Facebook as an Avenue to News: A Comparison and Validation of Approaches to Identify Facebook Referrals

Felix Schmidt, Frank Mangold, Sebastian Stier & Roberto Ulloa

View supplementary material ☑

Published online: 25 Apr 2024.

Submit your article to this journal ☑

Article views: 1525

View related articles ☑

View Crossmark data ☑

Citing articles: 1 View citing articles ☑

Routledge
Taylor & Francis Group

⌿ OPEN ACCESS   Check for updates

Facebook as an Avenue to News: A Comparison and Validation of Approaches to Identify Facebook Referrals

Felix Schmidt [iD][a], Frank Mangold [iD][a], Sebastian Stier [iD][a,b], and Roberto Ulloa [iD][c]

[a]Department of Computational Social Science, GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany; [b]School of Social Sciences, University of Mannheim, Mannheim, Germany; [c]Cluster "The Politics of Inequality", University of Konstanz, Konstanz, Germany

ABSTRACT

Given that Facebook is still the most widely used social networking site in the world, its influence on democratic processes is under constant scrutiny. Academics have put a special focus on Facebook's role in inhibiting or enhancing citizens' news exposure. Recent studies using digital behavioral data have analyzed the prevalence and effects of "Facebook news referrals." Using a web tracking tool that captures general browsing behavior as well as public posts seen on Facebook, this paper lays the groundwork for the field by assessing the validity of previously proposed operationalizations. We validate news referrals by investigating whether different measures actually reflect exposure to a news URL a user saw on Facebook. We furthermore assess the effects of news referrals on central outcomes in extant literature, contingent on different operationalizations. The results show that the most precise measure of news referrals are click identifiers attached to news URLs by Facebook. Different operationalizations of referrals have theoretically impactful consequences for the substantive understanding of Facebook's role in high-choice online environments. The paper demonstrates the need for academics to constantly innovate in order to measure citizens' online behavior in an ecologically valid manner.

The assumption that social media plays an integral role in news access is pivotal to theories of information exposure in high-choice online environments. While television and newspapers traditionally guaranteed common exposure to news, 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). In order to disseminate their coverage, news producers are increasingly dependent on social media to attract and steer larger fractions of the online audience to their content (Webster, 2014). To comprehensively study contemporary digital media environments, researches need valid conceptualizations and measures of so-called news referrals by social media (Kümpel, 2018; Wojcieszak et al., 2021). Empirical data on who follows URLs embedded in social media to news websites and the prevalence of the phenomenon are especially pertinent for understanding Facebook's role in democratic processes (Guess et al., 2023). After all, Facebook is still the

most widely used social networking site with almost 3 billion users around the globe (Statista, 2022b).

While the study of Facebook news referrals has seen growing popularity in recent years, their investigation has remained a considerable theoretical and methodological challenge. To mitigate the limitations of survey self-reports when it comes to studying fine-grained aspects of digital behavior (Parry et al., 2021), researchers have resorted to web browsing data to study Facebook news referrals (Cardenal et al., 2019; Flaxman et al., 2016; Fletcher, Kalogeropoulos, et al., 2021; Guess, 2021; Guess et al., 2020; Möller et al., 2020; Nelson & Webster, 2017; Scharkow et al., 2020; Stier et al., 2022; Wojcieszak et al., 2021). Nonetheless, these studies have only inferred referrals from indirect and not yet validated proxy measures.

Therefore, the present study addresses the lack of methodological research on Facebook as a pathway to news by providing the first systematic conceptual and empirical assessment of approaches to identify referrals, including their validity as well as their downstream implications for the substantive results in applied research, both in terms of how prevalent news access via Facebook is, the role of Facebook as a distinct pathway to different types of news and the platform's role in reinforcing or counteracting inequalities in news exposure. Critically, our validation integrates an alternative to previously employed approaches that promises more direct accounts of Facebook referrals, the referrer-ID approach. To these ends, our study uses a web tracking data set collected from August to October 2021 in Germany ($N = 739$ participants). The present dataset provides more fine-grained measures thanks to a scraping of the public posts encountered by participants on Facebook in addition to surveys and web browsing histories that have been used in previous research. The results show that operationalizations of Facebook news referrals differ in their validity. We derive methodological recommendations for the study of Facebook news referrals and discuss the theoretical implications of our analysis for Facebook's role in high-choice online environments.

Facebook as a Pathway to News

Extant recent debate has revolved around the Facebook's broader democratic impact as an intermediary between information producers and citizens in digital high-choice media environments. Various popular accounts by politicians, journalists, etc., but also scholars have spotted Facebook and social media more generally as a severe obstacle to a more widely informed electorate – criticized them for inhibiting news use, trapping citizens in enclaves of like-minded speech, catalyzing widespread dissemination of disinformation, and as the reason for other societal ills (e.g., Haidt, 2022; Vaidhyanathan, 2018). At the same time, scholars have prominently called for more differentiated assessments (e.g., Guess et al., 2019; Kümpel, 2018; Thorson & Wells, 2015). In fact, recent studies have often pointed to positive normative implications of Facebook as a facilitator not as an inhibitor of news use (e.g., Fletcher & Nielsen, 2018; Nelson & Webster, 2017; Scharkow et al., 2020). Still, research has continued to disagree in three central regards which will provide an important substantive background for our study: (a) the prevalence of news access via Facebook, (b) Facebook's significance as a pathway to different types of news, and (c) the broader implications of (algorithmic) curation on Facebook for news use gaps in democratic societies and the diversity of citizens' news diets.

One contentious issue in recent research has been how often and many people actually access news via Facebook. Recent survey studies have coined social media as a game changer in a media ecology otherwise increasingly characterized by news avoidance (for reviews, see Kümpel, 2018; van Aelst et al., 2017). Facebook was portrayed as a de-facto front page through which the modal social media user navigates the Internet more broadly and still gets (incidentally) exposed to news by default. Mere encounters of news previews and headlines on social media are arguably of restricted democratic value by themselves. However, if citizens click on links of news articles within Facebook to arrive at the full story on original news sites, there is potential for gains in political knowledge and participation traditionally attributed to newspaper use (Lee & Kim, 2017; Wieland & Kleinen von Königslöw, 2020).

Studies using behavioral data instead of self-report measures of media use have cautioned against overestimating the prevalence of Facebook referrals. Only a minority of news visits follows from visits of Facebook (and social media more generally), with estimates ranging from about 10% in the USA (e.g., Guess, 2021; Guess et al., 2020; Nelson & Webster, 2017; Wojcieszak et al., 2021) to 5 or less percent in European democracies like the UK or Spain (e.g., Cardenal et al., 2019; Fletcher, Kalogeropoulos, et al., 2021; Scharkow et al., 2020). Regular news access via Facebook seems confined to a small fraction of its users, amid a vast majority of users who only very rarely or never arrive at news sites via the platform (e.g., Flaxman et al., 2016; Möller et al., 2020).

A second issue has been to what types of news Facebook steers its users to. One line of research has cautioned that when referred by Facebook, news users more often arrive at untrustworthy or hyperpartisan niche media that disseminate disinformation and so-called alternative (populist) worldviews. Following recent evidence, this extends from the USA (Guess et al., 2020) to Europe (Müller & Bach, 2021), although here visits to hyperpartisan media are generally less prevalent (Stier, Kirkizh, et al., 2020). Another school borrowed the notion of a "persistence of popularity" to contend that Facebook referrals overwhelmingly go to mainstream media (Nelson & Webster, 2017; Taneja et al., 2018; Webster, 2014). By that, news access via Facebook is enshrined in larger winner-takes-all dynamics, cementing the popularity of already popular news providers. Depending on the media system, this can have strikingly different implications. Instead of getting routed to hyperpartisan news, British users might, for instance, primarily get exposed to BBC URLs on Facebook, while Facebook referrals in Germany should be dominated by legacy press outlets, the bedrocks of Germany's online news ecology (Mangold et al., 2022).

The third prominent debate related to our work has revolved around algorithmic and other forms of curation on Facebook (Guess et al., 2023). Scholars have vividly debated whether Facebook cements or even increases existing inequalities in news use or whether it acts as a gateway to news for all strata of the population. As users self-curate their social media feeds and algorithms might amplify these patterns, otherwise more avid news users – typically male, older, politically more interested citizens with high education – might ultimately get exposed to more news (e.g., Kümpel, 2018; Thorson, 2020; Thorson & Wells, 2015). Another recent school has contended that intermediaries do not amplify but, if anything, counteract news use gaps (e.g., Möller et al., 2020; Stier et al., 2022; Taneja et al., 2018). People may routinely refrain from using news yet still access them via social media, where popularity cues shape the perceived relevance of contents (Anspach, 2017; Messing & Westwood, 2014). This

resonates with a growing body of evidence showing that contrary to popular notions of algorithmically driven filter bubbles (Pariser, 2011), Facebook enhances the diversity in citizens' information sources instead of biasing them (Barberá, 2015; Fletcher & Nielsen, 2018; Scharkow et al., 2020). Meanwhile, it has remained less well understood whether this effect merely concerns the use of a larger number of news sources (i.e., diversity in the sense of variety) or also comes with news diets that are more balanced across sources (Fletcher, Kalogeropoulos, et al., 2021; Jürgens & Stark, 2022).

