

How social network sites and other online intermediaries increase exposure to news

Michael Scharkow^{a,1}, Frank Mangold^b, Sebastian Stier^c, and Johannes Breuer^c

^aDepartment of Communication, Johannes Gutenberg University Mainz, 55099 Mainz, Germany; ^bDepartment of Communication, University of Hohenheim, 70599 Stuttgart, Germany; and ^cGESIS Leibniz Institute for the Social Sciences, 50667 Cologne, Germany

Edited by Mary C. Waters, Harvard University, Cambridge, MA, and approved January 9, 2020 (received for review November 4, 2019)

Research has prominently assumed that social media and web portals that aggregate news restrict the diversity of content that users are exposed to by tailoring news diets toward the users' preferences. In our empirical test of this argument, we apply a random-effects within-between model to two large representative datasets of individual web browsing histories. This approach allows us to better encapsulate the effects of social media and other intermediaries on news exposure. We find strong evidence that intermediaries foster more varied online news diets. The results call into question fears about the vanishing potential for incidental news exposure in digital media environments.

news exposure | online media use | web tracking data

People can come across news and other internet offerings in a variety of ways, for example, by visiting their favorite websites, using search engines, or following recommendations from contacts on social media (1). These routes do not necessarily lead people to the same venues. While traditionally considered as an important ingredient of well-functioning democratic societies, getting news as a byproduct of other media-related activities has been assumed to wane in the online sphere. Intermediaries like social networking sites (SNS) and search engines are regarded with particular suspicion, often criticized for fostering news avoidance and selective exposure (2). This assumption has been, perhaps most prominently, ingrained in the “filter bubble” thesis, positing that search and recommendation algorithms bias news diets toward users' preferences and, thus, decrease content diversity (3). On the other hand, incidental news exposure (INE) due to other online activities has received much scholarly attention for several decades (4). Contrary to widely held assumptions, recent INE research found that SNS users have more rather than less diverse news diets than nonusers. For example, one study showed that SNS users consumed almost twice the number of news outlets in the previous week as did nonusers (2). Similar results emerged regarding the use of web aggregators (portals) and search engines, although people may use search engines in a more goal-driven fashion compared to SNS (1).

In previous studies, SNS-based news exposure was typically measured by asking respondents whether they are (unintentionally) exposed to news via social media. Like many survey studies, this approach naturally suffers from the limited accuracy and reliability of self-reports (5). More specifically, recent work has criticized self-report measures for being biased toward active news choices and routine use (6) and being particularly inaccurate when people access news via intermediaries (7). To alleviate these limitations, some studies have used log data to estimate the quantity and quality of online news exposure, for example, in terms of exposure to cross-cutting news (8, 9). However, these studies have focused only on single social media platforms instead of different intermediary routes to news. Other recent studies (1, 10) have traced direct and indirect pathways to online news using browser logs, but have not distinguished nonregular—and therefore possibly incidental—news exposure from regular, typically more intentional or routinized forms of news consumption online. In other words, the question whether

visiting SNS more often (than usual) actually leads to more varied news exposure (than usual) essentially remains unanswered. This problem concerns almost all studies on the use and effects of online media, and has received considerable attention in recent communication research (11). We argue that positive within-person effects of visiting intermediary sites on online news exposure are a necessary (although not sufficient, since even nonregular visits could be intentional) precondition for INE, and, therefore, testing for such effects is a useful endeavor. We address this question using a statistical model that distinguishes between stable between-person differences and within-person effects, that is, the random-effects within-between (REWB) model (12). Investigating within-person effects has additional value by safeguarding causal inferences against bias due to (previously) unmeasured person-level confounders. We apply the REWB model to two large, representative tracking datasets of individual-level browsing behavior in Germany, collected independently in 2012 and 2018. This allows us not only to compare within- and between-person effects but also to analyze possible changes in the effects of SNS (Facebook, Twitter) and intermediaries (Google, web portals) over recent years. Specifically, we investigate their effects on the amount and variety of online news exposure. Using this approach enables us to replicate and extend two recent survey studies (2, 13) that looked at the effects of SNS, web portals, and search engines on 1) overall online news exposure and 2) the diversity of people's online news diets.

Results

News accounted for only a very small portion of total site visits, with an average of 1.51 news visits (SD = 5.61) per respondent and day in 2012. News visits were more frequent in the 2018 sample, with a daily average of 2.51 (SD = 8.41), which is largely in line with previous research (10). Accounting for between-person differences using a logistic regression model with person-level random intercepts yielded an estimated probability for a news visit of 0.009 (99% CI 0.009 to 0.010) in 2012, and 0.014 (99% CI 0.013 to 0.015) in 2018. Following previous studies (1, 10), we estimated the effect of the previously visited site on the probability of a news visit, and found positive and significant effects for all intermediary platforms, most notably Facebook and Google (Fig. 1). However, this clickstream-level analysis likely underestimates the effects of intermediaries, since indirect pathways (e.g., bookmarking, revisiting news later) are ignored, and it does not distinguish regular from nonregular online activities, as mentioned above.

