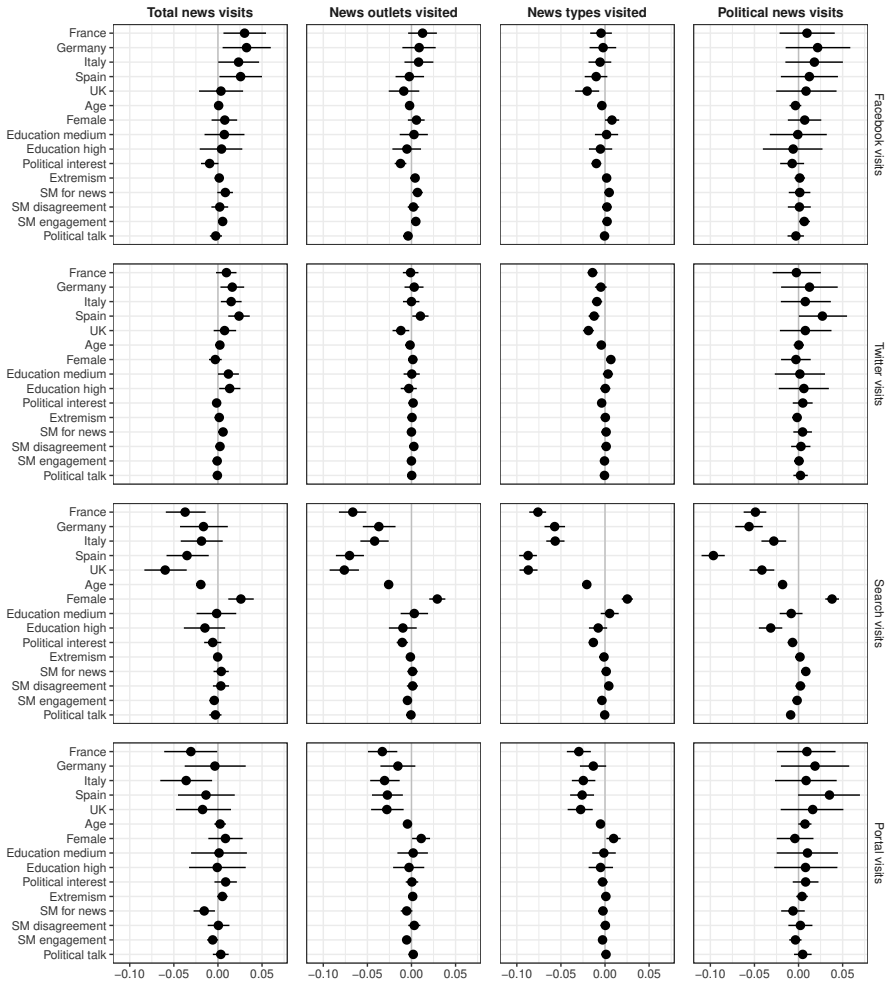


Figure S11: Regression coefficients and 99% confidence intervals from moderation analyses of the random within-person slopes of the REWB model with **additional covariates included**. Coefficients describe how, in any given subgroup, the effects of the random slopes deviated from the fixed effects in Figure 2 in the main paper. Reference categories are "US" and "Education low". Age was divided by 10 before the estimation to improve interpretation. "SM" = Social media. N = 6,408 persons; 413,978 person-days.

S8.3 Mobile use

In addition to desktop browsing, mobile data was available for 36% of study participants. The mobile tracking captures website visits in mobile browsers and app usage. Besides the domain codes used in the desktop analysis, we also coded the top 5,000 used apps as news or the different intermediaries). The analysis for political news visits cannot be replicated, as the mobile data captures the full URL for non-https traffic only and does not provide any information about the content seen in apps.

The results are similar to the patterns for desktop browsing in the main paper. However, there are two noteworthy differences: (1) the share of news in the media diet of the smartphone sample is lower than among desktop users; and (2) the effects of intermediaries are generally weaker, especially in the case of Facebook and portals (predominantly apps that provide direct access to emails without getting exposed to the starting pages of portals).