Methodological Challenges in Studying Facebook Referrals

The rise of digital media has not only created manifold options for citizens to create new media routines. It has also presented research with unprecedented methodological challenges in studying news use. Although still prevailing, survey measures have arguably always been affected by over-reporting of news use and social desirability (Prior, 2009). The limited validity and reliability of self-report measures is aggravated in the online sphere, because the relative ease by which people switch among the various options online makes it particularly hard to accurately recall media use (Parry et al., 2021; Scharkow, 2016). Critically, tendencies to misreport apply to situations when people arrive at news via social media (Kalogeropoulos et al., 2019). In fact, social media users are themselves often not even aware of the original sources of the news articles they engage with (Messing & Westwood, 2014; Möller et al., 2020).

To mitigate these concerns, a growing body of research has replaced self-report measures of media use with more direct digital behavioral measures. The most comprehensive way of investigating citizens' web browsing behaviors in natural real-world settings has been offered by studies that have tracked and combined the individual browsing histories of participants with their survey responses (Stier, Breuer, et al., 2020). Still, studies using these data to investigate Facebook's role as a pathway to news have suffered from an important drawback. The used tracking tools only trace the URLs visited and do not capture users' activity and content exposure on Facebook. As a solution, studies have inferred Facebook referrals from various indirect proxy measures (Guess et al., 2020; Möller et al., 2020; Müller & Bach, 2021; Scharkow et al., 2020). However, the validity of these proxy measures has not yet been established, due to the aforementioned measurement issues.

Behavioral studies have used three main approaches to infer Facebook referrals from participants' browsing histories. In the absence of an established terminological framework, we refer to these approaches as the sequence approach, the subpage approach, and the timeframe approach. These approaches can be broadly understood as subvariants of a single underlying time-series approach. Namely, they build on the common assumption that important information about whether news users were referred to a news website is discernible from the timeline of website visits and, more specifically, the temporal ordering of their starting times. In response to calls for more direct operationalizations of referrals (e.g., Chen et al., 2023; Nelson & Webster, 2017; Robertson et al., 2021; Ulloa & Kacperski, 2023), we will additionally validate an approach that has only played a peripheral role in empirical research thus far: the referrer-ID approach. Instead of the time series of website visits, this approach draws on the information contained in the URL query string tags of news visits that originate from clicks on links within Facebook (signified by the presence of the parameter fbclid in a news URL). These identifiers were introduced by Facebook in 2018 and – as a practical consequence – only after the data used in some previous studies had been collected.

Table 1. Timeline-based approaches for tracking Facebook referrals.

Approach	Operationalization	Previous Studies
Sequence	Facebook is domain whose visit was started immediately prior to news domain visit	Cardenal et al. (2019)
		Fletcher, Kalogeropoulos, et al. (2021)
		Kalogeropoulos et al. (2019)
		Müller and Bach (2021)
		Nelson and Webster (2017)
		Scharkow et al. (2020)
		Stier et al. (2022)
		Wojcieszak et al. (2021)
Subpage	Facebook is domain whose visit was started immediately prior to news domain visit which is not a visit of the domain's main page	Flaxman et al. (2016)
		Möller et al. (2020)
Timeframe	Facebook is among three domains visited within 30 seconds prior to news domain visit	Guess (2021)
		Guess et al. (2020)

Table 1 gives an overview of the established timeline-based approaches to inferring referrals, including prior papers employing each one. The most-widely used approach in previous research, which sticks most closely to the sequential ordering of website visits' starting times, is the sequence approach. Specifically, it generally counts a news visit as a Facebook referral if Facebook is the domain whose starting time directly precedes the news visit's starting time. To illustrate, Table 2 shows the browsing sessions of three hypothetical study participants (i.e., Ann, Michael, and Sarah) visiting the same domains, including both Facebook and a news domain (i.e., cnn.com). Since the starting times of Sarah's Ebay and Amazon visits fall in between the starting times of her Facebook and CNN visits, the sequence approach would only count Ann and Michael's CNN visits as Facebook referrals. The subpage approach builds on the sequence approach yet incorporates more information than just the sequential ordering of website visits. Specifically, it categorizes a news visit as being referred if Facebook is not only the domain whose visit was started directly prior to the news visit. In addition, the respective news visit can also not be a visit of the news domain's main ("index") page but a subpage. In our hypothetical example, the

Table 2. Hypothetical examples of browsing sessions.

	Starting Time	Domain	Full-URL	Classification of CNN Visit		
				Sequence	Subpage	Timeframe
Ann	09:35:32	amazon.com	amazon.com/a-product			
	09:35:42	ebay.com	ebay.com/a-product			
	09:35:52	facebook.com	facebook.com/a-string			
	09:35:59	**cnn.com**	cnn.com/a-news-story	Referred	Referred	Referred
Michael	09:35:32	amazon.com	amazon.com/a-product			
	09:35:42	ebay.com	ebay.com/a-product			
	09:35:52	facebook.com	facebook.com/a-string			
	09:35:59	**cnn.com**	**cnn.com (Main page)**	Referred	Not referred	Referred
Sarah	09:35:32	facebook.com	facebook.com/a-string			
	09:35:42	amazon.com	amazon.com/a-product			
	09:35:52	ebay.com	ebay.com/a-product			
	09:35:59	**cnn.com**	cnn.com/a-news-story	Not referred	Not referred	Referred

News visit (to cnn.com) in bold.

subpage approach would categorize Ann's CNN visit as being referred but not the one of Michael because Michael logged on the main page of CNN. This illustrates that the Facebook referrals identified by the subpage approach are, by definition, a fraction of those identified by the sequence approach. This stands in contrast to the timeframe approach that goes more than one step back in the timeline by categorizing a news visit as being referred by Facebook if the platform is among the three domains visited within 30 seconds prior to the news visit's starting time. In our example, the timeframe approach classifies Ann, Michael, and Sarah's CNN visits as being referred by Facebook.

The subpage approach promises to enhance the measurement precision compared to the sequence and timeframe approaches, at least when we assume that most of the time Facebook users will get referred to specific articles and not the main page of news website. Yet this approach requires full URL information that is not included in all web tracking datasets, whereas the sequence and timeframe approaches can be implemented solely based on domain-level data. Despite some individual pros and cons, the sequence, subpage, and timeframe approaches all suffer from the same two drawbacks:

(1) Facebook referrals presuppose that a user interacted with Facebook prior to a news visit. Yet, this does not necessarily imply subsequent starting times of her Facebook and news visits, because people may use multiple browser tabs and surf the web in a non-linear fashion. In such instances, the starting time of the Facebook visit from which a news visit was referred may actually date back hundreds of entries in the time series of website visits' starting times (Robertson et al., 2021; Wojcieszak et al., 2021). This happens to be the case when, for instance, a user starts a Facebook visit during early morning hours, subsequently uses other browser tabs, and eventually returns to and clicks on links of news articles within Facebook during lunch break.

(2) Even in instances in which people actually interacted with Facebook before visiting news sites, their news visits do not necessarily originate from Facebook as a referrer. Instead of having clicked on links of news articles within Facebook, people may have, for example, simply navigated to their favorite news website out of routine or visited a bookmarked article (Flaxman et al., 2016; Möller et al., 2020).

One promising alternative to identifying referrals that has been brought up in recent literature is the referrer-ID approach (e.g., Chen et al., 2023; Wojcieszak et al., 2021). This approach does not rely on indirect timeline-based proxies but directly takes into account the marketing information that is amended to the URLs originating from clicks on links on Facebook and other online platforms (like Google or Twitter). This bears the major advantages that measures of referrals are not confined to linear browsing behaviors and that there is no longer the need to conceptually distinguish between visits of news domains' main pages and subpages. Meanwhile, the referrer-ID approach has, to the best of our knowledge, never been validated nor fully exploited in an academic study. Most closely related to our study, Wojcieszak et al. (2021) considered the Referred-ID but only used it as supplementary approach to identify additional referrals, above and beyond those already identified by the sequence approach. They took into account the marketing information embedded by "utm_source"-strings (e.g., https://www.usatoday.com/article-title/?utm_-source=facebook&utm_medium=social). We more directly concentrate on the Facebook click-identifiers (fbclid) that are generated by Facebook itself (e.g., https://www.usatoday.

com/article-title/?fbclid=IwAR13t53iHz19...). The referrer-ID approach comes with the principal advantage that it follows an opt-out rather than opt-in logic. While news providers have to activate the generation of utm_source-string, fbclids are generated by default – leading to a much higher inclusiveness of the fbclid approach.

