Digital Journalism

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rdij20

Dissecting Non-Use of Online News – Systematic Evidence from Combining Tracking and Automated Text Classification

Michael V. Reiss

DIGITAL JOURNALISM
https://doi.org/10.1080/21670811.2022.2105243

ARTICLE

Dissecting Non-Use of Online News – Systematic Evidence from Combining Tracking and Automated Text Classification

Michael V. Reiss

Department of Communication and Media Research, University of Zurich, Zurich, Switzerland

ABSTRACT
A high proportion of non-users of news is considered a concern for a functioning democracy. However, existing empirical assessments on the share of news avoiders rely exclusively on survey data and the results vary drastically between studies, making it difficult to evaluate the severity of the issue. This study relies on tracking data of Swiss Internet users and applies and discusses two computational methods, identifying news at the domain and article level, to realistically assess the extent of non-users of online news. Results indicate that at least 14.2% of Internet users do not use news online. Furthermore, this study suggests that identifying news use solely based on tracking data at the domain level is distorted by a faux news effect, i.e., non-news use on news domains, and an invisible news effect, i.e., news use on small and unknown news domains. The parallel use of tracking data and supervised text classification allows to dissect and discuss these effects systematically. Similarly, it is found that not accounting for news use via apps overestimates the extent of non-use of online news. The findings provide valuable insights for future applications of these methods in similar contexts.

KEYWORDS
News avoidance; news consumption; supervised machine learning; automated text analysis; tracking data; faux news effect; invisible news effect

Introduction

News critically inform, educate or mobilize citizens, serving as a common ground and sense-making institution in society (Coronel 2003; Gans 2003; Van Aelst et al. 2017). The crucial role of news in society and for fulfilling basic requirements in a democracy is also acknowledged by many theoretical models of democracy (Christians et al. 2009; Habermas 1991; Street 2011). Furthermore, empirical findings show that news exposure is associated with increased civic engagement, political participation, and political knowledge—outcomes that are usually desired in democracies (Barabas and Jerit 2009; de Vreese and Boomgaarden 2006; Gil de Zúñiga, Jung, and Valenzuela 2012;

CONTACT Michael V. Reiss m.reiss@ikmz.uzh.ch

Supplemental data for this article is available online at https://doi.org/10.1080/21670811.2022.2105243.

Leeper 2020). But for this "important democratic resource" (Adoni et al. 2017, p. 227) to be effective, citizens are expected to consume news.

However, for a growing proportion of the society this is not the case. In today's high-choice media environment, especially online, individuals can selectively choose the media content they want to consume—for example depending on their political interest (Skovsgaard, Shehata, and Strömbäck 2016). This results in people that actively seek for and consume news on the one hand and people that do not consume news on the other hand (Blekesaune, Elvestad, and Aalberg 2012; Newman et al. 2019; Strömbäck, Djerf-Pierre, and Shehata 2013). In line with the theoretical role of news in democratic societies, news avoidance can therefore be considered an increasing problem for democracy (Skovsgaard and Andersen 2020). However, mixed results on the extent of news avoidance in society and imprecise measures make it difficult to assess the severity of the issue.

This study has the aim to advance our understanding on the prevalence of news non-use in the online context by applying and discussing two different computational methods that promise a realistic evaluation of the extent of non-use of online news. The investigation bases on a combination of tracking and web data. The main contributions of this study are the precise estimation of the extent of non-use of online news as well as the parallel application and comparison of two computational methods that have not been applied and compared in this context. The latter includes a systematic dissection of the invisible and faux news effect for the first time and an accurate quantitative assessment of the importance of apps for identifying the extent of non-users of online news.

Non-Use of News – Mapping the Issue

The scientific debate on news non-use is shaped by a number of related and intertwined conceptualizations, most of which relate to the notion of news avoidance which again can be distinguished into intentional and unintentional news avoidance.

For intentional news avoidance, Skovsgaard and Andersen (2020) identify three motives. First, people subjectively perceive an overload regarding the amount of news available to them (Holton and Chyi 2012; Villi et al. 2022). This perceived news overload can lead to news fatigue, eventually leading people to actively avoid news (Song, Jung, and Kim 2017). Second, people avoid news because news consumption can have negative effects on their well-being and mood (Boukes and Vliegenthart 2017; Newman et al. 2019; Villi et al. 2022; Woodstock 2014). In fact, news avoidance can even increase perceived well-being (de Bruin et al. 2021). The third reason identified for intentional news avoidance is a lack of trust in news (Palmer, Toff, and Nielsen 2020; Toff and Kalogeropoulos 2020).

Unintentional news avoidance, by contrast, is usually no active practice but rooted in personal preferences and characteristics: individual preferences, political interest, gender norms, upbringing, socialization, or social position affect the media type, content or level of news exposure individuals choose (Lindell 2018; Prior 2005; Skovsgaard, Shehata, and Strömbäck 2016; Stroud 2010; Toff and Palmer 2019; Valenzuela, Bachmann, and Aguilar 2019). Although choices and selective behavior are no new phenomena (Stroud 2008), the choices—most importantly—in and because of the online context are more extensive today compared to times before and apart from the Internet. Hence, due to increased choices, individual and socio-economic differences reveal themselves in diverging news consumption patterns. Consequently, the group of people that avoid news increased over time (Blekesaune, Elvestad, and Aalberg 2012; Newman et al. 2019; Strömbäck, Djerf-Pierre, and Shehata 2013).

Although the underlying causes of news avoidance differ, both types, intentional and unintentional news avoidance, share the communality that findings on the extent of news avoidance vary drastically depending on the study and operationalization of news avoidance. The proportion of news avoiders range from 3 to 73% across different studies (Lee and Yang 2014; Palmer, Toff, and Nielsen 2020). Skovsgaard and Andersen (2020) partly attribute these differences to variations in the conceptualization and operationalizations of news avoidance: Individuals qualify as news avoiders when not using news for example on a typical day (Karlsen, Beyer, and Steen-Johnsen 2020), once a week (Blekesaune, Elvestad, and Aalberg 2012; Trilling and Schoenbach 2013), or once a month (Palmer, Toff, and Nielsen 2020; Toff and Palmer 2019). Other studies apply relative cut off points (Wolfsfeld et al. 2016) or identify clusters of similar news use characteristics (Lee and Yang 2014). For example, the fög (2020) uses the term news deprived, grouping individuals with low news exposure compared to other groups of news consumers. Finally, some rely on verbalized categories of news consumption frequency (e.g., "sometimes," "often") (Newman et al. 2019; Toff and Kalogeropoulos 2020), on tailor-made scores to identify news avoiders (Strömbäck, Djerf-Pierre, and Shehata 2013), or let interviewees self-identify as news resisters (Woodstock 2014). It becomes apparent that news avoidance—depending on the study—comes in various forms and does not necessarily mean that people do never use news.[1]

The large inconsistencies in results and conceptualizations make it difficult to assess the extent of the problem. When concerns exist regarding inequalities from peoples' diverging news use (Van Aelst et al. 2017), such as a decrease in political knowledge and participation for news avoiders (Barabas and Jerit 2009; Gil de Zúñiga, Jung, and Valenzuela 2012), it makes a big differences for assessing the situation whether the share of potentially vulnerable individuals is 3 or 73% and whether news avoidance implies comparatively low news use or never using news. Additionally, all of the above-mentioned empirical studies, rely on self-reported, mostly survey data for identifying news avoiders. As has repeatedly and consistently been shown, self-reported data does not very reliably represent actual media use and news exposure (Parry et al. 2021; Prior 2009; Scharkow 2019). Particularly for the online context, participants tend to overestimate Internet use (Araujo et al. 2017; Festic, Büchi, and Latzer 2021; Scharkow 2016) and visits to online news websites, or sharing of political articles on social media (Haenschen 2020; Vraga and Tully 2020).