Author contributions: M.S. and F.M. designed research; M.S., F.M., S.S., and J.B. performed research; M.S. and F.M. analyzed data; and M.S., F.M., S.S., and J.B. wrote the paper.

The authors declare no competing interest.

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

Data deposition: Replication data, code, and supplementary materials are available on the Open Science Framework, <https://osf.io/pqd9f/>.

¹To whom correspondence may be addressed. Email: scharkow@uni-mainz.de.

First published January 27, 2020.

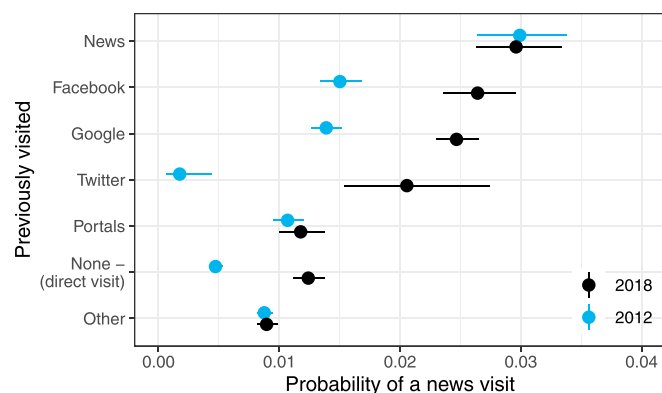


Fig. 1. Probability of news site exposure, conditional on the previously visited site. Estimated marginal probabilities and 99% CIs from two random-effects logistic regression models with random intercepts and slopes for respondents; $n_{2012} = 1,254,490$ site visits, and $n_{2018} = 1,276,519$ site visits.

In a next step, we estimated four REWB panel models using daily aggregated web tracking data. The point estimates and CIs of the REWB models are displayed in Fig. 2. Looking at the within-person effects in the upper part of the figure, there is strong and consistent evidence that visiting SNS more often on a given day increased the chance of being exposed to more news outlets and more news overall. For example, for a respondent in 2012 who visited Twitter twice as often as usual, we would expect her to consume about 41% more news from about 28% more online outlets. This positive effect, however, was significantly smaller in 2018. Using Google and other web portals also had a strong positive effect on the amount and variety of news exposure. Google visits had the largest and most consistent within-person effect overall, while the large positive effects of portals declined somewhat between 2012 and 2018. There was also a positive effect of overall site visits on news exposure.

The REWB models also indicate that much online news exposure is not incidental, but related to stable interindividual differences in browsing behavior. These between-person effects are

displayed in the lower part of Fig. 2, and show that differences in respondents' news diets as well as overall news consumption were strongly associated with the respondents' regular use of Facebook, Google, and other web portals. Respondents who used these sites more often than other respondents were significantly more likely to have a larger and more varied online news diet. The large positive effects for SNS and online intermediaries support previous findings indicating that respondents regularly use these sites to get their news (2).

Overall, our results provide robust evidence for increased news exposure due to SNS and intermediaries on the within-person level. Looking at the differences in effect sizes of the within- and between-person coefficients, we conclude that nonregular intermediary use is not the main route to online news, but that the intraindividual effects are substantial across two different samples and time points. The two platforms often accused of fostering selective exposure (14), Facebook and Google, were consistently associated with a more varied news diet, on both the within- and between-person levels, in addition to increasing the overall amount of news exposure.

Conclusion

We used large-scale observational data to avoid the limited reliability and validity of self-reports on news exposure. Leveraging the potential of such data with the REWB model, our study provides strong evidence that getting more and more-diverse news as a consequence of other media-related activities is a common phenomenon in the online sphere. The findings contradict widely held concerns that social media and web portals specifically contribute to news avoidance and restrict the diversity of news diets. Note that we followed previous studies and measured the variety of news diets by counting the number of outlets visited. Given the overall low frequency of news visits, intermediaries add diversity to the news diets of the large majority of participants with a small news repertoire (2). While we cannot say that outlet variety always equals viewpoint variety, prior research has shown that using a larger number of online news sources typically translates into more-diverse overall news exposure (15). In contrast to previous studies (9, 10), we cannot quantify diversity in terms of cross-cutting exposure, but note that previous