Research Questions

Overall, we identified four main approaches to tracking Facebook referrals to news: the sequence approach, the subpage approach, the timeframe approach, and – as an alternative to traditional timeline-based measures – the referrer-ID approach. Various considerations may drive researchers' choices among these approaches, including matters of data availability. Yet the single most important – so far unresolved – question from a methodological viewpoint is how validly the approaches reflect the target concept of interest: news visits that follow from encounters and clicks of the respective articles within Facebook. Considering the previous evidence by Wojcieszak et al. (2021) that merely going back more than one step in the timeline of website visits diminishes rather than enhances the measurement validity, the timeframe approach will likely yield less valid measures of Facebook referrals than the sequence approach. Since it is a plausible assumption that visits of subpages (which tend to have longer URLs) are specifically likely to result from clicks on links (Flaxman et al., 2016; Möller et al., 2020), the subpage approach should – in turn – be more valid than the sequence approach that also incorporates main pages of news websites. Still, the referrer-ID promises the most valid measurements as it captures linear and non-linear browsing behaviors and is agnostic to the type of URL (main page, subpage). To shed more light on these guiding propositions, our most central research question asks:

RQ1. How validly do different approaches capture news visits that originate from exposure to news URLs on Facebook?

Besides questions of measurement validity, how the approaches impact results in substantive applications is another important matter. If scholars aim at safeguarding their inferences, it is essential to establish the robustness of results across approaches. When different findings based on variations in research methods are published, it is crucial to explicitly pit alternative choices against each other using the same data, and establish their consequences systematically. It otherwise remains speculative whether studies arrive at different conclusions because, for instance, people have changed their Facebook and news use routines or simply because the studies rest on incompatible approaches. Therefore, the second main goal of our study corresponds to what has also been called a "multiverse" analysis in recent methodological research (Mangold & Scharkow, 2020; Steegen et al., 2016). That is, we examine whether and in what specific regards the varied validity of the different approaches to measuring referrals makes a difference for the substantive understanding of Facebook as a pathway to news. Despite, or precisely because of, previous behavioral studies that used timeline-based approaches and cautioned against overestimating the prevalence of referrals, it has not only remained contentious how often and how many people are referred to news sites by Facebook. As outlined in greater detail above, it has also remained controversial whether Facebook foremost routes its users to already popular news outlets or specifically inclines them to visit hyperpartisan news

outlets. Yet another vital line of ongoing scholarly debate has revolved around the role of curation processes in shaping inequalities in news exposure. While some researchers have argued that Facebook reinforces preexisting gaps in news exposure by fortifying its associations with demographic and political characteristics, others have seen reason for the reverse mechanism as Facebook may actually enhance the news diet diversity of its users. Our research questions for investigating the substantive consequences of methodological choices therefore cover the most prominent research foci in the literature:

RQ2. Do different approaches lead to different results on the prevalence of Facebook referrals to news?

RQ3. Do different approaches lead to different results on the distribution of Facebook referrals across types of news outlets?

RQ4. Do different approaches lead to different results on the associations of demographic and political characteristics with Facebook referrals to news?

RQ5. Do different approaches lead to different results on the associations of Facebook referrals with news diet diversity?

Methods

Our study rests on data collected as part of a larger research project on media effects during the 2021 German federal election. We used a web tracking tool programmed by academics (Aigenseer et al., 2019) that enables both more granular accounts of users' browsing sequences and data on the public contents that users encounter on Facebook. The data collection was reviewed and approved by the GESIS Ethics Committee (decision 2021–6).

Country Case Selection

The German media system has undergone significant developments observable around the world in recent years. Growing parts of media use have been shuffled from the offline to the online sphere where social media have become increasingly popular (Newman et al., 2022). 93% of Germans are Internet users. About 60% of German Internet users are Facebook users (Statista, 2022a), which is somewhat less than in, e.g., the USA (with about 75%; Statista, 2023). The German media system has a (continued) history in news use rooted in political interest rather than partisanship (Mangold et al., 2017). Ideological selective exposure plays a relatively restricted role in Germany (Fletcher et al., 2020). In line with most other democracies, only a small proportion of less than 5% of German news users live in ideological online echo chambers (Fletcher, Robertson, et al., 2021). At the same time, cross-national comparative audience research has highlighted that Germany is one of the countries where non-use of online news is most pronounced. About 30% of German Internet users only rarely to virtually never visit news websites (Fletcher et al., 2020).

Participants

Our analysis is based on 739 participants recruited from the respondent pool of *dynata*, a market research company. The study participants were invited based on German population margins, but some cells could not be fully filled due to nonparticipation. As a consequence, the sample is underrepresented in terms of lower educated, younger and older people (see Online Appendix Table S1 for the demographic composition of the sample). Participants gave their informed consent and were incentivized to install a plugin in their main browser on their desktop computer. The data contain all logged website visits from the participants' web browser between 19 August 2021 and 27 October 2021, including the visits' domain, full-URL, contents as well as the order in which participants looked at websites – above and beyond the standard information of their visits' starting times. A wide range of sensitive domains were excluded from the tracking, and participants could temporarily pause the tracking tool at any time. The median number of active days was 61 out of a maximum of 69. The data comprise a total of 8,356,462 website visits. 60.2% of participants accessed Facebook at least once. Facebook visits accounted for 1.5% of total website visits. The average number of Facebook visits per participant and day was 3.4 ($SD = 11.9$).

Surveys

Our surveys of the same participants' captured the most relevant demographic and political covariates of news exposure. Standard demographics included gender (46.3% female participants), age (in years; $M = 48.1$, $SD = 13.8$, $MIN = 18$, $MAX = 74$), and education (coded into a 3-point measure following the country-comparative ISCED scheme; 1 = primary/lower secondary [low], 2 = upper secondary [medium], 3 = tertiary [high]; $M = 2.4$, $SD = 0.7$, $MIN = 1$, $MAX = 3$). Political interest was captured using a 5-point scale from 1 (= not at all interested) to 5 (= very interested; $M = 3.7$, $SD = 0.9$, $MIN = 1$, $MAX = 5$). Political ideology was operationalized with an 11-point scale from 1 (left) to 11 (right; $M = 5.5$, $SD = 1.9$, $MIN = 1$, $MAX = 11$). An additional measure of political extremity with range from 0 (= low extremity) to 5 (= high extremity) was calculated as the distance from the political ideology scale's midpoint ($M = 1.4$, $SD = 1.3$, $MIN = 0$, $MAX = 5$). All covariates used within the present investigation were captured in the baseline survey prior to the collection of the web tracking data. Online Appendix Table S2 shows the descriptive statistics for all used variables.

Identification and Categorization of News Visits

To identify news visits in study participants' website visits, we compiled a list of 742 news websites. 309 of them were visited at least once by participants. The news site list was generated by synthesizing lists used in several prior web tracking studies (Scharkow et al., 2020; Stier, Kirkizh, et al., 2020) and an additional manual inspection of the top 1,000 visited domains in our data, ensuring that the most relevant news sites were covered and adequately categorized. In line with other recent studies (e.g., Cardenal et al., 2019; Scharkow et al., 2020; Stier et al., 2022), news visits accounted for only a small proportion of total website visits (1.24%), with an average number of 2.8 ($SD = 7.2$) news visits per participant and day. 17% of participants did not have a single news visit during the two-month period.

Besides identifying news websites, we needed to categorize these in order to speak to previous literature. Given the relatively restricted role of ideological selective exposure in

Germany, we refrained from inferring the ideological leanings of news outlets but consider the most relevant variation in German news coverage instead: the news type. To this end, we followed Stier, Kirkizh, et al. (2020) by grouping the individual news websites into six categories: legacy press, public broadcasting, tabloid press, commercial broadcasting, digital-born outlets, and hyperpartisan news. Beyond the number of news outlets visited, we will use the news types as another feature to characterize the diversity of news diets, following the general understanding of diversity as an over-arching twofold concept (Fletcher, Kalogeropoulos, et al., 2021). News diet variety was calculated by counting the number of news outlets ($M = 10.8$, $SD = 12.0$) and news types ($M = 3.2$, $SD = 1.5$) visited by each participant (see also, e.g., Fletcher & Nielsen, 2018; Scharkow et al., 2020). The balance of participants' news diets across news outlets ($M = 1.3$; $SD = 0.8$) and news types ($M = 0.6$; $SD = 0.4$) was calculated using Shannon's H (see also Fletcher, Kalogeropoulos, et al., 2021; Jürgens & Stark, 2022).