Furthermore, retrospective measures can miss rare and unintentional news consumption. This is true for people with a news-finds-me perspective in particular. Such people do not actively seek news but feel, in contrast to news avoiders, that news will find them (Gil de Zúñiga, Weeks, and Ardèvol-Abreu 2017). Such behavior and incidental news exposure in general, where people stumble upon news online as a byproduct of their other online activities (Kim, Chen, and Gil de Zúñiga 2013), pose a challenge for reliably measuring news use based on self-reports.

In order to realistically assess the extent of non-use of news, more precise measures beyond the sole focus on self-reported data is required. For this reason, this study introduces and discusses two computational methods that are based on tracking data and automated text analysis to assess the extent of individuals that are non-users of online news. This approach does not rely on peoples' self-reports but on behavioral data, promising a more realistic assessment and evaluation of the extent of non-use of online news. However, both methods are comparatively young in the toolbox of social science research and have not yet been applied or compared in this context. The first research question therefore is: How do different computational approaches contribute to identifying non-users of online news?

The study focuses on online news because concerns regarding an increase in news avoidance are predominantly about and rooted in the online context. This is because the Internet context, a high-choice media environment, is especially susceptible to selective behavior and the associated consequences. Additionally, online sources are the most popular news source in many countries today while other news media with a lesser degree of choice loose popularity (Newman et al. 2020). This increases the likelihood that people who are not much interested in or socialized with news, for example, will have no news exposure at all, making it important to investigate the extent of news non-use in the online context.

This study does not investigate individual motivations of or reasons behind news non-use and can therefore not draw conclusions on peoples' motivation in their Internet use. To emphasize this, this study uses the term news non-use instead of news avoidance which is not consistently conceptualized but—often misleadingly—implies an active or motivated choice by the users. When focusing on directly observable behavior, the term news use is less presupposing and normative (Villi et al. 2022)[2] and therefore adequate for the context of this study.

Tracking Data and Supervised Text Analysis for Assessing Non-Use of Online News

This study makes use of a combination of tracking and web data to answer the research question. Based on this data two methods are applied and compared. The first method mainly relies on the tracking data and identifies news use on the domain and app level. For this purpose, users' tracked Internet use is compared with an extensive list of news sources. The second method identifies news use at the article level and therefore focuses on the content of each tracked website. Automated text analysis based on supervised machine learning is used to classify the content of websites into news and non-news.

The use of tracking of individuals' online behavior is a comparatively young method and allows, compared to survey data, a more fine-grained and less biased assessment of users' actual online behavior. Recent examples of tracking studies include investigating the influence of populist attitudes on online news use (Stier, Kirkizh, et al. 2020), improving our understanding on peoples' habits, diets, and patterns of online news use (Guess 2021; Möller et al. 2020; Vermeer et al. 2020) as well as generational gaps therein (Mangold et al. 2021), comparing different measures of news exposure (Vraga and Tully 2020), prediction of voting behavior based on voters' online activities (Bach et al. 2021), exposure to untrustworthy political websites (Guess, Nyhan, and Reifler 2020), and interaction with political content on social media (Haenschen 2020).

Similarly, with the advent of automated text classification methods and the mounting availability of large digitalized text corpora, automated text analysis became increasingly popular in the context of political communication (Barberá et al. 2021; Boumans and Trilling 2016; Grimmer and Stewart 2013). Automated text analysis comprises of different computational methods and their applicability varies depending on the task and research problem at hand. Identifying non-users of online news based on the consumed online content requires distinguishing this content into news and non-news. Facing a classification problem and large amount of data makes supervised machine learning an appropriate choice. Supervised text classification requires human labeled training data on which a classifier is trained before being applied to the full data set to automatically classify the data. A variety of different topics related to news use already made use of this type of automated text analysis such as for the classification of news and news topics (Flaxman, Goel, and Rao 2016; Guess 2021) and policy issues (Burscher, Vliegenthart, and De Vreese 2015), for predicting the share worthiness of news articles (Trilling, Tolochko, and Burscher 2017) and the voting behavior based on social media posts (Ceron, Curini, and Iacus 2015), and for quantifying news media bias (Budak, Goel, and Rao 2016). An alternative method is based on dictionaries where text content is compared with predefined wordlists that indicate certain categories. However, as has been shown repeatedly, supervised machine learning outperforms dictionary based text analysis for various contexts (Barberá et al. 2021; Hartmann et al. 2019; L. K. Nelson et al. 2021; van Atteveldt, van der Velden, and Boukes 2021).

Although both proposed methods—tracking and supervised text classification—have limitations (Grimmer and Stewart 2013; Stier, Breuer, et al. 2020), a parallel application and combination of the two methods addresses and resolves some of the limitations of the respective other method: First, for the list-based approach to capture all of the tracked news use requires the list of news sources to be exhaustive. Otherwise, news use of individuals might go unnoticed, overestimating the share of non-users of online news. Several studies on news consumption—independently of the method—focus on a number of popular or selected news sources, capturing the news consumption for a majority of the population (Newman et al. 2020; Stier, Kirkizh, et al. 2020). While this proves sufficient for many research objectives, it potentially misses smaller online news websites and news content on websites not listed as news source (e.g., blogs). We call the structural bias of this approach *invisible news effect*.

Second, even on popular news websites not all content is news. Recipes, crosswords, and other non-news content are popular content for some news sites. Budak, Goel, and Rao (2016) and Flaxman, Goel, and Rao (2016) find that the share of actual news in the overall content of news websites is only 42 and 46%, respectively. In addition, some websites, especially portal sites like gmx.ch, bluewin.ch, and yahoo.com, offer some extent of news next to a variety of additional non-news services such as email or games. Hence, assessing news use at the domain level could lead to overestimating users' online news use. We call such systematic overestimation on the domain level *faux news effect*. Supervised text classification prevents these shortcomings by identifying news content on the article level for each tracked website, theoretically allowing an even finer grained news detection compared to the list-based approach.

Third, supervised text classification, however, requires large amount of labeled training data to provide meaningful classification results (Barberá et al. 2021; van Atteveldt, van der Velden, and Boukes 2021). It can be cumbersome and difficult to sample sufficient data for each class for manual coding; particularly, as in this case, when having severe imbalance in the class distribution. Here, it proves helpful to have the tracking data and a list of online news sources to assist the sampling of texts for both categories for manual coding.

The parallel analysis of news use at the domain and article level provides the opportunity to systematically investigate the extent of the invisible and faux news effect. When only having data on the domain level, both effects are likely to distort findings on the extent of news use or non-use of news. In such a scenario, theoretical reasoning suggests that the invisible news effect contributes to overestimating the extent of non-users of news while the faux news effect contributes to underestimating it. This study aims to systematically dissect the non-use of online news, estimating the size of these effects for the first time. Such investigations are important because many tracking studies (on news use) are limited to the domain level (Haim, Breuer, and Stier 2021; Mangold et al. 2021; Stier, Kirkizh, et al. 2020), and therefore affected by the invisible and faux news effect. In turn, other tracking studies do not incorporate mobile or app use (Guess 2021; Möller et al. 2020; Vermeer et al. 2020). This could also distort results because mobile and app news use is different from desktop news use and the only access to news for some users (Dunaway et al. 2018; J. L. Nelson 2020). It can therefore be expected that the extent of non-users of online news is overestimated when taking app news use not into account. Hence, the second research question is: What is the effect of faux news use, invisible news use and app use on the extent of non-users of online use?

