Article

More diverse, more politically varied: How social media, search engines and aggregators shape news repertoires in the United Kingdom

Richard Fletcher
University of Oxford, UK

Antonis Kalogeropoulos
University of Liverpool, UK

Rasmus Kleis Nielsen
University of Oxford, UK

Abstract
There is still much to learn about how the rise of new, 'distributed', forms of news access through search engines, social media and aggregators are shaping people's news use. We analyse passive web tracking data from the United Kingdom to make a comparison between direct access (primarily determined by self-selection) and distributed access (determined by a combination of self-selection and algorithmic selection). We find that (1) people who use search engines, social media and aggregators for news have more diverse news repertoires. However, (2) social media, search engine and aggregator news use is also associated with repertoires where more partisan outlets feature more prominently. The findings add to the growing evidence challenging the existence of filter bubbles, and highlight alternative ways of characterizing people's online news use.

Corresponding author:
Richard Fletcher, Reuters Institute for the Study of Journalism, University of Oxford, 13 Norham Gardens, Oxford OX2 6PS, UK.
Email: richard.fletcher@politics.ox.ac.uk

The growth of 'distributed' news access through search engines, social media and news aggregators, constitutes one of the most important recent changes to online information environments. However, this growth has been accompanied by concerns over the potential for people to be trapped inside echo chambers and filter bubbles that narrow their exposure to news and information (Pariser, 2011; Sunstein, 2017).

Researchers have responded by exploring the extent to which people experience cross-cutting news exposure – where they are exposed to views that oppose their own – on a variety of different platforms (Bakshy et al., 2015; Barberá et al., 2015; Cardenal et al., 2019; Flaxman et al., 2016). Although this work has not typically found strong evidence of filter bubbles or echo chambers for most people (Barberá, 2020; Bruns, 2019), this focus on cross-cutting news exposure – with its emphasis on the democratic value of exposure to content that challenges, opposes and provokes – has emerged as the dominant way of understanding the effects of distributed news access. However, measuring the extent to which news recommender systems provide users with critical perspectives through cross-cutting exposure is just one way of assessing distributed news use (Helberger, 2019). We might also ask, for example, about the extent to which these systems leave users with *diverse* news repertoires that contain a broad but balanced range of perspectives and sources (Loecherbach et al., 2020).

To better understand the effect of distributed news use on the diversity of people's news repertoires, we analyse desktop and laptop web tracking data (collected by YouGov) that recorded the web use of a representative sample of 3071 UK news users for 1 month in 2017. The dataset contains information about every visit made to selected news outlets during this period, as well as data on how that visit was initiated. By combining measures of diversity used in communications research (McDonald and Dimmick, 2003) and audience-based measures of news outlet slant (Flaxman et al., 2016), we compare the diversity of individual news repertoires of people with different degrees of reliance on direct access, search engines, Facebook, Twitter and Google News. This approach allowed us to analyse more fine-grained data on people's news repertoires, avoid problems associated with recall, and analyse social media, search engine and aggregator use in the context of other online behaviours.

This approach also allowed us to explore in detail the role of distributed news use in a media system that is fundamentally different from the United States – the focus of most current research on the topic. As we will see, in the United Kingdom by far the most widely used news outlet is the BBC, a public service media organization that is required to be both independent and impartial. At the same time, the United Kingdom has a highly partisan press meaning that even small changes to news repertoires as a result of distributed news use can be consequential.

Literature review

By 2018, around two-thirds of people across 37 media markets said their main way of accessing online news was through 'distributed' access (Newman et al., 2018). This term differentiates between direct access (i.e. going straight to the homepages and apps of branded news outlets), and arriving at news through platforms, for example, by

clicking through to a news story from a search results page, a social media feed or a news aggregator.