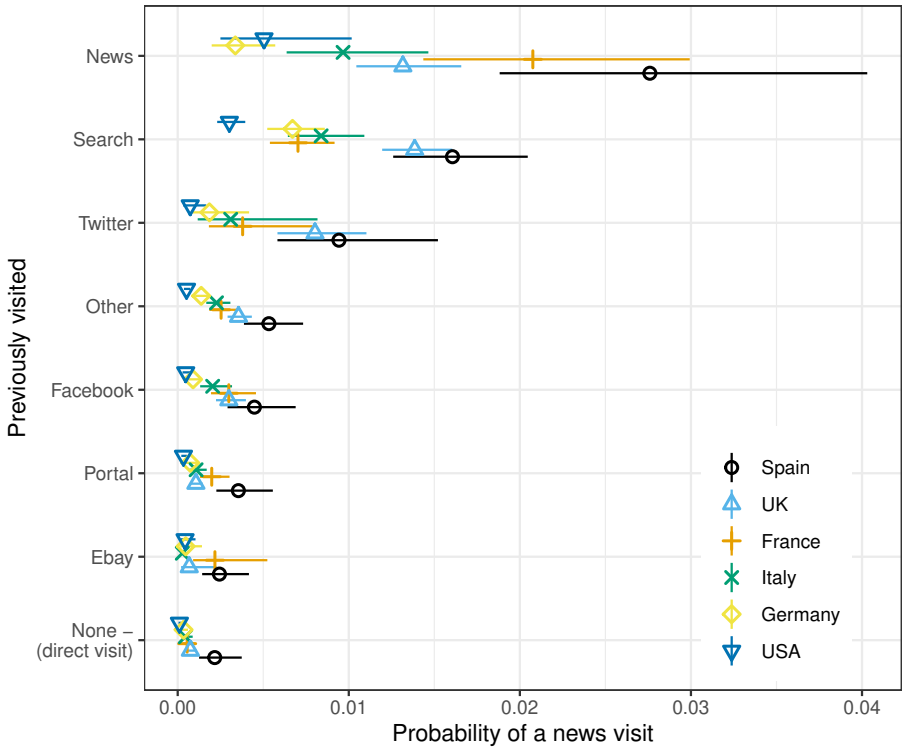


Figure S12: Probability of exposure to **news on mobile apps or mobile browsers**, conditional on the previously visited website. Estimated marginal probabilities and 99% confidence intervals from a logistic regression model with person-level random intercepts. $N = 9{,}056{,}404$ domain or app visits (subsequent visits of the same domain or app were merged).

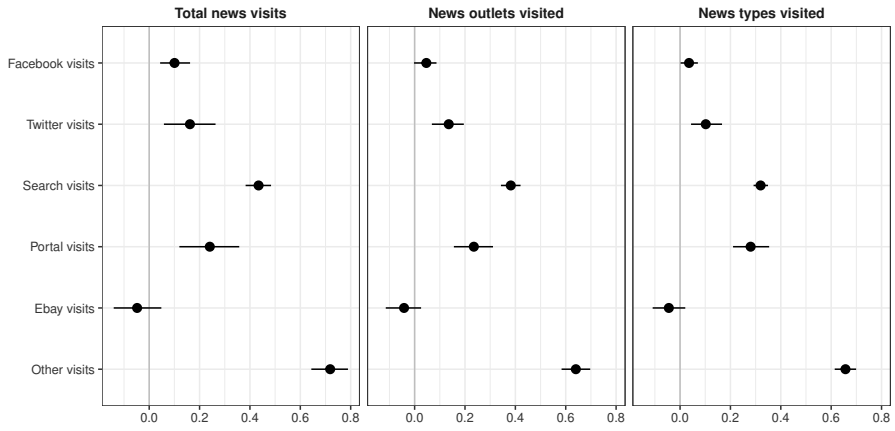


Figure S13: Within-person Poisson regression coefficients and 99% confidence intervals from REWB models, estimated on **use of mobile apps or mobile browsers**. N = 2,830 persons; 173,071 person-days.