Method and Analysis

Sample

This study relies on a combination of tracking, web, and survey data which were collected stepwise. Tracking data was collected in fall 2018 by tracking the website visits on desktop devices (desktop and laptop computers) as well as website visits and app use on mobile devices (smartphones and tablets) of Swiss Internet users. All participants were already part of an actively recruited tracking panel by an independent market and social research company. See Festic, Büchi, and Latzer (2021) for a more detailed description on the sampling and tracking of this data. The collected variables were the specific URL of a visited website or the name of a used app (mobile only), duration and time of visit. Tracking lasted for 30 days, participants with available tracking data of less than three days were excluded. Subsequently, participants were invited to complete a survey questionnaire that lasted 30 min on average and covered

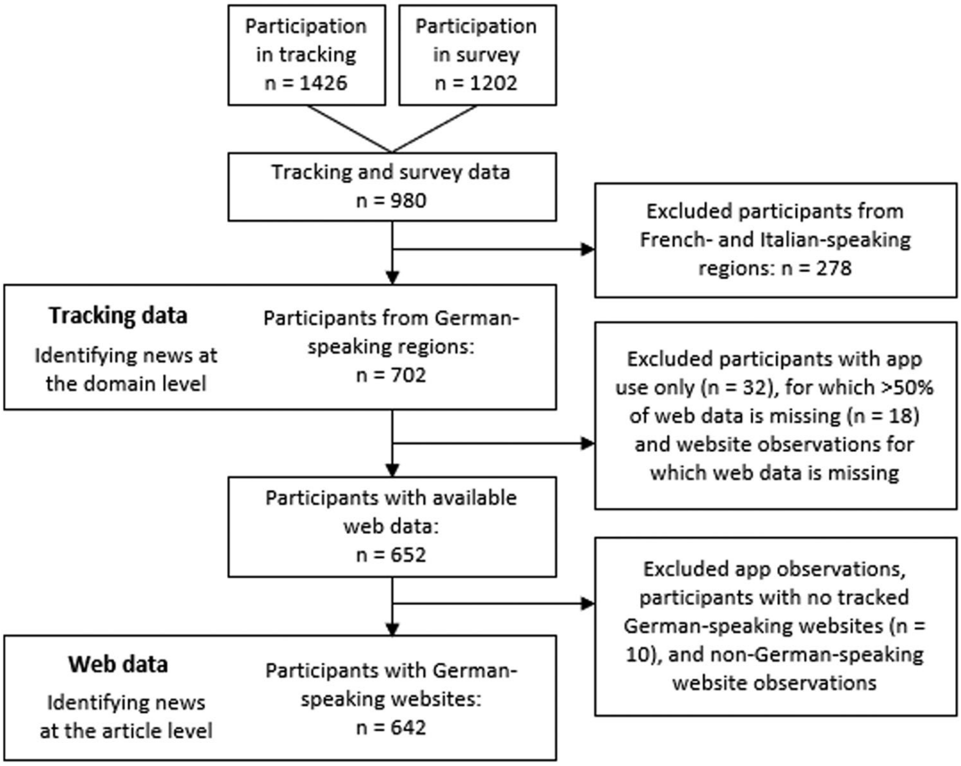

Figure 1. Flowchart for sample composition and data cleansing.

questions on Internet-use related topics. This survey sample of 1202 participants is non-probability based but representative for the Swiss online population over the age of 16 with respect to age, gender, language region, household size, and employment status. Participants received a small pecuniary incentive for their participation and all participants in the tracking and survey gave informed consent on their participation. The study's research design was approved by the university's ethics review board and was preregistered online.[3] For ethical reasons, participants could temporarily disable tracking at any time. As Festic, Büchi, and Latzer (2021) show for the same data, it does not appear that participants used this option frequently.

This study focuses on two overlapping subsets of the tracked and surveyed participants. Figure 1 informs about the process of sample composition and data cleansing. The first subset—*tracking data*—comprises of 702 participants from the German-speaking regions in Switzerland and their tracked Internet use. The tracking data is used as a basis for comparing the tracked website and app use with a list of news sources to identify news use at the domain and app level. Based on this tracking sample, starting in spring 2019, an automated web scraper collected the HTML data and a complete screenshot of each tracked website via its URL. For apps, only the app name was tracked because most apps do not allow or provide a technical possibility for scraping their content, rendering the scraping of specific content for app use sessions impossible. Subsequently, the collected HTML files were converted to plain text (e.g., excluding all HTML tags) using the python package inscriptis (1.1.2) (Weichselbraun 2021).

Table 1. Sample characteristics.

	Tracking data	Web data	2018 Census
Participants	702	642	–
only desktop tracking	38	71	–
only mobile tracking	323	286	–
desktop and mobile tracking	341	285	–
Mean tracking duration[1]	28.7 days	24.4 days	–
website observations	1,280,761	694,316	–
App observations	1,096,649	–	–
Observ. per particip. (mean / median)	3,387 / 1,781	1,081 / 362	–
Female	47.3%	47.2%	50.4%
Age	44.6	44.8	42.4
Secondary education	69.6%	69.4%	59.6%
Tertiary education	23.3%	23.5%	28.8%
Income[2]	6,001–8,000 CHF	8,001–10,000 CHF	9,560 CHF
Political interest[3]	3.45	3.46	–
Self-reported internet use per day	3.5h	3.4h	–
Tracked internet use per day	2.02h	0.45h[4]	–

1Participants with a tracking duration of less than 3 days were excluded in this study.
2Median monthly household income. Means of nine categories are 4.50 (tracking data) and 4.52 (web data). Census provides mean income.
35-point likert scale, 5 = high interest.
4App use and non-German-speaking websites excluded for this ample.

Inscriptis provided comparable or better results than packages with a similar purpose such as beautiful soup. Based on these texts, the fastText classifier (lid.176.bin) (Joulin et al. 2016) was used to identify the language for each text. A manual check of a subset of the automatically classified texts indicated a good performance of the language classification (see Online Appendix A, supplementary materials, for details on this validation). Texts identified as German-speaking and the corresponding tracked website observations constitute the second dataset. This data—subsequently called *web data*—comprises of tracked observations and website texts by 642 participants and represents the basis for identifying news at the article level based on supervised text classification. Table 1 informs about sample characteristics of the two datasets (see Figure A1 in Online Appendix A for the distribution of website/app use per user and information on the tracking duration). The restriction on German-speaking texts was made as a simplification to improve the performance of the supervised text classifier. It is plausible to assume that the costs of this restriction are low for the task of identifying non-users of online news because most people who largely consume news in a non-dominant language like English would probably use news in the dominant German language at least sometimes.

Identifying News at the Domain Level

News websites and apps were identified by comparing the domains and app names with an extensive list of relevant online news sources for Switzerland. It was aimed to obtain a high degree of completeness for the list of news sources to ideally capture all news sources that are used by Swiss Internet users. The list is based on analyses regarding news use in Switzerland by NET-Metrix (2018),

Udris and Eisenegger (2019), and WEMF (2019). Furthermore, data from the German Audit Bureau of Circulation (IVW), by Schwaiger (2022) on alternative news media in the German-speaking countries, and by Stier, Kirkizh, et al. (2020) on popular global online news sources were included. Furthermore, the list includes the first 200 apps that were listed in the charts of the Apple app store and Google Play store for the category news in spring 2020. For each website, the corresponding app was added, if available, and vice versa. Unknown news sources were manually checked to ensure their news character. Duplicates were removed. All in all, the list contains 2266 unique entries of online news sources; 1398 websites and 868 apps. See the supplemental material in Online Appendix B for the full list of news sources.