Behavioral Benchmarks for Validating Facebook Referrals

We utilize a web tracking tool that combines two unique advantages over previously used instruments (Aigenseer et al., 2019). First, the web tracking tool captures tab activity, as opposed to simply the sequence of website visits' starting times. In essence, this means that we can determine whether Facebook was – as a necessary precondition for referrals – actually the domain that a user interacted with prior to a news visit, irrespective of whether the tracked starting time of the Facebook visit directly precedes the news visit's starting time or whether it dates back further (because participants show non-linear browsing behavior). Second, the tool immediately captures the content of the webpages visited by users (in addition to their visited domains and URLs). This comes with the principal advantage that we can crack open the central black box in previous web tracking studies: the (news) contents encountered by participants on social media (for technical backgrounds, see Christner et al., 2021). Meanwhile, ethical and legal reasons required us to capture only exposure to public posts on Facebook. For our methodological goals, this means that whereas our design allows to establish which news visits reflect Facebook referrals, we cannot conversely rule out for other visits that they do not in fact reflect clicks on links of news articles encountered on Facebook because we lack insights in private posts and messages on the platform.

Results

Validation of Facebook Referrals to News

RQ1 asked how validly different approaches capture news visits that originate from exposure to a news URL on Facebook. We derive our answer in two steps that build on the conditions for Facebook referrals outlined in the literature review. First, with the capabilities of our tool we examined for how many of the referrals identified by each approach Facebook was actually the domain that participants engaged with before visiting a news article, as opposed to simply the domain whose starting time preceded the news visit's starting time. Second, we examined in how many of these instances participants had encountered the visited news URL on their Facebook feed. It is important to bear in mind

that we were only capturing public posts on Facebook. This does not impact the relative differences between approaches yet means that we generally cannot expect to find all news URLs in the HTMLs of participants' Facebook visits.

Figure 1, Panel A shows that the biggest difference can be identified between the time-frame approach and the remainder of approaches. Facebook was the previously active domain for less than 60% of the referrals identified by the timeframe approach, as opposed to about 90% of those identified by the referrer-ID approach and about 80% and 86% of those identified by the sequence and subpage approaches. Strikingly, the results of the second analysis step in Panel B show a much higher validity of the referrer-ID approach compared with the other approaches. Almost 90% of the referrals identified by the referrer-ID approach could be verified as instances in which participants had encountered the respective news URL during the preceding Facebook visit, as opposed to less than 70% with the subpage approach and less than 60% with the sequence and timeframe approaches. Overall, the results supported the benefits of the referrer-ID approach over the timeline-based approaches, especially due to its capability to more validly capture instances in which users do not only look at their Facebook feed before visiting a news article but actually get exposed to the very news article on Facebook.

With this answer to RQ1 in mind, we investigate whether and how the approaches' differing validity impacts the substantive conclusions in applied research of Facebook as a pathway to news (RQ2 to RQ5).
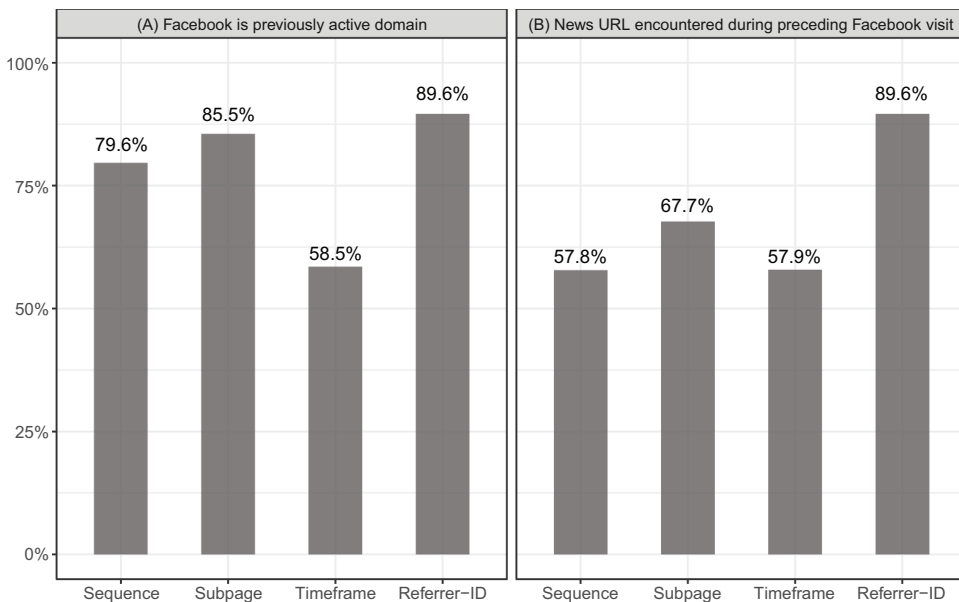

Figure 1. Benchmarks for validating facebook referrals to news. The bars represent the shares of referrals identified by each approach for which (A) Facebook was the previously active domain (i.e., the domain participants interacted with prior to a news visit) and (B) participants encountered the visited news URL during the preceding Facebook visit (within public posts). With exception of the shares for the sequence and timeframe approaches in Panel B, the estimates within each panel were statistically different throughout, $p < .05$. Statistical significance was tested using Derrick et al.'s test for comparing proportions with partially overlapping samples (taking into account that the approaches categorized partially the same and partially different news visits as referrals).

Prevalence of Facebook Referrals to News

In light of scholarly debates about how often and how many people access news via Facebook, RQ2 asked whether different approaches lead to substantially the same or different results regarding the prevalence of Facebook referrals to news sites. We first calculated the share of total news visits that each approach categorized as Facebook referrals. While the share was highest with the timeframe approach (4.4%) – followed by the sequence and referrer-ID approaches (3.3% each) – and lowest with the subpage approach (2.6%), all approaches shared the principal understanding that only a small fraction of less than 5% of news visits originated from Facebook as a referrer.

Figure 2 demonstrates that the approaches more strongly disagreed about how many Facebook users (i.e., participants who visited the platform at least once) were referred to news.[1] According to the timeframe approach, approximately 40% of the Facebook users in our sample were referred to news sites during the two-month period covered by our study. This remained confined to only about one-fourth of Facebook users according to the referrer-ID approach. The sequence approach and – to a lesser extent – the subpage approach also yielded higher overall shares of referred Facebook users than the referrer-ID approach. The lower overall share of
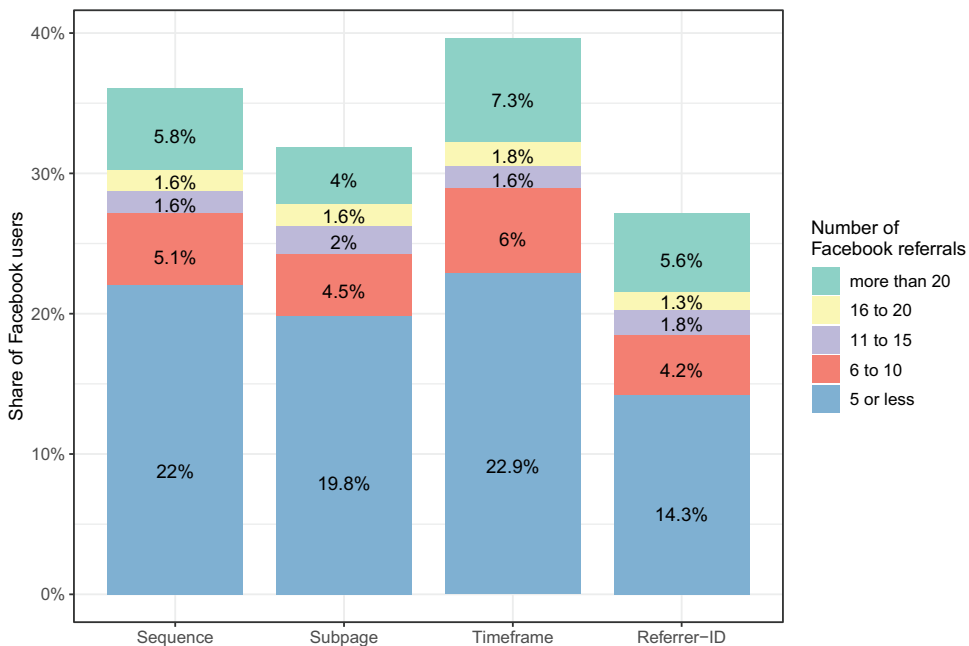

Figure 2. Share of Facebook Users Referred to News Sites. Facebook users were operationally defined as participants with at least a single Facebook visit (60.2% of participants; see also "Methods"). The overall height of the bars represents the overall share of Facebook users with at least a single referral to news websites identified by each approach. The colored stacks differentiate the overall shares based on the number of referrals per individual. Both the overall share of referred Facebook users ($X^2 \geq 7.0$, $df = 1$, $p < .05$) and the share of Facebook users with 5 or less referrals ($X^2 \geq 8.3$, $df = 1$, $p < .05$) were significantly smaller with the referrer-ID approach than the other approaches. Statistical significance was tested using McNemar X^2-test for comparing proportions with dependent samples (taking into account that the estimates reflect alternative categorizations of the same Facebook users).

referred Facebook users identified by the referrer-ID approach was specifically due to the identification of a smaller fraction of Facebook users with a relatively low number of five or less referrals to news sites. The results for RQ2 show that timeline-based approaches overestimate the prevalence of news access via Facebook, though not necessarily in terms of the share of referred news visits but more crucially in terms of how many social media users are referred to news websites.