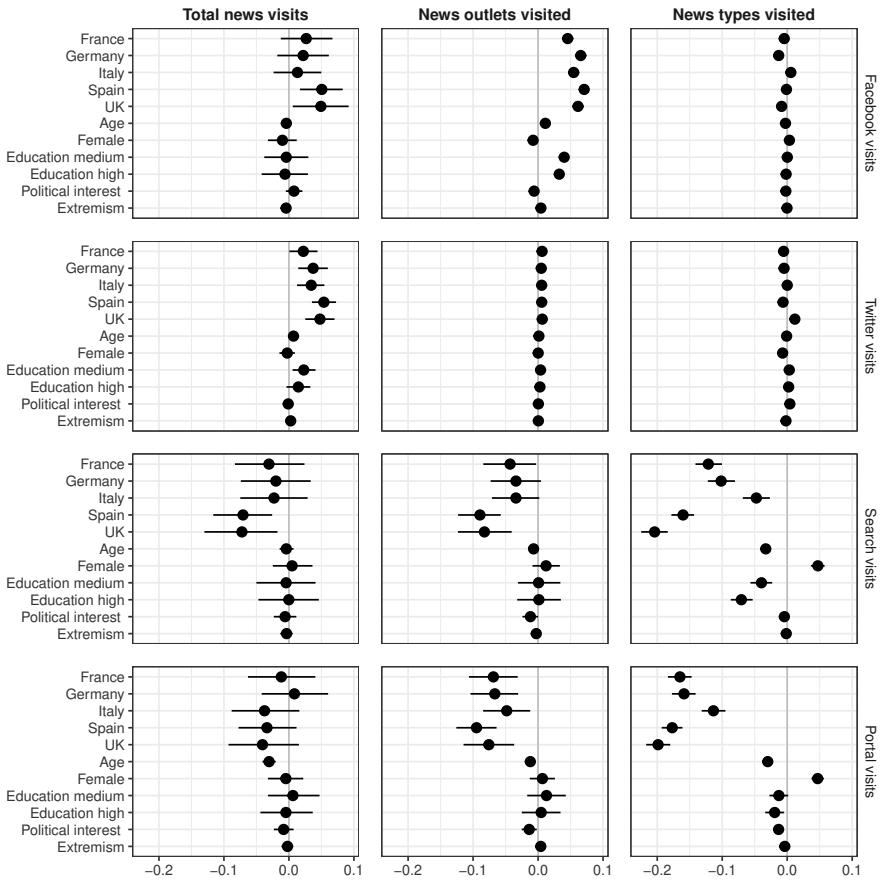


Figure S14: Regression coefficients and 99% confidence intervals from moderation analyses of the random within-person slopes of the REWB model, estimated on **use of mobile apps or mobile browsers**. Coefficients describe how, for any level of the moderating variable, the within-person effects of using intermediaries on news exposure deviated from the fixed effects displayed in Figure S13. Reference categories are "US" and "Education low". Age was divided by 10 before the estimation to improve interpretation. N = 2,803 persons; 171,410 person-days.

S8.4 Ideological alignment of news domains

As a direct test of whether online intermediaries narrow the ideological diversity of media diets, we used the alignment scores of Bakshy et al. (2015) to calculate the slant of study participants' media diets. We restrict this analysis to the US sample as the domain alignment scores were constructed based on the ideology of US Facebook users and are therefore only available for American website domains. The platforms `amazon.com`, `twitter.com` and `youtube.com` that got assigned an alignment score were excluded, while the alignment scores for the portals `aol.com`, `msn.com` and `yahoo.com` were only assigned to their news sections instead of the entire domain. In addition to the individual-level covariates from the main paper, we use party identification for this analysis (38% Democrats, 37% Independents including other parties, 25% Republican). The share of partisans is similar to the American National Election Studies 2016 (ANES, 2021), with a slight over-representation of Democrats.

Figure S15 plots the average daily media slant, with negative values representing a more liberal and positive values a more conservative media diet. The distribution resembles the results of Guess (2021), but is more bumpy due to the aggregation at the daily level instead of the respondent level and is on average shifted slightly more to the right.