Subsequently, to identify news use in the tracking sample, the list of news sources was compared with the tracked domains of each website and the app names (e.g., nytimes.com for the website and NYTimes for the app). Website news sources that appear in the tracking data are checked whether news content on this domain is restricted to some specific subdomains or subdirectories. Some cases, for example srf.ch, are very popular among Swiss Internet users but the tracking data makes apparent that a large proportion uses it to watch series or films. News is available only at the homepage as well as at some specific subdirectories (e.g., srf.ch/news). For domains where news are limited to some subdomains or subdirectories, only these subdomains or subdirectories and the home page were taken as the basis for identifying news.

Identifying News at the Article Level

News at the article level were identified by means of supervised text classification. A stratified random subsample from the Web data was used to manually label website texts into the category news or non-news. This manual labeled data built the training data for the classifier. To get sufficiently large training data for both classes and against the backdrop of a highly imbalanced prevalence of news versus non-news texts in the data, texts of a domain listed as news were oversampled for manual coding based on the list of news sources. The inclusion of identical URLs was capped at 25 to avoid a bias of the training data by popular websites. Three independent and trained coders performed the manual coding. Based on the collected screenshot of each website, a website was coded as news when at least one category of newsworthiness criteria suggested by Harcup and O'Neill (2017) was present. Krippendorff's alpha was 0.85, indicating a reasonable inter-coder reliability (Krippendorff 2004) and conflicts were settled by majority ruling; no website was excluded. In total, 3819 websites were categorized as news, and 3353 websites as non-news. See Online Appendix C for details on the manual coding.

Based on this labeled data, the Python scikit-learn machine learning library (0.23.2) (Pedregosa et al. 2011) was used to train, test, and apply machine learning-based text classification. Random sampling was used to split the labeled data into a training set (80 percent) and test set (20 percent). Subsequently, because news account for approximately 5.8% of the data, the test set was stratified to resemble a realistic test situation. The parameters for the preprocessing of the texts, the classification algorithm and its hyperparameters were chosen based on their performance comparing the average F1 values[4] from at least 50 train/test runs for a multitude of different

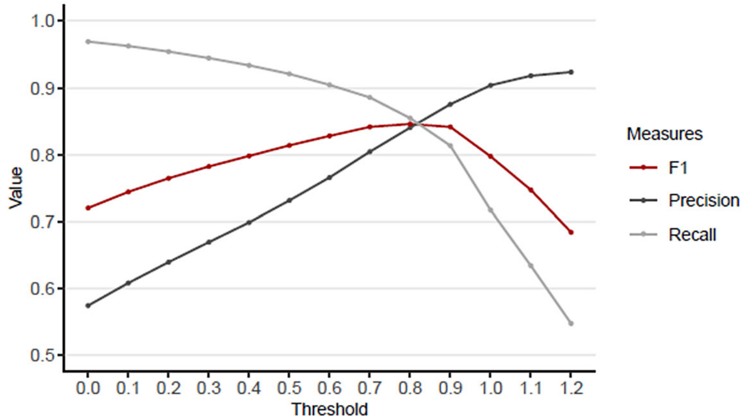

Figure 2. Classifier performance for different classification thresholds.

parameter and hyperparameter combinations; effectively resembling an extended grid search and cross-validation. See Online Appendix D for details on the supervised text classification process.

The performance was evaluated using the F1 measure because false positives and false negatives (i.e., predicting news when it is actually not news and predicting non-news when it is actually news) are, in principle, equally costly for determining news use and non-use of news. However, because of the sensitive definition of non-use of online news in this study, see below, even a single false positive could wrongly turn an actual non-user into a user. For this reason, when evaluating models, precision was given slightly more weight than recall. Figure 2 presents the tradeoff between precision and recall fordifferent classification thresholds of the best performing model; 0 being the default threshold (see Figure D2 in Online Appendix D for additional information on this relation). Linear Support Vector Classification demonstrated to perform best and the scores for the best performing model are: precision of 0.87, recall of 0.81, and F1 of 0.84 for a threshold of 0.9; which maximizes F1 while providing a reasonable prediction/recall tradeoff for the task at hand. This constitutes a good performance which is similar or slightly better compared to text classifications for related tasks (Budak, Goel, and Rao 2016; Burscher, Vliegenthart, and De Vreese 2015; Flaxman, Goel, and Rao 2016; Vermeer et al. 2020). Accuracy is not reported as this measure is not meaningful for imbalanced class distributions. Finally, the classifier was trained on the full labeled data and was applied to all non-labeled texts of the Web data.

Identifying Non-Users of Online News

As discussed above, a plethora of approaches are suggested in the literature as to when someone is counted as a non-user of news. However, the ongoing discussion shows, there exists no single best approach. Instead, this study applies a *conservative* approach to identify non-users of online news. In line with Palmer, Toff, and Nielsen (2020) and Toff and Palmer (2019), non-users of online news are defined as users who never use news during the period of one month. Palmer, Toff, and Nielsen (2020, p. 5) call such users "extreme news avoiders," which renders this approach well suited to establish a conservative baseline of non-users of online news. Focusing on this period furthermore corresponds well to the tracking, as users were on average tracked for 28.7 days for the tracking data (domain level) and 24.4 days for the Web data (article level).

This conceptualization results in the following operationalization: For the domain level, a user was counted as a non-user of online news when none of the user's tracked observations on the domain and app level matched with the list of news sources. Similarly, for the article level, a user was counted as a non-user of online news when none of the user's tracked websites were classified as news by the trained text classifier. In contrast, having one tracked observation that is identified as news results in the user being counted as a user of online news. For such users, the frequencies of online news use are investigated, too.

The second research question is approached by calculating the number of users that use news as identified on the article level but not on the domain level (invisible news effect), the number of users that use news as identified on the domain level but not on the article level (faux news effect), and the amount of news users that can only be identified as such when having app use data (i.e., no news use on mobile and desktop browsing).

Results

Reporting on research question one, Table 2 presents the frequency of online news use including non-users of online news for both methods.[5] When matching the tracked domains of each user with a list of news sources, 14.5% of users do not use any website/app that is classified as news within the investigation period. For the Web data, which allows analysis by both approaches, supervised text classification identifies a share of 23.1% of non-users of online news. In comparison, 23.4% are identified as non-users of online news when comparing the tracked domains with the list of news sources for the Web data. Furthermore, an additional share of 12.1 to 14.5% of users, depending on the method, use news at one to three days within 30 days; resulting in a monthly news use of a few minutes.

Table 2. Frequency of online news use including non-use of online news (news use on 0 out of 30 tracked days, in bold) for the tracking and web data; the latter seperated by method of news identification.

Data	n	Method of news identification	Online news use within 30 days and mean news use time in minutes				
			0 days	1–3 days	4–10 days	11–24 days	25–30 days
tracking	702	list-based (domain level)	**14.5%** 0 min	14.5% 7.2 min	19.5% 27.7 min	26.6% 148.5 min	24.8% 483.8 min
web	642	supervised class. (article level)	**23.1%** 0 min	13.7% 2.8 min	29.4% 10.3 min	20.2% 60.4 min	13.6% 209.3 min
		list-based (domain level)	**23.4%** 0 min	12.1% 4.2 min	29.4% 11.5 min	22.6% 76.7 min	12.5% 218.1 min

95% confidence intervals for population proportions range from +/− 2.5%–3.7% for the frequency and from +/− 0.8–66 for the means of news use time in minues. For details, see Figure E1 in the Online Appendix.