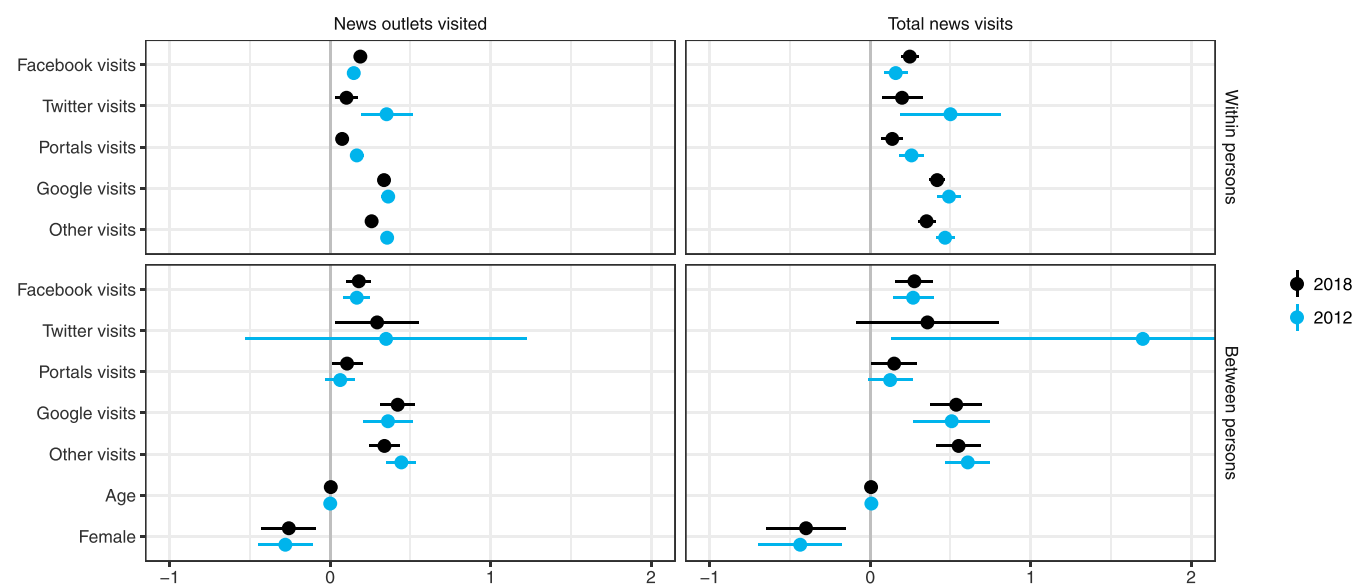


Fig. 2. Relation between browsing behavior and news exposure. Regression coefficients and 99% CIs from Poisson REWB models. Four separate models were estimated for the two outcomes (number of news outlets visited and number of news visits) and the two samples (2012 and 2018). The models include random intercepts and slopes for respondents, random intercepts for days, and OLR to account for overdispersion. All predictors except age and gender were log-transformed before estimation; $n_{2012} = 48,919$ person-days, and $n_{2018} = 40,158$ person-days.

studies have shown little evidence for strong partisan alignments of news audiences in Germany (16) on the outlet level, so that variety would have to be measured on the level of individual news items, which requires URL-level tracking and content analysis data. In addition, future combinations of web tracking with experience sampling surveys are needed to disentangle in what instances nonregular news use is entirely nonintentional and how the respective contents specifically affect the diversity in news diets.

Materials and Methods

Samples. Our analysis is based on data from two representative samples of German internet users aged 14 to 65 y who agreed to use a tracking software that hooks into the web browsers on their desktop computers and/or smartphone. For privacy reasons, respondents could temporarily disable the tracking. The 2012 dataset was originally collected as part of a large, nationally representative household panel (5). It contains the desktop browser logs of $n = 2,970$ respondents in November 2012. The average age of the respondents was $m = 44.6$ y, and 53.1% were female. The 2018 dataset was drawn from an online access panel, with browser logs (desktop and mobile) collected from $n = 2,035$ respondents in December 2018. The average age of the 2018 respondents was a little lower ($m = 41.6$ y), as was the proportion of female respondents (50.6%). In order to simplify comparisons across samples, the datasets were matched in terms of the respondents' age range.

Data collection was carried out by market research companies in accordance with the ICC/ESOMAR International Code of marketing and Social Research Practice (<https://www.esomar.org>). Informed consent was obtained from the participants prior to participation.