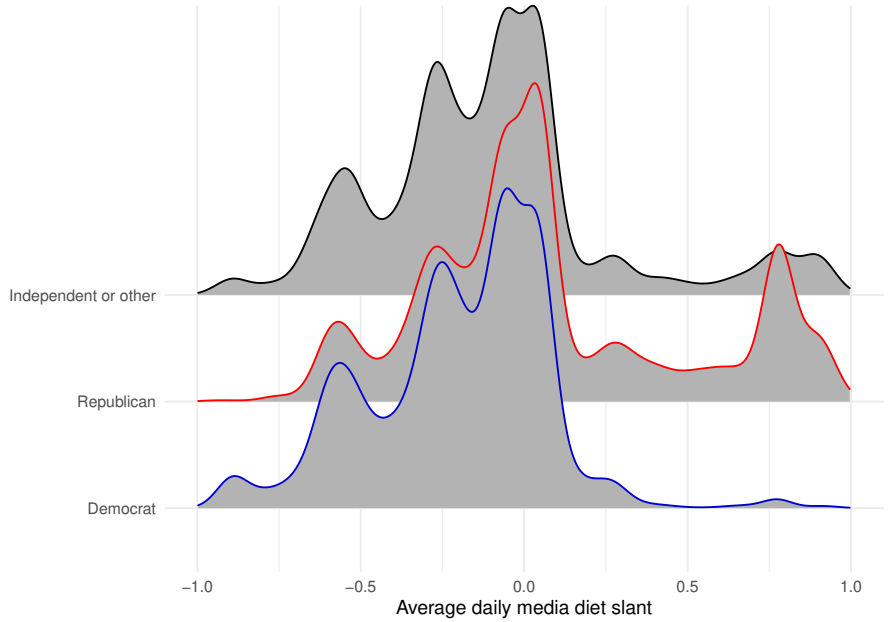


Figure S15: Average daily media diet slant. 23,153 person-days for US study participants with at least one visit to a news website with an alignment score by Bakshy et al. (2015).

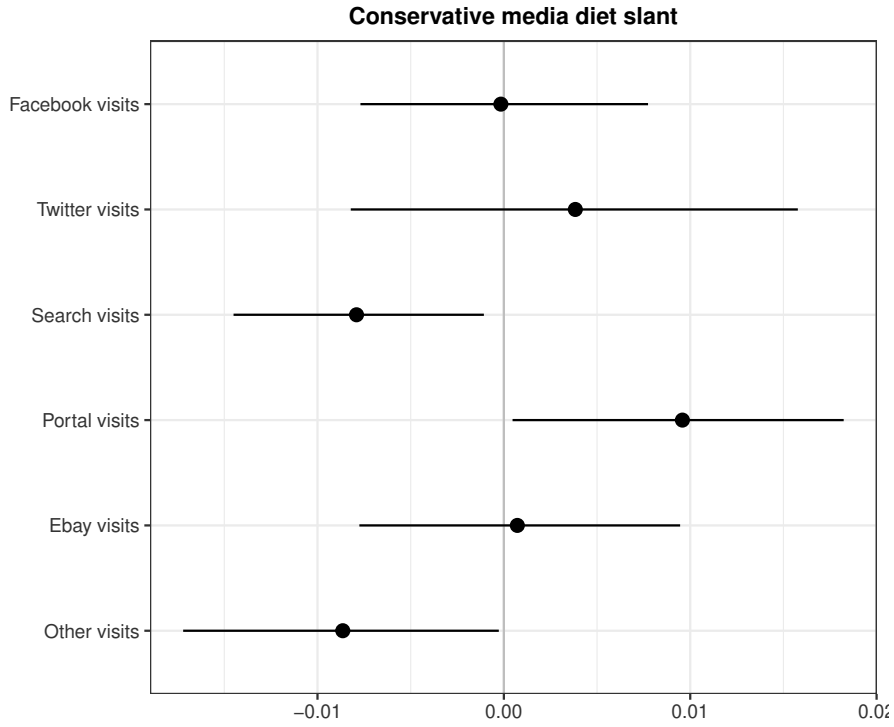


Figure S16: Within-person linear regression coefficients and 99% confidence intervals of daily intermediary use on media diet slant. REWB models estimated for US study participants. N = 1,131 persons; 23,153 person-days.

We used the mean domain alignment for each person-day when a panelist visited at least one website with an alignment score as the dependent variable in linear regressions.[6] Figure S16 shows that daily intermediary use is only weakly related to the ideological slant of media diets. Daily search frequencies and having more daily visits overall is associated with a more liberal media diet. The positive effect of portals on conservative media diets can be explained by the overall left-leaning audience distribution. The alignment scores of all news sections of portals are to the right of the mean ideology of news domains visited by US study participants (-0.097). Therefore using portals (e.g., for checking emails), which frequently results in visits of portals' news sections (see also the log-level results in Figure 1 in the main paper), shifts media diets towards the conservative pole, on average.