Table 3 Invisible news and faux news effect as well as apps for news use as contributers to assessing the extent of non-use of online news.

Effect	News identification	Contribution to the extent of non-users of online news
Invisible news use	On the article but not on the domain level	−4.7%
Faux news use	On the domain but not on the article level	+ 4.4%
App-only news use	Via apps but not on websites	−8.9%

Overall, 3.6% of observations for the tracking data are identified as being news based on comparing the tracked domains and apps with the list of news sources (See Table A2 in Online Appendix A for the most popular news sources). For the Web data (i.e., no app data, only German-speaking websites), the share is 5.8% for both the classification based on the supervised text classification and the list-based approach. The correlation between news as being identified by the text classification and by the list-based approach for the Web data is 0.75.

Addressing the second research question, Table 3 presents the invisible and faux news effect and the contribution of app-only news use for assessing the extent of non-use of online news for the Web sample. Accordingly, this invisible news effect amounts to 4.7%, while the faux news effect has an opposing effect direction and a size of 4.4%. Finally, 8.9% of users use news only via apps. Taking these distortions into account, the corrected extent of non-users of online news for the Web sample is 14.2%. See Online Appendix F for more a more detailed account.

Discussion

This study applied two methods to investigate the extent of non-users of online news complementing existing findings that predominantly build on self-reported survey data which demonstrated to be an imprecise measure of online media exposure. The first method relied on the comparison of tracked domain and app use with an extensive list of news sources identifying that 14.5% of participants do not use any online news within the period of one month. The second method identified non-use of online news at the article level by classifying each website text into news or non-news. This method revealed a higher share of 23.1% of non-users of online news. Both methods therefore confirm previous findings from the literature, which, despite huge differences depending on the study, also conceptually, indicate that a considerable share of the population does not use news.

The difference in the results between the two methods can be attributed to various factors. Three important factors were investigated systematically by making use of the parallel identification of news at the domain and at the article level. Most striking, it was shown that apps are crucial for online news use for some users. When not taking app use into account, the extent of non-users of online news is overestimated by around 9 percentage points. Although this might not seem much, not considering app use increases the extent of non-users of online news by 62%. Future tracking studies on news consumption should therefore be aware of this finding when conceptualizing the research design. Second, when identifying news at the domain level, some content is in fact not news but crosswords, cooking recipes, etc. Not accounting for this faux news effect will overestimate news use and therefore underestimate non-use of news. Additionally, utilizing supervised text classification that is more powerful and precise in identifying news events at the article level suggests that the faux news effect contributes to underestimating the extent of non-users of online news by 4.4 percentage points. On the one hand, this is partially in line with studies that identify a high share of non-news content on news websites (Budak, Goel, and Rao 2016; Flaxman, Goel, and Rao 2016). On the other hand, this is surprising because it is reasonable to assume that people at least occasionally also use or stumble upon actual news when being on news websites for other reasons. A clear evaluation of this relation also depends on the goodness of the supervised text classifier. The third factor relates to small or unknown news sources that can go unnoticed when focusing on a list-based approach, leading to an overestimation of non-use of news. Utilizing news identification at the article level, this invisible news effect is estimated to account for 4.7% of users. It is possible, that the invisible news effect involves some people with a news-finds-me perception (Gil de Zúñiga, Weeks, and Ardèvol-Abreu 2017). These people do not have a routine news consumption and fixed sources but might incidentally encounter news (Kim, Chen, and Gil de Zúñiga 2013), also aside from popular news sources. However, whether this is in fact the case and to what extent requires more specific and additional investigations beyond the scope of this study. Overall, taking these effects into account, the identified extent of non-use for the Web data is very similar to the identified extent of non-use for the tracking data. This supports the validity of the findings.

Comparing the manual labeling of texts with the list-based classification furthermore informs about the potential goodness of news identification by the latter. For the 7172 texts that were labeled manually, in 84% of cases, the manual labeling and the list-based approach had a consistent classification (i.e., both classified as non-news or news, respectively). For a small proportion of 1%, the manual coding classified texts as news while the list-based approach did not, indicating a high degree of completeness for the list of news sources and a low prevalence of the invisible news effect. A larger share of 15%, however, was not classified as news by manual coding but identified as news by the list-based approach, indicating a faux news effect. While the results above suggest that both effects are almost similar in size, this comparison with the manual coding indicates, that the faux news effect might be larger. It is important to note that this is no contradiction because the former applies to the user level while the latter refers to the news use level. In fact, when also analyzing the results of Table 3 on the news use level, the faux news effect is larger than the invisible news effect. Future tracking studies should be aware of this.

This study applied a conservative conceptualization of non-use of news and a sensitive cut-off between news users and non-users. Having the different news use frequencies of online news users, however, makes it easy to assess how the extent of non-users would change when applying a more relaxed conceptualization of online news non-use. For example, the fög (2020) identifies a share of 37% of the Swiss population as news deprived, grouping people with a low news use compared to other people. In parallel for this study, when including people that use news at one to three days withing 30 days, an arguable infrequent news use (and low news use, when looking at the time) the extent of non-users of online news would double to 29%.

Overall, it is difficult to judge the performance of the supervised text classification, which is, arguably, the more complex method. Two related factors in particular have a great influence on the goodness of the classifier and inform on the difficulty of the classification. The first factor is the definition of news. This study applied a broad definition of news, rendering the cut-off point rather vague for some topics. While this broad definition has the advantage to capture the entirety of what people use for their political and social orientation—and not only hard news—the goodness of the classifier would probably have benefited from a more closely defined news concept. This is, however, not only a problem for the text classifier, but for human coders as well. Irrespective of the reasonable inter-coder reliability, for some, for example soft news topics and infotainment, distinguishing news and non-news proved to be difficult. Excluding texts where not all coders agreed on the class increased the F1 value of the classifier by around 4 percentage points. Second, related to that and as Figure 2 displays, for the default classification threshold, recall is very high while precision is low. This means that the classifier makes almost all of its errors by falsely predicting a text to be news (that is actually not news). For an unbiased classifier that faces highly imbalanced data, it is therefore logical to more often predict a wrong label for the majority class than a wrong label for the minority class, because, assuming an approximately even distribution of difficult texts for both classes, hard to distinguish texts will more frequently appear for the majority class. By moving the classification threshold, it was possible to correct for this bias that is not induced by the classifier itself, but by the imbalance in the data and a vague cut-off between news and non-news for some texts. Especially when facing the alternative of identifying news by a list-based approach, to profit by the advantage of article level news identification requires a very good performing classifier which, in this case but depending on the definition of non-users of news, needs a very high precision, in particular.