Distribution of Facebook Referrals Across News Types

Recent research has highlighted the role of Facebook as a distinct pathway to hyperpartisan (niche) media. Other research followed the notion of a "persistence of popularity" to argue that Facebook reinforces the popularity of already popular news media such as in the German case the legacy press. In light of this duality, RQ3 asked whether different approaches lead to substantially the same or different results on the distribution of Facebook referrals across news types. Figure 3 differentiates the referrals identified by each approach in terms of the share per news type. For instance, if we look at the estimates for the sequence approach, we see that about 60% of the referrals identified by the approach were visits to the legacy press, 17% of the referrals identified visits to the tabloid press, and so forth. As a reference, we added the distribution of (other) news visits that were not referred by Facebook. This comparison demonstrates the specifics of Facebook news referrals or, alternatively speaking, whether visits to a given news type were distinctively more likely when users were referred by Facebook. To illustrate, we see that while the legacy press accounted for about 40% of non-referred news visits, about 60% of the referrals identified by all four referral approaches went to the legacy press. Thus, news use routed by Facebook was by 20% points more likely to benefit the legacy press compared with other pathways.

More generally, the results show three main patterns. (1) All approaches agreed that the vast majority of Facebook referrals led participants to arrive at the domains of established media organizations. (2) All approaches agreed that visits to not only the legacy press but also hyperpartisan news were specifically more likely when users were referred by Facebook. (3) The probability of a Facebook news referral being a visit to hyperpartisan news was almost twice as high with the referrer-ID approach (6%) than with the other approaches (3%). Overall, while all approaches yielded evidence in favor of both the persistence of popularity and the relevance of Facebook as a distinct pathway to hyperpartisan media, the referrer-ID approach empirically reflected both theoretical streams of thought most closely.

Demographic and Political Correlates of Facebook Referrals to News

Theoretical accounts have put forth the argument that curation mechanisms on Facebook should reinforce the preexisting associations of news exposure with demographic and political characteristics. Yet other research has identified null effects or even mitigated inequalities in news exposure thanks to Facebook use. Therefore, RQ4 asked whether different approaches lead to substantially the same or different results
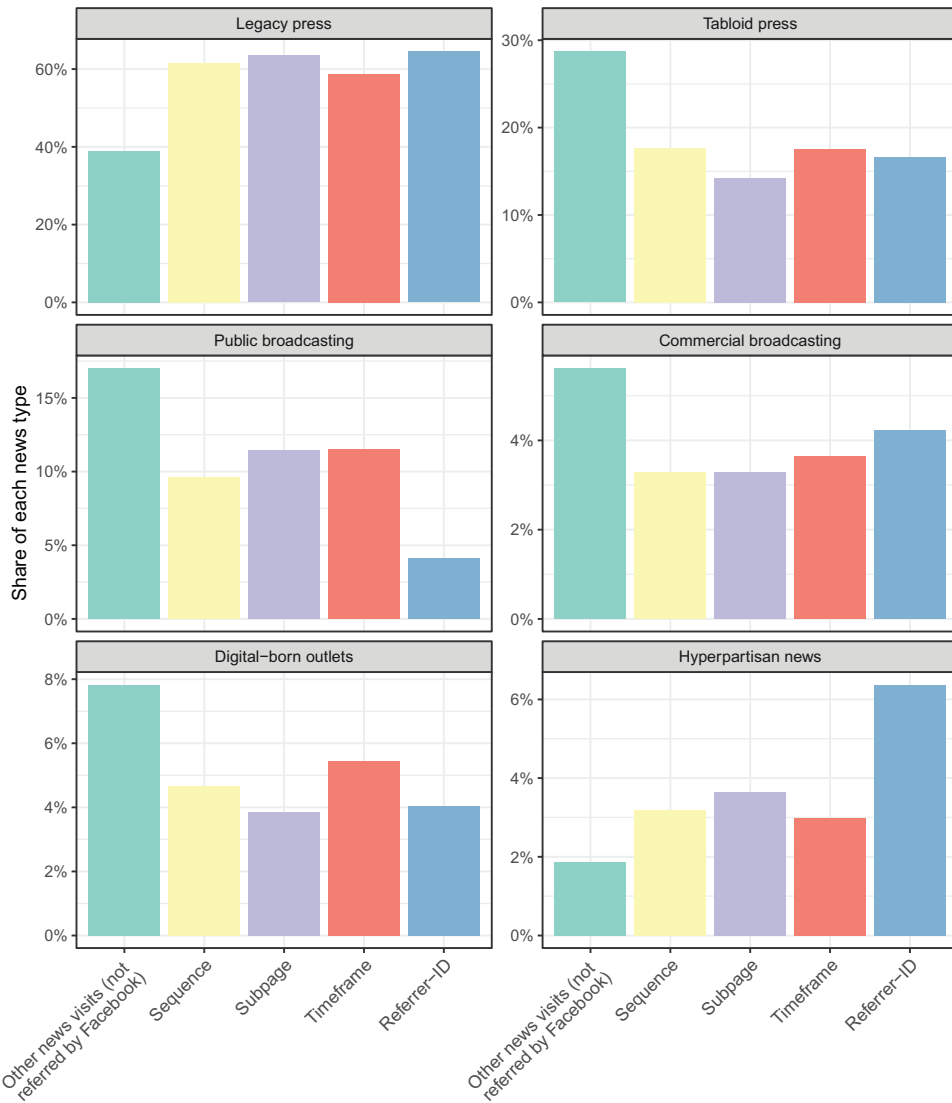
Figure 3. Distribution of Facebook referrals across news types. The figure breaks down the Facebook referrals identified by each approach in terms of the proportion of referrals to each of the six news types. Other news visits were included as a reference and operationally defined as visits that – all four approaches to identifying Facebook referrals agreed – were not referred by Facebook. The shares of Facebook referrals to the legacy press ($X^2 \geq 445.8$, $df = 1$, $p < .05$) and to hyperpartisan news ($X^2 \geq 16.6$, $df = 1$, $p < .05$) were significantly higher for Facebook referrals than other news visits with all approaches.

regarding the associations of Facebook referrals with demographic and political characteristics.

The results in Figure 4 speak for similar associations across different strata of the population. All approaches shared an understanding that in contrast to common notions of selective exposure, neither being male or older nor high levels of education or political
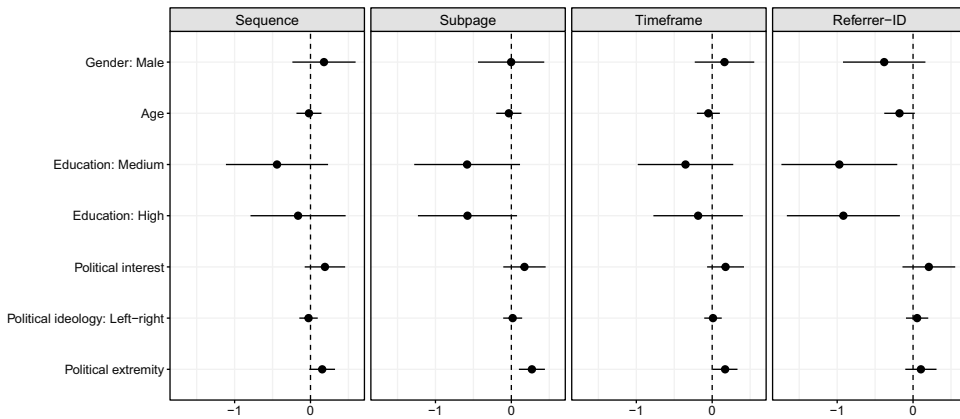
Figure 4. Associations of demographic and political characteristics with facebook referrals to news. Regression coefficients and 95% confidence intervals from negative binomial models, where the criterion variables are the number of Facebook referrals per approach (see Online Appendix S1 for results across news types). Total number of website visits and total number of Facebook visits controlled but not reported. Reference category is "Education: Low." Age was divided by 10 before the estimation to improve interpretation.

interest were positive predictors of Facebook referrals.[2] This pattern extended from news in general (Figure 4) to most of the six news types (see Online Appendix Figure S1).