These fixed effects are again broken down by personal characteristics in a moderation analysis. Figure S17 first shows the random intercepts. Compared

[6]The analysis covers 24.5% of all US person-days. 81% of US study participants visited a domain with an alignment score at least once during our research period.

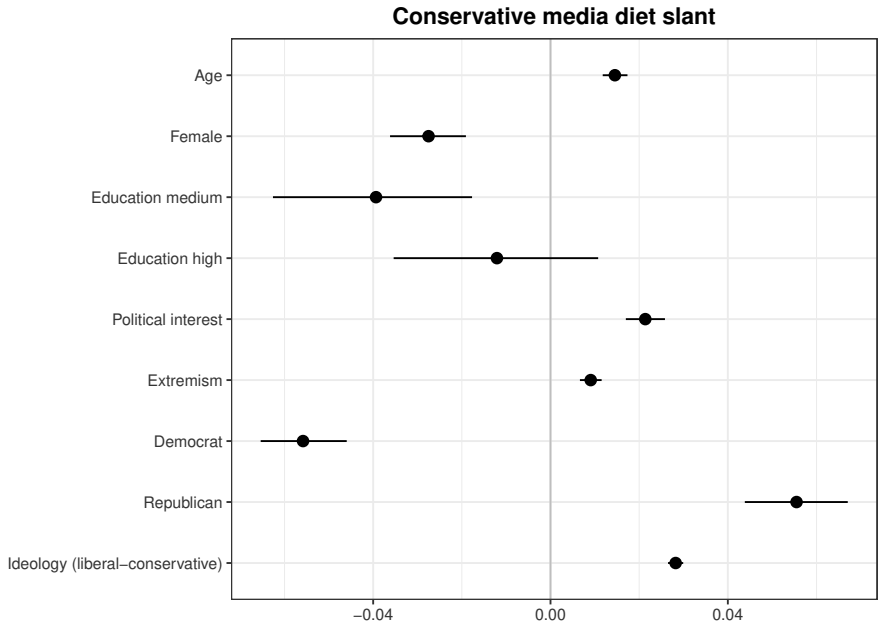


Figure S17: Regression coefficients and 99% confidence intervals from moderation analyses of the **within-person random intercepts** of the REWB model among US study participants (see Figure S16). The reference categories are "Education low" and "Independent" (party identification). Age was divided by 10 before the estimation to improve interpretation. N = 1,126 persons; 23,096 person-days.

with self-identified Independents, Democrats' media diet is shifted to to the liberal end of the full ideological distribution by 27%, while Republicans's media diet skews to the conservative side by roughly the same amount.[7] As expected, ideologically more conservative persons had a more conservative media diet.

If widespread assumptions about supposed echo chambers bear some semblance of reality, the random within-person slopes should show that using more intermediaries on a given day reinforces the slant of the daily media diet, depending on individual-level factors. Yet Figure S18 again reveals no noteworthy micro-level heterogeneity. Most importantly, neither being a Democrat or Republican nor ideological self-placement significantly moderated the effects of intermediaries on the daily media diet slant.[8]

[7] While our data source and regression models differ from Guess (2021), it appears that the ideological spread of media diets has grown since 2016.

[8] The coefficients are very similar when the models are estimated only with party identification but without political ideology.