A few limitations should be considered when interpreting the results of this study. First, due to restricted access, Internet use on social media or messaging apps (Kümpel 2022) could not be considered in this investigation. However, several findings suggest that news use on Facebook, the by far most popular online social network for Swiss Internet users (Latzer, Festic Noemi, and Kappeler 2020), does not greatly distort the results. In the tracking data only 3.4% of observations are dedicated to Facebook and research suggests that when on Facebook, only a small share of 4–6% of users' news feed content is news (Clegg 2021; Haim, Breuer, and Stier 2021). Moreover, Swiss Internet users assign Facebook a low relevance and popularity for their news use (Reiss et al. 2021; fög 2020). Second, although Internet use was tracked both for mobile and desktop devices, participants were not tracked when using another device or different browser, for example. Furthermore, a share of survey participants did not participate in the collection of tracking data and tracking participants might have altered their Internet use due to the monitoring. Such unobserved participation bias, which is typical for tracking studies (Makhortykh et al. 2021), and having a non-probability sample limits the generalization of findings. In addition, mainly for the Web data, and albeit great care to prevent for that, there is a chance that data processing (i.e., the automated scraping of websites and identification of German-speaking texts) might have introduced undesired and unaccounted bias. Finally, on the article level non-German-speaking content was not considered and the news identification was limited to website content because content of the app use could not be collected. The former does not seem to have affected the results much, as only 2.5% of the news use for websites in the tracking data is accounted for by non-German-speaking domains (e.g., bbc.co.uk) and all but one individual that use non-German-speaking news also used German-speaking news in the tracking data.

Conclusion

When increasing non-use of news is considered a problem for democracy, an accurate understanding on the extent of the problem is crucial. Because existing studies predominantly use imprecise measures and come to vastly diverging results, this study introduced and discussed two computational methods that are based on a combination of tracking, web, and survey data to realistically determine the share of non-users of online news. The results indicate that at least 14.2% of Swiss Internet users do not use news online.

This study applied a broad conceptualization of news and had a comparatively high threshold to count as non-user of online news. The identified share of non-users of online news can therefore be considered a lower boundary of non-use of online news. It is very likely that the proportion of non-users of online news is higher when focusing on hard news or when relaxing the definition of non-use of news, for example by also including users with a low and infrequent news use.

Alongside the realistic findings on the extent of non-use of online news, the main contribution of this study lies in the methodical application and discussion. The parallel application of two methods proved to be very fruitful, allowing a critical reflection and a systematic dissection of non-use of online news. Including news identification on the article level allowed for the precise estimation of the invisible news effect and the faux news effect. Moreover, the high importance of apps for correctly assessing the extent of non-use of online news was demonstrated. This study provides valuable insights for future applications of these methods in similar contexts. Future tracking studies can profit from these findings by better assessing bias in their results when additional app tracking is not viable or when news user identification is limited to the domain level as both is often the case.

Overall, while being potentially more powerful, supervised text classification comes with higher costs regarding time, resources, and skill. In addition, depending on the objective, a very good performance of the text classification is required to bring out the advantages over the list-based approach that provides good results itself. This is especially the case when the domain list is as exhaustive as possible and when the content of websites (or subdomains and subdirectories) of a domain can be expected to be rather homogenous regarding the dimension(s) of interest and a domain therefore clearly attributable. In contrast, utilizing supervised machine learning when having tracking data is sensible when being interested in details on the content level (e.g., topics or sentiments of a text, actors or sources in a text) or when domains are hard to categorize and heterogeneous regarding the dimension(s) of interest. In a borderline case or when facing limited resources, we recommend using a list-based approach but to additionally investigate the validity of the list-based categorization by manually coding a random subset of websites (e.g., 200 to 1000 texts depending on the number of classes and balance thereof). Such manual validation—checking for potential biases of the approach—should be considered a best-practice not only for automated text analyses, but for every list-based approach in the context of tracking data.

This study focused on non-use of news in the online context. However, on a societal level, having a realistic understanding of the extent of non-use of online news constitutes only one side of the coin. Future studies should integrate such evaluations with equally reliable measures for the offline context. Furthermore, future deliberations must also consider the issue that other sources than news or other institutions than journalism can have informative value regarding the political orientation of individuals (Moe and Ytre-Arne 2022, Swart et al. 2022), for example face-to-face communication (Peters et al. 2022, Reiss et al. 2021). This includes the theoretical untangling of what is considered of informative value, to whom, in what situation and to what end as well as developing and using methodological avenues that enable empirical investigations of these relatinships. Only a complete but nuanced and context-sensitive understanding allows a meaningful evaluation of the magnitude of non-use of news or non-information in society and for democracy and can function as a basis for potential policy measures.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Notes

1. In fact, one could argue that no recipient is entirely isolated from accidental exposure of (some form of) news or relevant information regarding their political orientation (Swart et al. 2022), making news use and avoidance of news rather a gradual and complex (Moe and Ytre-Arne 2022) than a dichotomous distinction.
2. Villi et al. (2022) seem to suggest the term news non-use as a substitute for unintentional news avoidance. In contrast, this study suggests non-use of news as a neutral observable phenomenon while reasons and motivations (intentional or unintentional) are a separate dimension.
3. See https://osf.io/sv8fj/?view_only=2f3b767fbca54363b45d92cdc1851a25 for the preregistration protocol and Appendix G for explanations on how the final study deviated from it.
4. The F1 score, a statistical analysis for binary classification, is the harmonic mean of the measures precision and recall. In this context, recall measures the share of news texts that are correctly classified as news text. Precision measures the share of correctly classified news texts out of all texts classified as news.
5. Supplemental material and data for this study is accessible via Online Appendix B.

Acknowledgments

The author wants to sincerely thank the three anonymous reviewers and the guest editors for their invested effort and the very constructive feedback and valuable comments that helped much in improving the article.

Funding

This work was funded by Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung.This research project received funding from the Swiss National Science Foundation (SNF Grant No. 176443).

ORCID

Michael V. Reiss http://orcid.org/0000-0002-6094-9985

References

Adoni, H., Z. Peruško, H. Nossek, and K. C. Schrøder. 2017. “Introduction: News Consumption as a Democratic Resource – News Media Repertoires across.” *Europe. Participations. Journal of Audience and Reception Studies* 14 (2): 226–252.

Araujo, T., A. Wonneberger, P. Neijens, and C. de Vreese. 2017. “How Much Time do You Spend Online? Understanding and Improving the Accuracy of Self-Reported Measures of Internet Use.” *Communication Methods and Measures* 11 (3): 173–190.

Bach, R. L., C. Kern, A. Amaya, F. Keusch, F. Kreuter, J. Hecht, and J. Heinemann. 2021. “Predicting Voting Behavior Using Digital Trace Data.” *Social Science Computer Review* 39 (5): 862–883.

Barabas, J, and J. Jerit. 2009. “Estimating the Causal Effects of Media Coverage on Policy-Specific Knowledge.” *American Journal of Political Science* 53 (1): 73–89.

Barberá, P., A. E. Boydstun, S. Linn, R. McMahon, and J. Nagler. 2021. “Automated Text Classification of News Articles: A Practical Guide.” *Political Analysis* 29 (1): 19–42.

Blekesaune, A., E. Elvestad, and T. Aalberg. 2012. “Tuning out the World of News and Current Affairs—An Empirical Study of Europe’s Disconnected Citizens.” *European Sociological Review* 28 (1): 110–126.

Boukes, M, and R. Vliegenthart. 2017. “News Consumption and Its Unpleasant Side Effect: Studying the Effect of Hard and Soft News Exposure on Mental Well-Being Over Time.” *Journal of Media Psychology* 29 (3): 137–147.

Boumans, J. W, and D. Trilling. 2016. “Taking Stock of the Toolkit: An Overview of Relevant Automated Content Analysis Approaches and Techniques for Digital Journalism Scholars.” *Digital Journalism* 4 (1): 8–23.

Budak, C., S. Goel, and J. M. Rao. 2016. “Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis.” *Public Opinion Quarterly* 80 (S1): 250–271.