The reason the distinction between direct news access and distributed news access is important is because direct access is primarily governed by self-selection, whereas distributed access – to varying degrees depending on the specific circumstances – combines self-selection with algorithmic selection in the form of news recommenders (Zuiderveen Borgesius et al., 2016). Some find the use of news recommenders concerning, in part because of a lack of transparency about how they work (Pasquale, 2015), and in part because they help cement the power of big platform companies (Moore and Tambini, 2018) – but also because it is possible that algorithmic selection will be so responsive to people's interests and desires that it will narrow their news and information exposure.

'Filter bubbles' and 'echo chambers' are the terms often used as shorthand for this concern. There is a lack of clarity about the precise meaning of these terms, and how they differ from one another (if at all), but the basic idea underpinning both is that algorithmic selection will draw on data about past behaviour to reinforce preference-based consumption patterns. This, the theory goes, could create 'echo chambers', where media users tend to be over-exposed to like-minded perspectives, and 'filter bubbles', where counter-attitudinal information is filtered out.

In response, empirical studies have looked for filter bubbles and echo chambers by measuring the extent to which people experience cross-cutting news exposure on different platforms (Bakshy et al., 2015; Barberá et al., 2015). Cross-cutting news exposure simply refers to situations where people come into contact with news and information that runs counter to their own views – something that has long been seen as good for democracy.

So far, most studies have not found evidence of strong filter bubble and echo chamber effects (Barberá, 2020; Bruns, 2019), with some even suggesting that distributed news use increases cross-cutting news exposure. Flaxman et al. (2016) used data collected from Bing toolbar users in the United States to compare the level of cross-cutting exposure from direct access to that from search engines, social media and news aggregators. They found that distributed access resulted in a higher degree of exposure to news from cross-cutting outlets, suggesting that algorithmic selection is less closely aligned with partisan preferences than self-selection. Similarly, using tracking data from Spain, Cardenal et al. (2019) measured the probability that each news visit in the data would be to an ideologically aligned outlet, and found that the probability was higher than average for direct news accesses and lower when people used Google Search.

One possible explanation for these findings is that many automated recommender systems are designed to occasionally surface random content, because the engineers that design them know that people can get bored by predictable results (Kotkov et al., 2016). This means that when people use platforms like search engines and aggregators for news, they may experience 'automated serendipity', where they are shown news from outlets they would not normally use (Fletcher and Nielsen, 2018b). A second reason is that some platforms, particularly social networks like Facebook, Twitter and YouTube, can drive incidental exposure to news (Boczkowski et al., 2018; Fletcher and Nielsen, 2018a; Valeriani and Vaccari, 2016; Weeks et al., 2017). Incidental exposure, which can occur in a wide variety of different online and offline situations, happens when people come into

contact with news while primarily intending to do something else (Tewksbury et al., 2001). Relatively, few people log on to social networks just to look at the news, but because many people's networks consist of weak ties to people with different interests and persuasions, they might still be exposed to news content shared by others.

Diversity

Cross-cutting news exposure is an important dimension of people's news repertoires, but it is not the only way of characterizing them. Natali Helberger (2019) has outlined four democratic roles that algorithmic news recommenders could fulfil – 'liberal', 'participatory', 'deliberative' and 'critical' – and the types of news and information they should surface in order to do so. This typology offers a useful starting point for thinking about other ways of characterizing distributed news use and the news repertoires they enable. Recommenders arguably need to provide at least some degree of cross-cutting exposure in order to partially fulfil all four roles, but the 'deliberative' role, in particular, places a stronger emphasis on *diversity* through surfacing the full range of available sources and perspectives, with no special emphasis on news content that provokes, opposes and challenges the user's views.

In the context of media and communication research, the broad concept of diversity has been applied in many different ways. In particular, it has been used to characterize media content, the structure of the media environment as a whole and – as is our focus here – media exposure (Loecherbach et al., 2020). According to Napoli (1999: 26), exposure diversity can be understood horizontally through 'the distribution of audiences across all available content options', or vertically through 'the diversity of content consumption within individual audience members'. Recent studies of audience fragmentation suggest a high degree of horizontal diversity in many media systems (Fletcher and Nielsen, 2017; Webster and Ksiazek, 2012), but the vertical diversity of people's repertoires is less well understood.