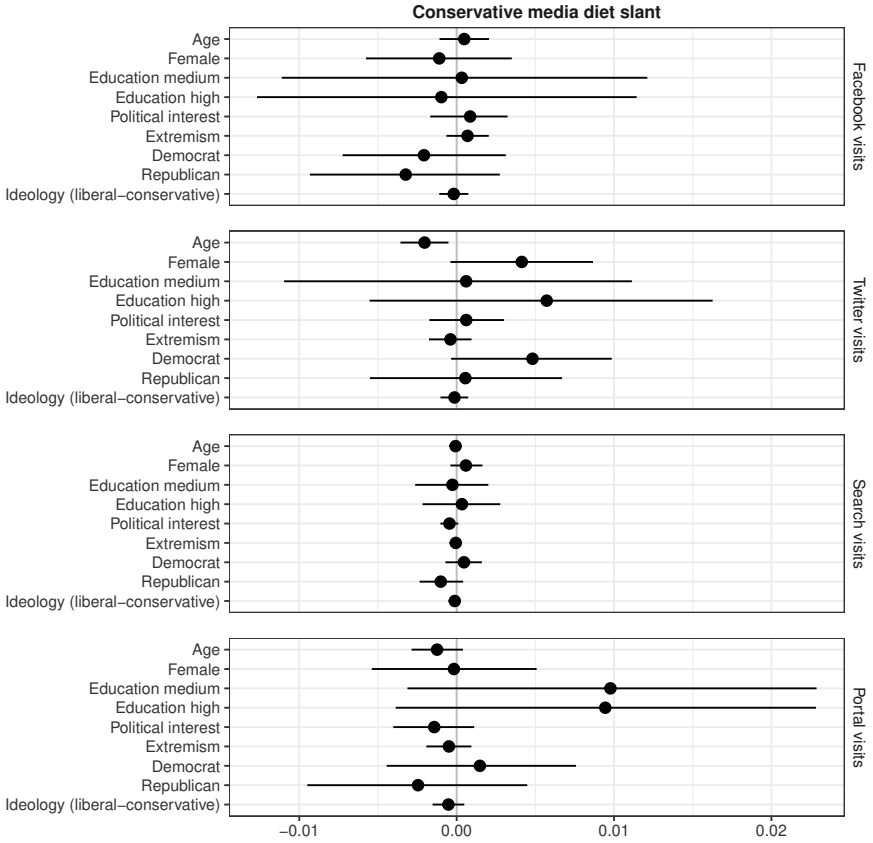


Figure S18: Regression coefficients and 99% confidence intervals from moderation analyses of the **within-person random slopes** of the REWB model among US study participants. Coefficients describe how, for any level of the moderating variable, the within-person effects deviated from the fixed effects displayed in Figure S16. Reference categories are "Education low" and "Independent" (party identification). Age was divided by 10 before the estimation to improve interpretation. N = 1,126 persons; 23,096 person-days.

References

ANES. 2021. Party identification 7-point scale 1952-2016. https://electionstudies.org/resources/anes-guide/top-tables/?id=21.

Bakshy, Eytan, Solomon Messing, and Lada Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science 348*(6239), 1130–1132.

Barberá, Pablo. 2015. *How social media reduces mass political polarization. Evidence from Germany, Spain, and the U.S.*

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*(1).

Bell, Andrew, Malcolm Fairbrother, and Kelvyn Jones. 2019. Fixed and random effects models. *Quality & Quantity 53*(2), 1051–1074.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software 3*(30).

Boulianne, Shelley and Karolina Koc-Michalska. 2021. The Role of Personality in Political Talk and Like-Minded Discussion. *The International Journal of Press/Politics*.

Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal 10*(1), 395–411.

Dutton, William H, Bianca Reisdorf, Elizabeth Dubois, and Grant Blank. 2017. Search and politics. *SSRN*.

ERIC, European Social Survey. 2017. European Social Survey (ESS), Round 8 - 2016. Publisher: NSD - Norwegian Centre for Research Data.

European Commission. 2018. Eurobarometer 83.3 (2015). GESIS Data Archive, Cologne. ZA5998 Datafile Version 2.0.0, https://doi.org/10.4232/1.13133.

Flaxman, Seth, Sharad Goel, and Justin Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly 80*(1), 298–320.

Guess, Andrew M. 2021. (Almost) Everything in moderation: New evidence on Americans' online media diets. *American Journal of Political Science*.

Harrison, Xavier A. 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ 2*.

Lee, Jae Kook and Eunyi Kim. 2017. Incidental exposure to news: Predictors in the social media setting and effects on information gain online. *Computers in Human Behavior 75*, 1008–1015.

Newman, Nic, Richard Fletcher, Antonis Kalogeropoulos, and Rasmus Kleis Nielsen. 2019. Reuters Institute Digital News Report 2019.

Ou-Yang, Lucas. 2013. *Newspaper3k: Article scraping & curation*.

Thorson, Kjerstin. 2020. Attracting the news: Algorithms, platforms, and reframing incidental exposure. *Journalism 21*(8), 1067–1082.

Vaccari, Cristian, Augusto Valeriani, Pablo Barberá, John T. Jost, Jonathan Nagler, and Joshua A. Tucker. 2016. Of echo chambers and contrarian clubs: Exposure to political disagreement among German and Italian users of Twitter. *Social Media + Society 2*(3).

Wickham, Hadley. 2020. *rvest: Easily Harvest (Scrape) Web Pages*.