Burscher, B., R. Vliegenthart, and C. H. De Vreese. 2015. “Using Supervised Machine Learning to Code Policy Issues: Can Classifiers Generalize across Contexts?” *The Annals of the American Academy of Political and Social Science* 659 (1): 122–131.

Ceron, A., L. Curini, and S. M. Iacus. 2015. “Using Sentiment Analysis to Monitor Electoral Campaigns: Method Matters—Evidence From the United States and Italy.” *Social Science Computer Review* 33 (1): 3–20.

Christians, C. G., T. L. Glasser, D. McQuail, K. Nordenstreng, and R. A. White. 2009. “Normative Theories of the Media: Journalism in Democratic Societies.” http://site.ebrary.com/id/10532349

Clegg, N. 2021. “You and the Algorithm: It Takes Two to Tango.” https://nickclegg.medium.com/you-and-the-algorithm-it-takes-two-to-tango-7722b19aa1c2

Coronel, S. S. 2003. “The Role of the Media in Deepening Democracy.” http://www.changecommunications.org/docs/dib/dibengunpan010194.pdf

de Bruin, K., Y. de Haan, R. Vliegenthart, S. Kruikemeier, and M. Boukes. 2021. “News Avoidance during the Covid-19 Crisis: Understanding Information Overload.” *Digital Journalism* 9 (9): 1286–1302.

de Vreese, C. H., and H. Boomgaarden. 2006. “News, Political Knowledge and Participation: The Differential Effects of News Media Exposure on Political Knowledge and Participation.” *Acta Politica* 41 (4): 317–341.
Dunaway, J., K. Searles, M. Sui, and N. Paul. 2018. “News Attention in a Mobile Era.” *Journal of Computer-Mediated Communication* 23 (2): 107–124.
Festic, N., M. Büchi, and M. Latzer. 2021. “How Long and What For? Tracking a Nationally Representative Sample to Quantify Internet Use.” *Journal of Quantitative Description: Digital Media* 1: 1–23.
Flaxman, S., S. Goel, and J. M. Rao. 2016. “Filter Bubbles, Echo Chambers, and Online News Consumption.” *Public Opinion Quarterly* 80 (S1): 298–320.
fög. 2020. Forschungsinstitut Öffentlichkeit und Gesellschaft. *Jahrbuch Qualität der Medien 2020 Schweiz*, Schwabe.
Gans, H. J. 2003. *Democracy and the News*. Oxford: Oxford University Press.
Gil de Zúñiga, H., N. Jung, and S. Valenzuela. 2012. “Social Media Use for News and Individuals’ Social Capital, Civic Engagement and Political Participation.” *Journal of Computer-Mediated Communication* 17 (3): 319–336.
Gil de Zúñiga, H., B. Weeks, and A. Ardèvol-Abreu. 2017. “Effects of the News-Finds-Me Perception in Communication: Social Media Use Implications for News Seeking and Learning About Politics: News Finds Me Perception.” *Journal of Computer-Mediated Communication* 22 (3): 105–123.
Grimmer, J, and B. M. Stewart. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” *Political Analysis* 21 (3): 267–297.
Guess, A. M. 2021. “(Almost) Everything in Moderation: New Evidence on Americans’ Online Media Diets.” *American Journal of Political Science* 65 (4): 1007–1022.
Guess, A. M., B. Nyhan, and J. Reifler. 2020. “Exposure to Untrustworthy Websites in the 2016 US Election.” *Nature Human Behaviour* 4 (5): 472–480.
Habermas, J. 1991. “The Public Sphere.” In *Rethinking Popular Culture: Contemporary Perspectives in Cultural Studies*, edited by C. Mukerji and M. Schudson, 398–404. Berkeley: University of California Press.
Haenschen, K. 2020. “Self-Reported Versus Digitally Recorded: Measuring Political Activity on Facebook.” *Social Science Computer Review* 38 (5): 567–583.
Haim, M., J. Breuer, and S. Stier. 2021. “Do News Actually “Find Me”? Using Digital Behavioral Data to Study the News-Finds-Me Phenomenon.” *Social Media + Society* 7 (3): 205630512110338.
Harcup, T, and D. O’Neill. 2017. “What is News?: News Values Revisited (Again).” *Journalism Studies* 18 (12): 1470–1488.
Hartmann, J., J. Huppertz, C. Schamp, and M. Heitmann. 2019. “Comparing Automated Text Classification Methods.” *International Journal of Research in Marketing* 36 (1): 20–38.
Holton, A. E, and H. I. Chyi. 2012. “News and the Overloaded Consumer: Factors Influencing Information Overload Among News Consumers.” *Cyberpsychology, Behavior and Social Networking* 15 (11): 619–624.
Joulin, A., E. Grave, P. Bojanowski, and T. Mikolov. 2016. “Bag of Tricks for Efficient Text Classification.” *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics* 2: 427–431.
Karlsen, R., A. Beyer, and K. Steen-Johnsen. 2020. “Do High-Choice Media Environments Facilitate News Avoidance? A Longitudinal Study 1997-2016.” *Journal of Broadcasting & Electronic Media* 64 (5): 794–814.
Kim, Y., H.-T. Chen, and H. Gil de Zúñiga. 2013. “Stumbling upon News on the Internet: Effects of Incidental News Exposure and Relative Entertainment Use on Political Engagement.” *Computers in Human Behavior* 29 (6): 2607–2614.
Krippendorff, K. 2004. *Content Analysis. An Introduction to Its Methodology*. Thousand Oaks: SAGE.
Kümpel, A. S. 2022. “Using Messaging Apps in Audience Research: An Approach to Study Everyday Information and News Use Practices.” *Digital Journalism* 10 (1): 188–199.

Latzer, M., M. Festic Noemi, and K. Kappeler. 2020. *Use and Assigned Relevance of Algorithmic-Selection Applications in Switzerland. Report 1 from the Project: The Significance of Algorithmic Selection for Everyday Life: The Case of Switzerland*. Universität Zürich.

Lee, H, and J. Yang. 2014. "Political Knowledge Gaps Among News Consumers with Different News Media Repertoires Across Multiple Platforms." *International Journal of Communication* 8: 597–617.

Leeper, T. J. 2020. "Raising the Floor or Closing the Gap? How Media Choice and Media Content Impact Political Knowledge." *Political Communication* 37 (5): 719–740.

Lindell, J. 2018. "Distinction Recapped: Digital News Repertoires in the Class Structure." *New Media & Society* 20 (8): 3029–3049.

Makhortykh, M., A. Urman, T. Gil-Lopez, and R. Ulloa. 2021. "To Track or Not to Track: Examining Perceptions of Online Tracking for Information Behavior Research." *Internet Research* 32 (7): 260–279.

Mangold, F., S. Stier, J. Breuer, and M. Scharkow. 2021. "The Overstated Generational Gap in Online News Use? A Consolidated Infrastructural Perspective." *New Media & Society*. DOI: https://doi.org/10.1177/1461444821989972.

Moe, H, and B. Ytre-Arne. 2022. "The Democratic Significance of Everyday News Use: Using Diaries to Understand Public Connection over Time and beyond Journalism." *Digital Journalism* 10 (1): 43–61.

Möller, J., R. N. van de Velde, L. Merten, and C. Puschmann. 2020. "Explaining Online News Engagement Based on Browsing Behavior: Creatures of Habit?" *Social Science Computer Review* 38 (5): 616–632.

Nelson, J. L. 2020. "The Persistence of the Popular in Mobile News Consumption." *Digital Journalism* 8 (1): 87–102.