Previous work has found that distributed news use is associated with accessing a greater *variety* of different news outlets – that is, the number of different news outlets in a repertoire. Studies based on survey and tracking data across several countries have shown that those who use social media, search engines and news aggregators have news repertoires that consist of more news outlets (Fletcher and Nielsen, 2018a, 2018b; Scharkow et al., 2020). However, variety is one of the simplest ways of measuring news repertoire diversity (Loecherbach et al., 2020), and can mask patterns of news use that are heavily skewed towards particular outlets. Therefore, it is difficult to use measures of variety to confidently conclude that news recommenders are really leaving their users with the level of diversity required to fulfil a 'deliberative' role (Helberger, 2019). Therefore, our aim here is to build on existing work by using measures of news repertoire diversity that also capture the *balance* of people's news use across a broad range of outlets.

News use in the United Kingdom

Comparative research has concluded that the United Kingdom has a 'liberal' media system, characterized by a high degree of commercialization, strong professionalization – and

Table 1. Tracked articles read by news outlet.

Outlet	Articles read (N)	Articles read (%)
bbc.co.uk/news	69,401	41.1
theguardian.com	20,825	12.3
msn.com	19,113	11.3
dailymail.co.uk	17,673	10.5
buzzfeed.com	8029	4.8
telegraph.co.uk	6354	3.8
huffingtonpost.co.uk	3787	2.2
mirror.co.uk	3668	2.2
independent.co.uk	3644	2.2
express.co.uk	3621	2.1
thesun.co.uk	2875	1.7
metro.co.uk	2789	1.7
news.sky.com	1915	1.1
thetimes.co.uk	1869	1.1
standard.co.uk	1067	0.6
uk.news.yahoo.com	763	0.5
itv.com/news	552	0.3
thecanary.co	335	0.2
breitbart.com	319	0.2
theladbible.com	234	0.1
channel4.com/news	24	<0.01
Total	168,857	100.0

somewhat uniquely – a highly partisan press alongside strong and well-funded public broadcasting (Hallin and Mancini, 2004). However, to better understand the dynamics of online news use, we can examine descriptive statistics from the dataset we analyse here. We include this analysis as part of the literature review because similar analysis has been published previously (Newman and Kalogeropoulos, 2017). These statistics take news outlets as the unit of analysis, but will be useful later for interpreting the individual-level data. See the 'Data' and 'Measures' sections below for more information on how this was done.

Table 1 displays the number of news articles read during a 1-month period in 2017 across 21 popular online news outlets (identified using subdomains or directories) in the United Kingdom (Newman et al., 2017).[1] Over the course of the tracking period, a representative sample of 3071 news users read 168,857 news articles across the 21 outlets.[2] It is important to point out here that we use the term 'news articles' broadly to refer to many of the other types of content that news outlets produce, such as sports coverage, opinion pieces, reviews and other cultural or lifestyle content, that are usually packaged in the form of articles.

These descriptive statistics highlight one of the most important features of the United Kingdom's online news landscape; the prominence of the BBC News website. Around 40% of all news articles read on desktop or laptop computers are from the BBC, with the

Table 2. Tracked articles read by access mode.

Mode	Articles accessed (N)	Articles accessed (%)
Direct	130,248	77.1
Facebook	9433	5.6
Search	9301	5.5
Twitter	2271	1.4
Google News	826	0.5
Other	16,778	9.9
Total	168,857	100.0

Guardian, MSN News and MailOnline (dailymail.co.uk), the only other outlets that account for more than a 10% share. None of the other 17 outlets account for more than 5% each, meaning that the five most-used outlets account for around 80% of all news articles. This level of inequality among news outlets is unusual, at least in relatively free media systems. Comparative data are rare, but one study from the United States found that the most popular news outlet accounted for 30% of news visits, and the top five outlets made up 64% of the total (Guess, 2021). More generally, country-by-country lists of the news outlets with the largest online weekly reach are almost never so heavily skewed towards a single provider (Newman et al., 2019).