Meanwhile, the approaches disagreed about the extent to which Facebook effectively counteracts news use gaps, such that the referrer-ID approach yielded the clearest association of Facebook news referrals with low education (the reference category). More generally, the distinctiveness of the referrer-ID approach was also supported by a direct comparison between Facebook referrals and other news visits: the referrer-ID approach was the only approach that yielded associations of age and education with Facebook referrals that were, by themselves, significantly weaker than these correlates' associations with news visits not referred by Facebook (see Online Appendix Figure S2). Overall, while all approaches agreed that Facebook does not aggravate gaps in news exposure, the referrer-ID approach most strongly supported that Facebook counteracts such gaps.

Facebook Referrals and News Diet Diversity

Another main focus in the literature is whether Facebook fosters or restricts the diversity of news diets and, more specifically, their variety and/or balance. In this vein, RQ5 asked whether different approaches lead to substantially the same or different results regarding the associations of Facebook referrals with news diet diversity. We operationalized the variety and balance of news diets both on the level of the individual news outlets and distinct news types, leading to a total of four outcome measures and thereby accounting for both the number of news outlets and news types visited by participants (i.e., variety) and how equally their visits are distributed across news outlets and news types (i.e., balance).

Figure 5 summarizes the (partial) associations obtained when regressing these measures on the Facebook referrals identified by each approach, controlling for demographics, political characteristics, as well as for the total number of website visits and Facebook visits
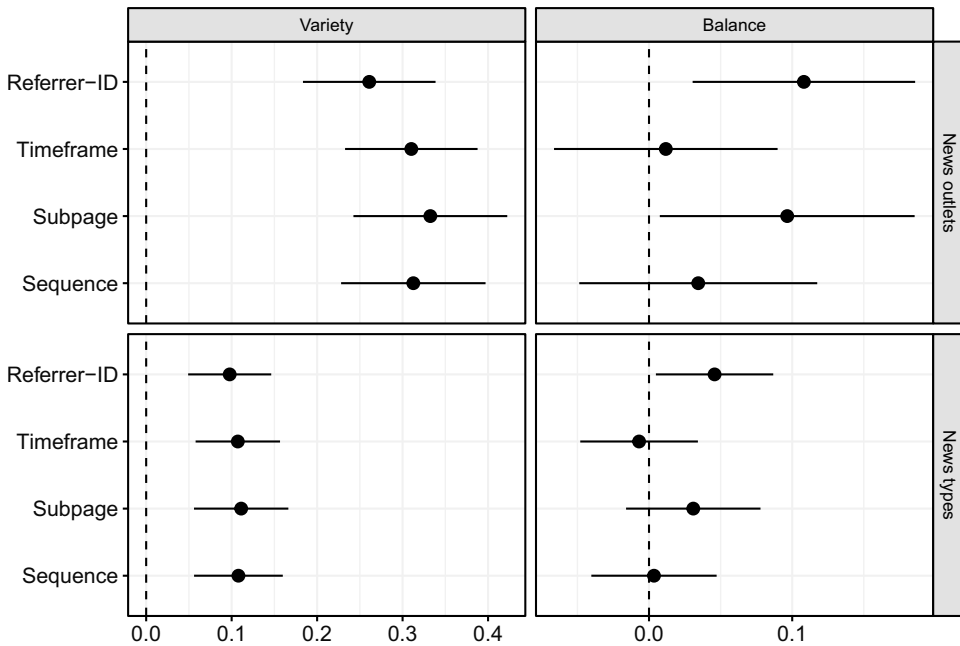
Figure 5. Facebook Referrals and News Diversity. Regression coefficients and 95% confidence intervals from multivariate models, where the focal predictor variable is the number of Facebook referrals and the criterion variables are the diversity measures differentiated by the panels. Each panel contains the associations of Facebook referrals with the respective diversity measure obtained from the four approaches. Gender, age, education, political interest, political ideology, political extremity, Facebook visits in general, and total visits included as controls but not reported. Associations of referrals with the variety measures were estimated using negative binomial models to account for the dependent count variables; associations with the balance measures using linear models (see also Jürgens & Stark, 2022).

in general. Looking at the left-hand panels, we see that despite some variation, all approaches shared the understanding that news access via Facebook corresponds to more varied news diets, both on the level of news outlets and news types. The estimates in the right-hand panels show that the approaches disagreed more substantially about whether news access via Facebook also translates into more balanced news diets. While the referrer-ID approach yielded significantly positive associations of Facebook referrals with news diet balance on both the outlet and news type level, the associations were non-significant with the sequence and timeframe approaches throughout. The subpage approach yielded a significantly positive association of Facebook referrals with news diet balance on the news outlet level but not on the news type level.[3] Overall, all approaches supported that news access via Facebook does not limit news diet diversity yet disagreed about the extent to which Facebook fosters news diet diversity. The referrer-ID approach yielded the most consistent empirical support for positive effects of news access via Facebook on news diet diversity, most distinctively compared to the sequence and timeframe approaches.

Discussion

While much public and scholarly debate is concerned with Facebook's role in contemporary online media ecologies, the methodological underpinnings of academic research on the topic have remained understudied. This paper provided the first systematic assessment of behavioral measures for identifying Facebook referrals to news, including both their validity and substantive implications. To this end, we used data collected in Germany that combined the individual browsing histories and survey responses of participants with a scraping of public posts on Facebook. In the following, we outline the methodological and substantive conclusions that can be derived from our research.

Methodological Conclusions

The first core methodological finding is that we observe profound differences in the validity of alternative approaches to measuring referrals (RQ1). The referrer-ID most validly captured our benchmark instances in which participants had not only interacted with Facebook prior to visiting a news article but had actually encountered the very article on their Facebook feed, followed by – though with quite some distance – the subpage approach. This result resonates with calls for more direct operationalizations of referrals instead of using proxy measures (e.g., Chen et al., 2023; Wojcieszak et al., 2021). Simply the circumstance that a Facebook visit's starting time precedes the starting time of a news visit does not sufficiently establish a referral. The second methodological finding of our study comes from its multiverse analysis (RQ2 to RQ5). We probed the substantive implications of operationalizing referrals in either one way or the other to find the same basic pattern across various core applications in previous research: while the approaches did not produce downright contradictory substantive results, the theoretical implications for Facebook's role in high-choice online media environments differ in some respects depending on the chosen approach (see more details further below).

In conjunction, we recommend that researchers using browsing data for constructing referrals carefully consider the substantive implications of each approach. If the data allows, researchers should use the approach with the highest validity, the referrer-ID approach, though this approach is not entirely free of error and should therefore not be mistaken with a ground truth. Given that none of the approaches radically called into question the bigger substantive picture, researchers may still consider timeline-based approaches if only less detailed data are available but should clearly acknowledge their restricted validity. We recommend employing the sequence approach in studies working with web tracking data that only include the starting time and domain information, instead of the more liberal timeframe approach that introduces more false positives. The subpage approach still serves as a valuable supplement to the sequence approach, especially when systematic platform referrer identifiers are not available, unlike in the case of Facebook. Reassuringly, the analysis also revealed that when Facebook users followed links of news articles, up to 90% of the resulting visits could be traced back to public posts, providing support for the previously untested assumption that news is mostly shared publicly on Facebook (Breuer et al., 2022). For academic researchers who need to get insights into Facebook content but should not scrape private posts on Facebook due to legal and ethical prohibitions, it is essential that the vast majority of news content on Facebook stems from public posts.

Substantive Conclusions

The results of our analysis also have implications for the theoretical understanding of online media environments. Introducing the methodologically superior referrer-ID approach did not fundamentally alter the overall picture drawn by previous research on news access via social media. Instead, several of the theoretical inferences made in prior studies received stronger support when using the referrer-ID approach. Specifically, the substantive results of our analysis speak to three theoretical streams in the literature.

First, in terms of the prevalence and distribution of news referrals (Flaxman et al., 2016; Möller et al., 2020), the four approaches show that the practice is moderately prevalent. Approximately 25% to 40% of the Facebook users in our sample had at least one Facebook referral to a news website, with the true number being closer to the lower bound represented by the referrer-ID approach. There is also agreement that for around 10% of Facebook users who had more than 10 referrals during our two-month research period, the social network is a central gateway to news. Taken together, it is fair to say that a rather narrow fraction of Internet users is relying on Facebook for the (pre)selection of news URLs. Of course, in addition to referrals, research on news exposure within social media – a black-box in online audience research so far – would also benefit from more precise behavioral measures. Only a comprehensive, theoretically and empirically grounded perspective on exposure within and outside of platforms can fully elucidate the role of social media in contemporary digital media environments.