Measures. Both the 2012 and 2018 data contain all logged visits from the respondents' web browser on the domain level (e.g., <https://www.spiegel.de>). For the present analysis, we defined a visit with a threshold of 10-s exposure (rather than 1 s or 3 s as in previous studies; ref. 6), in order to establish that respondents had at least somewhat engaged with

a website. We compiled a list of 319 general interest news domains using data from the German Audit Bureau of Circulation as well as a manual inspection of the most frequently visited domains. Similarly, we defined the relevant domains for web portal use (seven sites, such as Yahoo or T-Online), as well as Facebook, Twitter, and Google. For the clickstream analysis, the coded log entries were analyzed. For the REWB model, the number of news visits and number of news outlets visited, as well as Facebook, Twitter, Google and portal visits, were counted per respondent and day, yielding a panel dataset with a maximum of $k = 31$ repeated measurements of browser use.

Data Analysis. In order to account for between- and within-person associations of browsing behavior and news exposure, we estimated an REWB model (12). For a single predictor variable x , i respondents, and t repeated measurements, the REWB model is specified as

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

with β_{1W} as the within-person effect and 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, following current recommendations for conservative fixed effects estimates (12). We expanded the linear REWB model in two regards: Since the outcomes are count variables, we estimated a Poisson rather than a linear model. Moreover, we log-transformed all predictor variables from the web tracking data, because we expected diminishing returns for additional site visits. The models also include random intercepts for days, as well as observation-level random effects (OLRE) to account for overdispersion. Total visits (minus the intermediary sites) were included as a predictor to control for the conflation of the associations between news website use and the use of intermediaries with general internet use intensity.

Data Availability. Replication data and code are available on the Open Science Framework, <https://osf.io/pqd9f/> (17).

ACKNOWLEDGMENTS. M.S. thanks GfK Germany for providing survey and log data from their Media Efficiency Panel. J.B. and S.S. thank Pascal Siegers for his contribution to the collection of the 2018 data. We thank Marko Bachl for helpful comments on the statistical models.

1. J. Möller, R. N. van de Velde, L. Merten, C. Puschmann, Explaining online news engagement based on browsing behavior: Creatures of habit? *Soc. Sci. Comput. Rev.* **10**, 089443931982801 (2019).
2. R. Fletcher, R. K. Nielsen, Are people incidentally exposed to news on social media? A comparative analysis. *New Media Soc.* **20**, 2450–2468 (2018).
3. E. Pariser, *The Filter Bubble: What the Internet Is Hiding from You* (Penguin UK, London, 2011).
4. D. Tewksbury, A. J. Weaver, B. D. Maddex, Accidentally informed: Incidental news exposure on the World Wide Web. *J. Mass Commun. Q.* **78**, 533–554 (2001).
5. M. Scharnow, The accuracy of self-reported internet use—A validation study using client log data. *Commun. Methods Meas.* **10**, 13–27 (2016).
6. H. Taneja, A. X. Wu, S. Edgerly, Rethinking the generational gap in online news use: An infrastructural perspective. *New Media Soc.* **20**, 1792–1812 (2018).
7. A. Kalogeropoulos, R. Fletcher, R. K. Nielsen, News brand attribution in distributed environments: Do people know where they get their news? *New Media Soc.* **21**, 583–601 (2019).
8. E. Bakshy, S. Messing, L. A. Adamic, Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).
9. G. Eady, J. Nagler, A. Guess, J. Zilinsky, J. A. Tucker, How many people live in political bubbles on social media? Evidence from linked survey and Twitter data. *SAGE Open* **9**, 2158244019832705 (2019).
10. A. S. Cardenal, C. Aguilar-Paredes, C. Galais, M. Pérez-Montoro, Digital technologies and selective exposure: How choice and filter bubbles shape news media exposure. *Int. J. Press Politi.* **24**, 465–486 (2019).
11. A. Orben, T. Dienlin, A. K. Przybylski, Social media's enduring effect on adolescent life satisfaction. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10226–10228 (2019).
12. A. Bell, M. Fairbrother, K. Jones, Fixed and random effects models: Making an informed choice. *Qual. Quantity* **53**, 1051–1074 (2019).
13. R. Fletcher, R. K. Nielsen, Automated serendipity: The effect of using search engines on news repertoire balance and diversity. *Digital Journalism* **6**, 976–989 (2018).
14. J. G. Webster, *The Marketplace of Attention: How Audiences Take Shape in a Digital Age* (MIT Press, 2014).
15. F. Mangold, M. Bachl, New news media, new opinion leaders? How political opinion leaders navigate the modern high-choice media environment. *J. Commun.* **68**, 896–919 (2018).
16. M. Bachl, Selective exposure and hostile media perceptions during election campaigns. *Int. J. Public Opin. Res.* **29**, 352–362 (2016).
17. M. Scharnow, F. Mangold, S. Stier, J. Breuer, How social network sites and other online intermediaries increase exposure to news. <https://osf.io/pqd9f/>. Deposited 16 January 2020.