The second notable feature of the United Kingdom online news environment is that a large percentage of news articles are accessed by people going directly to news websites rather than distributed access through social media, search engines, and news aggregators (see Table 2). Around 77% of news articles are accessed directly – when someone clicks onto a news article after navigating to a publisher's homepage, or from another news article on the same website – compared to around 5% each through Facebook and search engines. Twitter and Google News are relatively marginal, with less than 2% combined. Of course, people can arrive at news in many other ways, such as through email, messaging apps and other websites, and this accounts for around 10% of news visits in the United Kingdom. It is important to note that, in the analysis presented here, if someone arrives at a news article through distributed access, but then reads another article on the same website, this second article (and those thereafter) is coded as having been accessed directly. An alternative approach would be to consider all subsequent articles to be part of one news 'session', characterized by how the first article was arrived at. Therefore, alternative descriptive statistics for news sessions can be found in the Supplemental material.

This reliance on direct access – which is probably linked to the prominence of the BBC – is not something we see in every country. Survey data suggest it can be found in countries like Finland, Norway and Sweden, where most people say that direct access in their main way of arriving at news, but elsewhere, we typically see markets that are either heavily reliant on social media (e.g. Chile, Brazil and Malaysia), search and aggregators (e.g. Japan and South Korea) or a relatively even mix of all three (e.g. the United States, Canada and Australia; Newman et al., 2019).

Table 3. Articles read from each outlet by access mode.

Outlet	Direct		Facebook		Search		Twitter		Google News	
	N	%	N	%	N	%	N	%	N	%
bbc.co.uk/news	64,111	49.2	1294	13.7	1019	11.6	374	16.5	157	19.0
theguardian.com	14,905	11.4	1268	13.4	1529	16.3	541	23.8	131	15.9
dailymail.co.uk	14,188	10.9	561	6.0	1265	13.5	175	7.7	107	13.0
msn.com	14,482	11.1	436	4.6	117	1.3	21	0.9	1	0.1
buzzfeed.com	6227	4.8	758	8.0	121	1.3	87	3.8	0	0
telegraph.co.uk	3403	2.6	358	3.8	1309	14	230	10.1	72	8.7
mirror.co.uk	1688	1.3	656	7.0	727	7.8	94	4.1	113	13.7
independent.co.uk	941	0.7	1111	11.8	643	6.9	219	9.6	52	6.3
express.co.uk	1872	1.4	364	3.9	565	6.0	38	1.7	79	9.6
thesun.co.uk	979	0.8	688	7.3	816	8.7	67	3.0	19	2.3
metro.co.uk	1647	1.3	437	4.6	384	4.1	45	2.0	17	2.1
huffingtonpost.co.uk	1644	1.3	397	4.2	302	3.2	67	3.0	10	1.2
news.sky.com	1589	1.2	99	1.1	65	0.7	44	1.9	10	1.2
thetimes.co.uk	1463	1.1	68	0.7	24	0.3	61	2.7	1	0.1
standard.co.uk	446	0.3	156	1.7	205	2.2	70	3.1	40	4.8
uk.news.yahoo.com	351	0.3	63	0.7	104	1.1	2	0.1	2	0.2
itv.com/news	87	0.1	249	2.6	91	1.0	59	2.6	14	1.7
thecanary.co	13	<0.1	265	2.8	1	<0.1	45	2.0	1	0.1
breitbart.com	189	0.2	16	0.2	7	0.1	14	0.6	0	0
theladbible.com	10	<0.1	187	2.0	2	<0.1	16	0.7	0	0
channel4.com/news	13	<0.1	2	<.1	5	0.1	2	0.1	0	0
Total	130,248	100.0	9433	100.0	9301	100.0	2271	100.0	826	100.0