Second, researchers investigated to which types of news content Facebook guides its users. Independent of the chosen approach, the distribution of news types across referred news visits differed widely from other news visits. Especially visits to legacy press outlets were much more prevalent when news users were referred – at least in the German media system. Besides the dominance of quality information sources, the referrer-ID most clearly supported previous findings that Facebook is a relevant gateway to hyperpartisan news. In fact, the estimates inferred from previously applied timeline-based approaches (Guess et al., 2020; Müller & Bach, 2021) uniformly underestimated the scale of the phenomenon. Whereas concerns that Facebook predominantly routes its users to untrustworthy niche websites that disseminate alternative worldviews and disinformation have certainly often been exaggerated, referrals to hyperpartisan websites from social media nevertheless matter and deserve further scholarly attention.

The third stream of research to which we contribute is the role of social media in exacerbating or reducing inequalities in news use. Especially when enhancing the measurement validity through the referrer-ID approach, the results supported that in contrast to concerns about negative effects of algorithmic and other forms of curation, Facebook news referrals enhance the diversity of citizens' news diets and even mitigate news use gaps. Both findings were compatible with, but more pronounced than previous results using timeline-based approaches (e.g., Fletcher, Kalogeropoulos, et al., 2021; Stier et al., 2022). The substantive analysis contextualized the meager prevalence of Facebook news referrals. Simply because referrals are less widespread according to the referrer-ID approach, they are not necessarily less important from a democratic point of view. Instead, the introduction of more valid measures yielded that Facebook fosters at least some news exposure among users otherwise inclined to refrain from online news.

Limitations and Directions for Future Research

Recent research has indicated that digitalization can have homogenizing effects on traditionally more disparate media systems, most notably because platforms like Facebook operate in a largely uniform manner across country borders (Fletcher & Nielsen, 2018; Stier et al., 2022). Still, any single-country study leaves need for follow-up research from a cross-national comparative perspective, especially with regard to the substantive conclusions and not least in light of our sample's restricted representativeness. Critically, this involves studies of countries where political ideology plays a greater role than in Central and Northern European countries (Fletcher et al., 2020; Mangold & Scharkow, 2022). Low baseline levels in ideological selective exposure and news audience polarization in Germany prevented an answer to the question to what extent news access via Facebook fosters or counteracts the emergence of ideologically skewed news diets (Scharkow et al., 2020).

Different approaches to operationalizing Facebook referrals can make meaningful differences for the substantive conclusions in applied research. Future studies should expand on our research to enable a more in-depth understanding of the specific mechanisms by which these differences come to exist. More generally, we urge readers to not mistake our results with evidence for the general superiority of the referrer-ID approach over timeline-based approaches across platforms. While similar systems exist for other social media platforms and online intermediaries (e.g., news portals), systematic assessments for each are needed to put applications of the referrer-ID approach on similarly solid methodological ground for platforms other than Facebook.

While our study expanded on the vast majority of previous web tracking research that has used desktop data for studying referrals (e.g., Cardenal et al., 2019; Flaxman et al., 2016; Guess, 2021; Wojcieszak et al., 2021), news access via mobile devices has become increasingly popular in recent years (Husin, 2018; Newman et al., 2022; Walker, 2019). Despite, or precisely in line with, the broader body of literature indicating that news exposure on mobile devices is often a less effortful endeavor than on desktop computers (Collier et al., 2021), research has indicated that referrals on mobile devices follow mechanisms similar to the ones observed in desktop environments (Stier et al., 2022). However, such generalizations are not straightforward, because in contrast to desktop computers (where most online activity can be captured through browser logs), social media use and news exposure on mobile devices often occurs within apps (Tyler et al., 2022). It is essential that academics develop tracking techniques that cover a broader range of social media and smartphone-based behaviors to enable a more comprehensive picture of news exposure in the digital age.

Notes

1. We report the share of referred Facebook users to rule out that we find only a relatively modest prevalence of Facebook news referrals just because Facebook use is less widespread in Germany than, e.g., the USA. The differences between approaches were similar when examining the share of total online users referred to news sites by Facebook.
2. The associations of news visits not referred by Facebook were in line with existing theory (see Online Appendix Figure S2): Male, older, and politically interested citizens consumed more news in general. Higher education more distinctly corresponded with more visits of public broadcasters and the legacy press rather than digital-born news outlets.

3. The differences in the associations of Facebook referrals with news diet balance produced by not only the referrer-ID approach but also the subpage approach on the one hand and the sequence and timeframe approaches on the other hand were statistically significant by themselves, $p < .05$, both on the level of news outlets and news types (Online Appendix Figure S4).

Notes on contributors

Felix Schmidt is a PhD student in a research project focusing on the effects of negative political communication during election campaigns and is affiliated with GESIS – Leibniz Institute for the Social Sciences. His research primarily relies on the analysis of tracking data and focuses on inequalities in citizens' news consumption.

Frank Mangold is a senior researcher in the Department of Computational Social Science at GESIS— Leibniz Institute for the Social Sciences. His research interests are audience structures and media repertoires, opinion leadership, and quantitative methods, especially data collection and statistical modeling.

Sebastian Stier is Scientific Director of the Department Computational Social Science at GESIS – Leibniz Institute for the Social Sciences and Professor of Computational Social Science at the University of Mannheim. His research interests are computational social science, digital society, political communication and comparative politics.

Roberto Ulloa is a postdoctoral researcher at the Cluster "The Politics of Inequality" at the University of Konstanz and at GESIS – Leibniz Institute for the Social Sciences. His interests include online behavior and the role of online platforms in shaping societies.

ORCID

Felix Schmidt http://orcid.org/0009-0009-1829-6165
Frank Mangold http://orcid.org/0000-0002-9776-3113
Sebastian Stier http://orcid.org/0000-0002-1217-5778
Roberto Ulloa http://orcid.org/0000-0002-9870-5505

Data availability statement

Replication materials including data and R scripts that enable the replication of all analyses are available on OSF: https://osf.io/pywz2/?view_only=98abd4d095e840a193c468efc1168126https://osf.io/pywz2/?

References

Aigenseer, V., Urman, A., Christner, C., Maier, M., Adam, S., Makhortykh, M., & Gil-Lopez, T. (2019, November 24). *Webtrack–desktop extension for tracking users' browsing behaviour using screen-scraping*. GESIS Computational Social Science (CSS) Seminar, Mannheim. https://boris.unibe.ch/id/eprint/139219

Anspach, N. M. (2017). The new personal influence: How our facebook friends influence the news we read. *Political Communication*, *34*(4), 590–606. https://doi.org/10.1080/10584609.2017.1316329

Barberá, P. (2015). How social media reduces mass political polarization. Evidence from Germany, Spain, and the U.S. *Job Market Paper*, *46*, 1–46.

Breuer, J., Kmetty, Z., Haim, M., & Stier, S. (2022). User-centric approaches for collecting facebook data in the 'post-API age': Experiences from two studies and recommendations for future research. *Information Communication & Society*, *0*, 1–20. https://doi.org/10.1080/1369118X.2022.2097015

Cardenal, A. S., Aguilar-Paredes, C., Galais, C., & Pérez-Montoro, M. (2019). Digital technologies and selective exposure: How choice and filter bubbles shape news media exposure. *The International Journal of Press/politics*, *24*(4), 465–486. https://doi.org/10.1177/1940161219862988

Chen, A. Y., Nyhan, B., Reifler, J., Robertson, R. E., & Wilson, C. (2023). Subscriptions and external links help drive resentful users to alternative and extremist YouTube channels. *Science Advances*, *9*(35). https://doi.org/10.1126/sciadv.add8080

Christner, C., Urman, A., Adam, S., & Maier, M. (2021). Automated tracking approaches for studying online media use: A critical review and recommendations. *Communication Methods and Measures*, *16*(2), 79–95. https://doi.org/10.1080/19312458.2021.1907841

Collier, J. R., Dunaway, J., & Stroud, N. J. (2021). Pathways to deeper news engagement: Factors influencing click behaviors on news sites. *Journal of Computer-Mediated Communication*, *26*(5), 265–283. https://doi.org/10.1093/jcmc/zmab009

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, *80*(S1), 298–320. https://doi.org/10.1093/poq/nfw006

Fletcher, R., Cornia, A., & Nielsen, R. K. (2020). How polarized are online and offline news audiences? A comparative analysis of twelve countries. *The International Journal of Press/politics*, *25*(2), 169–195. https://doi.org/10.1177/1940161219892768

Fletcher, R., Kalogeropoulos, A., & Nielsen, R. K. (2021). More diverse, more politically varied: How social media, search engines and aggregators shape news repertoires in the United Kingdom. *New Media & Society*, *25*(8), 2118–2139. https://doi.org/10.1177/14614448211027393