Nelson, L. K., D. Burk, M. Knudsen, and L. McCall. 2021. "The Future of Coding: A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods." *Sociological Methods & Research* 50 (1): 202–237.

NET-Metrix. 2018. "NET-Metrix-Profile: Online Reichweite 2018-2." https://netreport.net-metrix.ch/profile/index.php?l=de

Newman, N., R. Fletcher, A. Kalogeropoulos, and R. K. Nielsen. 2019. "Reuters Institute Digital News Report 2019." Reuters Institute for the Study of Journalism.

Newman, N., R. Fletcher, A. Schulz, S. Andi, and R. Nielsen. 2020. "Reuters Institute Digital News Report 2020." Reuters Institute for the Study of Journalism.

Palmer, R., B. Toff, and R. K. Nielsen. 2020. "The Media Covers Up a Lot of Things: Watchdog Ideals Meet Folk Theories of Journalism." *Journalism Studies* 21 (14): 1973–1989.

Parry, D., B. I. Davidson, C. Sewall, J. T. Fisher, H. Mieczkowski, and D. S. Quintana. 2021. "A Systematic Review and Meta-Analysis of Discrepancies Between Logged and Self-Reported Digital Media Use." Nature Human Behaviour *5: 1535–1547.*

Pedregosa, F., Varoquaux, G. Gramfort, Al, Michel, V. Thirion, B. Grisel, O. Blondel, M. Prettenhofer, P, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–2830.

Peters, C., K. C. Schrøder, J. Lehaff, and J. Vulpius. 2022. "News as They Know It: Young Adults' Information Repertoires in the Digital Media Landscape." *Digital Journalism* 10 (1): 62–86.

Prior, M. 2005. "News vs. Entertainment: How Increasing Media Choice Widens Gaps in Political Knowledge and Turnout." *American Journal of Political Science* 49 (3): 577–592.

Prior, M. 2009. "The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure." *Public Opinion Quarterly* 73 (1): 130–143.

Reiss, M. V., N. Festic, M. Latzer, and T. Rüedy. 2021. "The Relevance Internet Users Assign to Algorithmic-Selection Applications in Everyday Life." *Studies in Communication Sciences* 21 (1): 71–90.

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.

Scharkow, M. 2019. "The Reliability and Temporal Stability of Self-Reported Media Exposure: A Meta-Analysis." *Communication Methods and Measures* 13 (3): 198–211.

Schwaiger, L. 2022. *Gegen die Öffentlichkeit: Alternative Nachrichtenmedien im deutschsprachigen Raum*. transcript.
Skovsgaard, M, and K. Andersen. 2020. "Conceptualizing News Avoidance: Towards a Shared Understanding of Different Causes and Potential Solutions." *Journalism Studies* 21 (4): 459–476.
Skovsgaard, M., A. Shehata, and J. Strömbäck. 2016. "Opportunity Structures for Selective Exposure: Investigating Selective Exposure and Learning in Swedish Election Campaigns Using Panel Survey Data." *The International Journal of Press/Politics* 21 (4): 527–546.
Song, H., J. Jung, and Y. Kim. 2017. "Perceived News Overload and Its Cognitive and Attitudinal Consequences for News Usage in South Korea." *Journalism & Mass Communication Quarterly* 94 (4): 1172–1190.
Stier, S., J. Breuer, P. Siegers, and K. Thorson. 2020. "Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field." *Social Science Computer Review* 38 (5): 503–516.
Stier, S., N. Kirkizh, C. Froio, and R. Schroeder. 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.
Street, J. 2011. *Mass Media, Politics and Democracy*. 2nd ed. Basingstoke: Palgrave Macmillan.
Strömbäck, J., M. Djerf-Pierre, and A. Shehata. 2013. "The Dynamics of Political Interest and News Media Consumption: A Longitudinal Perspective." *International Journal of Public Opinion Research* 25 (4): 414–435.
Stroud, N. J. 2008. "Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure." *Political Behavior* 30 (3): 341–366.
Stroud, N. J. 2010. "Polarization and Partisan Selective Exposure." *Journal of Communication* 60 (3): 556–576.
Swart, J., T. Groot Kormelink, I. Costera Meijer, and M. Broersma. 2022. "Advancing a Radical Audience Turn in Journalism. Fundamental Dilemmas for Journalism Studies." *Digital Journalism* 10 (1): 8–22.
Toff, B, and A. Kalogeropoulos. 2020. "All the News That's Fit to Ignore." *Public Opinion Quarterly* 84 (S1): 366–390.
Toff, B, and R. A. Palmer. 2019. "Explaining the Gender Gap in News Avoidance: "News-Is-for-Men" Perceptions and the Burdens of Caretaking." *Journalism Studies* 20 (11): 1563–1579.
Trilling, D, and K. Schoenbach. 2013. "Skipping Current Affairs: The Non-Users of Online and Offline News." *European Journal of Communication* 28 (1): 35–51.
Trilling, D., P. Tolochko, and B. Burscher. 2017. "From Newsworthiness to Shareworthiness: How to Predict News Sharing Based on Article Characteristics." *Journalism & Mass Communication Quarterly* 94 (1): 38–60.
Udris, L, and M. Eisenegger. 2019. "Digital News Report—Switzerland." https://www.digitalnews-report.org/survey/2019/switzerland-2019/
Valenzuela, S., I. Bachmann, and M. Aguilar. 2019. "Socialized for News Media Use: How Family, Communication, Information-Processing Needs, and Gratification Determine Adolescents' Exposure to News." *Communication Research* 46 (8): 1095–1118.
Van Aelst, P., J. Strömbäck, T. Aalberg, F. Esser, C. de Vreese, J. Matthes, D. Hopmann, et al. 2017. "Political Communication in a High-Choice Media Environment: A Challenge for Democracy?" *Annals of the International Communication Association* 41 (1): 3–27.
van Atteveldt, W., M. A. C. G. van der Velden, and M. Boukes. 2021. "The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms." *Communication Methods and Measures* 15 (2): 121–140.
Vermeer, S., D. Trilling, S. Kruikemeier, and C. de Vreese. 2020. "Online News User Journeys: The Role of Social Media, News Websites, and Topics." *Digital Journalism* 8 (9): 1114–1141.
Villi, M., T. Aharoni, K. Tenenboim-Weinblatt, P. J. Boczkowski, K. Hayashi, E. Mitchelstein, A. Tanaka, and N. Kligler-Vilenchik. 2022. "Taking a Break from News: A Five-Nation Study of News Avoidance in the Digital Era." *Digital Journalism* 10 (1): 148–164.

Vraga, E. K, and M. Tully. 2020. “Who Is Exposed to News? It Depends on How You Measure: Examining Self-Reported Versus Behavioral News Exposure Measures.” *Social Science Computer Review* 38 (5): 550–566.

Weichselbraun, A. 2021. “Inscriptis—A Python-Based HTML to Text Conversion Library Optimized for Knowledge Extraction from the Web.” *Journal of Open Source Software* 6 (66): 3557.

WEMF. 2019. “Mach Basic—Titelauflistung.” https://wemf.ch/de/downloads/studien/mach-basic/zusammensetzung-der-titelkombinationen-und-gesamtausgaben.pdf

Wolfsfeld, G., M. Yarchi, and T. Samuel-Azran. 2016. “Political Information Repertoires and Political Participation.” *New Media & Society* 18 (9): 2096–2115.

Woodstock, L. 2014. “The News-Democracy Narrative and the Unexpected Benefits of Limited News Consumption: The Case of News Resisters.” *Journalism* 15 (7): 834–849.