If we combine the data from Tables 1 and 2 and look at the number of news visits to each outlet by access mode, we see important variation (see Table 3). While direct access is dominated by BBC News (accounting for nearly half of all articles read), the same is not true for Facebook, Twitter, search and Google News access. Here, news use is spread more evenly across outlets, with the BBC accounting for no more than 20%. Indeed, for some modes of distributed access, the BBC is not even the most-used outlet. Both the Guardian and the Telegraph receive more visits through search engines than the BBC, and the Guardian receives more through social media (Facebook and Twitter combined). Also, if Table 3 is read horizontally, it is clear how some outlets, such as ITV News, are particularly dependent on various forms of distributed access, whereas others like The Times – because of its hard paywall – rely on direct access.

Hypotheses

On the basis of previous research, both on news use in the United Kingdom and diversity, we can formulate a number of hypotheses about the relationship between how people

access news online news and their news repertoires. Thinking first about diversity, because previous research has shown that algorithmic selection can be a source of incidental exposure and automated serendipity, it seems likely that the more people use distributed access, the more diverse their news repertoires will be. At the same time, because direct access relies on self-selection, and because a large proportion of direct accesses are channelled towards a small number of news outlets, we expect that the more people use direct access, the less diverse their news repertoires will be:

> *H1a.* Direct news access is associated with lower news repertoire diversity.

> *H1b.* Distributed news access is associated with higher news repertoire diversity.

Because concern over the effects of distributed news use often focuses on politics, we also investigate political diversity specifically. To do this, we group together the use of articles from all relatively right-leaning outlets, all relatively left-leaning outlets, and outlets with no particular slant, and measure the balance of news use across these three groups. If outlets are grouped together in this way, people that, for example, *only* use a range of different right-leaning outlets, will have repertoires with low political diversity no matter how balanced their news use is across those outlets. Whereas those with news use spread evenly across right-leaning and left-leaning outlets will have repertoires with higher political diversity. Because previous research has found that distributed access is associated with cross-cutting news exposure, we expect that the more people use it the more politically diverse their news repertoires will be:

> *H2a.* Direct news access is associated with lower political news repertoire diversity.

> *H2b.* Distributed news access is associated with higher political news repertoire diversity.

If distributed news use is affecting the diversity of people's news repertoires, it is likely that their repertoires will change in other measurable ways as a result. For some, an increase in news repertoire diversity might necessarily require the use of outlets that are more partisan (i.e. more left-leaning or more right-leaning). There is already research from the United States to suggest that news from more partisan outlets is disproportionately amplified on Twitter (Hasell, 2020), and research from Germany finds that exposure to partisan media is associated with social media use (Müller and Schulz, 2019). But more importantly for the UK context, because we have already seen that so much direct news traffic goes to the BBC, and because the BBC is required to provide impartial news coverage, we expect that the more people rely on direct access, the less likely that partisan outlets will feature prominently in their news repertoires. At the same time, it is also clear that the UK media system as a whole is characterized by a particularly partisan press, meaning that increased use of distributed access will pull people away from the BBC and push them towards more partisan sources. Therefore, we also hypothesize that the more people use distributed access, the more likely that more partisan outlets will feature prominently in their repertoires. To be clear, this does not necessarily mean that

repertoires will be skewed towards a particular ideology, but rather that they will be more likely to contain a mixture of outlets that are closer to various ideological extremes.

> *H3a.* Direct news access is associated with news repertoires where more partisan outlets are less prominent.

> *H3b.* Distributed news access is associated with news repertoires where more partisan outlets are more prominent.