Fletcher, R., & Nielsen, R. K. (2018). Are people incidentally exposed to news on social media? A comparative analysis. *New Media & Society*, *20*(7), 2450–2468. https://doi.org/10.1177/1461444817724170

Fletcher, R., Robertson, C. T., & Nielsen, R. K. (2021). How many people live in politically partisan online news echo chambers in different countries? *Journal of Quantitative Description: Digital Media*, *1*, 1–56. https://doi.org/10.51685/jqd.2021.020

Guess, A. M. (2021). (Almost) everything in moderation: New evidence on Americans' online media diets. *American Journal of Political Science*, *65*(4), 1007–1022. https://doi.org/10.1111/ajps.12589

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R. , and Tucker, J. A. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, *381*(6656), 398–404. https://doi.org/10.1126/science.abp9364

Guess, A. M., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on facebook. *Science Advances*, *5*(1). https://doi.org/10.1126/sciadv.aau4586

Guess, A. M., Nyhan, B., & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behaviour*, *4*(5), 472–480. https://doi.org/10.1038/s41562-020-0833-x

Haidt, J. (2022). *Yes, social media really is undermining democracy*. The Atlantik. https://www.theatlantic.com/ideas/archive/2022/07/social-media-harm-facebook-meta-response/670975/

Husin, H. S. (2018). The effects of Facebook referrals and mobile devices on online news consumption. *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication*, Langkawi, Malaysia (pp. 1–5).

Jürgens, P., & Stark, B. (2022). Mapping exposure diversity: The divergent effects of algorithmic curation on news consumption. *Journal of Communication*, *72*(3), 322–344. https://doi.org/10.1093/joc/jqac009

Kalogeropoulos, A., Fletcher, R., & Nielsen, R. K. (2019). News brand attribution in distributed environments: Do people know where they get their news? *New Media & Society*, *21*(3), 583–601. https://doi.org/10.1177/1461444818801313

Kümpel, A. S. (2018). The issue takes it all? *Digital Journalism*, *7*(2), 165–186. https://doi.org/10.1080/21670811.2018.1465831

Lee, J. K., & Kim, E. (2017). Incidental exposure to news: Predictors in the social media setting and effects on information gain online. *Computers in Human Behavior*, *75*, 1008–1015. https://doi.org/10.1016/j.chb.2017.02.018

Mangold, F., & Scharkow, M. (2020). How do filtering choices impact the structures of audience networks? A simulation study using data from 26 countries. *Communication Methods and Measures*, *14*(2), 125–144. https://doi.org/10.1080/19312458.2020.1724274

Mangold, F., & Scharkow, M. (2022). Metrics of news audience polarization: Same or different? *Communication Methods and Measures*, *16*(3), 157–181. https://doi.org/10.1080/19312458.2022.2085249

Mangold, F., Stier, S., Breuer, J., & Scharkow, M. (2022). The overstated generational gap in online news use? A consolidated infrastructural perspective. *New Media & Society*, *24*(10), 2207–2226. https://doi.org/10.1177/1461444821989972

Mangold, F., Vogelgesang, J., & Scharkow, M. (2017). Nachrichtennutzung in Deutschland. Eine nutzerzentrierte Repertoireanalyse. *Medien & Kommunikationswissenschaft*, *65*(4), 704–723. https://doi.org/10.5771/1615-634X-2017-4-704

Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, *41*(8), 1042–1063. https://doi.org/10.1177/0093650212466406

Möller, J., van de Velde, R. N., Merten, L., & Puschmann, C. (2020). Explaining online news engagement based on browsing behavior: Creatures of habit? *Social Science Computer Review*, *38*(5), 616–632. https://doi.org/10.1177/0894439319828012

Müller, P., & Bach, R. L. (2021). Populist alternative news use and its role for elections: Webtracking and survey evidence from two campaign periods. *New Media & Society*, *25*(10), 2663–2683. https://doi.org/10.1177/14614448211032969

Nelson, J. L., & Webster, J. G. (2017). The myth of partisan selective exposure: A portrait of the online political news audience. *Social Media + Society*, *3*(3), 1–17. https://doi.org/10.1177/2056305117729314

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Nielsen, R. K. (2022). *Digital News Report 2022*. Oxford: Reuters Institute for the Study of Journalism.

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin UK.

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, *5*(11), 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Prior, M. (2009). Improving media effects research through better measurement of news exposure. *The Journal of Politics*, *71*(3), 893–908. https://doi.org/10.1017/S0022381609090781

Robertson, R. E., Green, J., Ruck, D., Ognyanova, K., Wilson, C., & Lazer, D. (2021). Engagement outweighs exposure to partisan and unreliable news within google search. *ArXiv:2201.00074 [Cs]*. http://arxiv.org/abs/2201.00074

Scharkow, M. (2016). The accuracy of self-reported internet use: A validation study using client log data. *Communication Methods and Measures*, *10*(1), 13–27. https://doi.org/10.1080/19312458.2015.1118446

Scharkow, M., Mangold, F., Stier, S., & Breuer, J. (2020). How social network sites and other online intermediaries increase exposure to news. *Proceedings of the National Academy of Sciences*, *117*(6), 2761–2763. https://doi.org/10.1073/pnas.1918279117

Statista. (2022a, October 27). *Number of Facebook users in germany from september 2018 to september 2022*. https://www.statista.com/statistics/1017402/facebook-users-germany/

Statista. (2022b, November 9). *Number of monthly active Facebook users worldwide as of 3rd quarter 2022*. https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

Statista. (2023, February 10). *Facebook usage penetration in the United States from 2018 to 2027*. https://www.statista.com/statistics/183460/share-of-the-us-population-using-facebook/

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, *38*(5), 503–516. https://doi.org/10.1177/0894439319843669

Stier, S., Kirkizh, N., Froio, C., & Schroeder, R. (2020). Populist attitudes and selective exposure to online news: A cross-country analysis combining web tracking and surveys. *The International Journal of Press/politics*, *25*(3), 426–446. https://doi.org/10.1177/1940161220907018

Stier, S., Mangold, F., Scharkow, M., & Breuer, J. (2022). Post post-broadcast democracy? News exposure in the age of online intermediaries. *American Political Science Review*, *116*(2), 768–774. https://doi.org/10.1017/S0003055421001222

Taneja, H., Wu, A. X., & Edgerly, S. (2018). Rethinking the generational gap in online news use: An infrastructural perspective. *New Media & Society*, *20*(5), 1792–1812. https://doi.org/10.1177/1461444817707348

Thorson, K. (2020). Attracting the news: Algorithms, platforms, and reframing incidental exposure. *Journalism*, *21*(8), 1067–1082. https://doi.org/10.1177/1464884920915352

Thorson, K., & Wells, C. (2015). Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory*, *26*(3), 309–328. https://doi.org/10.1111/comt.12087

Tyler, M., Grimmer, J., & Iyengar, S. (2022). Partisan enclaves and information bazaars: Mapping selective exposure to online news. *The Journal of Politics*, *84*(2), 1057–1073. https://doi.org/10.1086/716950

Ulloa, R., & Kacperski, C. S. (2023). Search engine effects on news consumption: Ranking and representativeness outweigh familiarity in news selection. *New Media & Society*, 146144482311549. https://doi.org/10.1177/14614448231154926

Vaidhyanathan, S. (2018). *Antisocial media: How Facebook disconnects us and undermines democracy*. Oxford University Press.

van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., de Vreese, C., Matthes, J., Hopmann, D., Salgado, S., Hubé, N., Ste̜pińska, A., Legnante, R., Berganza, G., Sheafer, C., Reinemann, T., Stanyer, J., & Stanyer, J. (2017). Political communication in a high-choice media environment: A challenge for democracy? *Annals of the International Communication Association*, *41*(1), 3–27. https://doi.org/10.1080/23808985.2017.1288551

Walker, M. (2019). *Americans favor mobile devices over desktops and laptops for getting news*. https://www.pewresearch.org/fact-tank/2019/11/19/americans-favor-mobile-devices-over-desktops-and-laptops-for-getting-news/

Webster, J. G. (2014). *The marketplace of attention: How audiences take shape in a digital age*. The MIT Press. https://doi.org/10.7551/mitpress/9892.001.0001

Wieland, M., & Kleinen von Königslöw, K. (2020). Conceptualizing different forms of news processing following incidental news contact: A triple-path model. *Journalism*, *21*(8), 1049–1066. https://doi.org/10.1177/1464884920915353

Wojcieszak, M., Menchen-Trevino, E., Goncalves, J. F. F., & Weeks, B. (2021). Avenues to news and diverse news exposure online: Comparing direct navigation, social media, news aggregators, search queries, and article hyperlinks. *The International Journal of Press/politics*, *27*(4), 860–886. https://doi.org/10.1177/19401612211009160