Data

We address these hypotheses using passive web tracking data. The data come from YouGov's Pulse panel, which at the time, was made up of 13,709 people in total, and 6811 active users. We tracked every visit from a desktop or laptop to 21 of the most popular UK news websites for a 1-month period from 13 March to 10 April 2017. During this time, 3071 panellists accessed at least one news article. For each, the previous URL was also recorded so that it was possible to infer how the panellist accessed it. Demographic data were provided by panellists upon joining, allowing YouGov to assemble a nationally representative panel in terms of age, gender and region. Panellists were offered incentives in return for installing Wakoopa tracking software. The software passively tracked each panellist's Internet use, but they were free to pause or uninstall it at any time.

Commercially available tracking data are typically aggregated at the subdomain level (e.g. the proportion of Internet users that visited a domain in a particular month; Taneja, 2016), but data collected by the YouGov Pulse panel contains web visits by specific panellists, allowing it to be analysed in much the same way as survey data. If used in this way, passive tracking has some clear advantages. Numerous studies have shown that people struggle to accurately recall their recent news use when asked about it later in surveys (Guess, 2015; Prior, 2009) – particularly when they have arrived at news through search engines and social media (Kalogeropoulos et al., 2019). However, tracking software automatically records every URL accessed by panellists in real time, so recall is not required. Relatedly, it is also possible to use passive tracking to obtain more fine-grained information. A survey can be used to ask people on how many of the last 7 days they have used a particular news outlet, but tracking data can record hundreds of specific news visits, and do so reliably over longer periods of time. This is particularly important for measures of diversity, because data from surveys may simply be too sparse for reliable estimates. There is also a particular advantage associated with this dataset, in that it contains individual-level demographic data about the age, gender, education and social grade of the panellists (social grade is explained below). These can be useful for both generating new measurements and for model building.

There are, however, some disadvantages associated with the tracking data. First, it is not drawn from a random sample. Quota sampling was used to build a panel that is representative of the UK population in terms of age group, gender and region. But even so, because people opt-in to the panel, it is likely biased towards particular groups, attitudes

and dispositions that in turn are associated with particular patterns of news use. Second, decisions about what information would usefully complement tracked web use were made by YouGov in advance of the study, and independently of the authors. Because the dataset did not contain information about individual political preferences, we draw on data from other sources (see 'Measures' section). Finally, and perhaps most importantly, in common with much research based on tracking data, only desktop or laptop use was tracked, and not news use on smartphones or tablets. Given that recent research has shown that mobile news attention differs from desktop or laptop attention (Dunaway et al., 2018), these findings cannot necessarily be extended to the world of mobile news.

Measures

We use three dependent variables. The first is diversity. In order to move beyond previous work that has measured diversity through variety, or the number of news outlets in people's repertoires, we instead measure the balance of people's news use across different news outlets.

Outside of media and communication research, diversity is measured in a wide range of different disciplines, and measures of diversity have been applied in many different contexts. Some of the most widely applied measures come from ecology, where scientists have attempted to characterize environments by the diversity of life they contain. Many measures of diversity are 'dual-concept', because they see diversity as a two-dimensional construct based on the number of categories present and how elements are distributed across those categories (McDonald and Dimmick, 2003). Here, we take news outlets to be the categories, and news articles as the elements.

We present two different measures of diversity: Simpson's *D* and Shannon's *H*. Simpson's *D* (1949) is one of the most widely used measures of diversity. It has been used by media and communication researchers to measure content diversity, for example comparing the issue diversity of online newspapers and citizen journalists by looking at the number of articles across different topics (Carpenter, 2010), or to measure the agenda diversity in the *New York Times* over time by looking at the number of articles across different issues (Tan and Weaver, 2013). Simpson's *D* can be expressed as follows:

$$D = 1 - \sum_i n_i (n_i - 1) / N(N - 1)$$

For our purposes, N refers to the total number of news outlets (categories) and n refers to the number of news articles (elements) for each outlet. This produces a figure between 0 and 1 inclusive, with higher values indicating greater diversity. One advantage if Simpson's *D* is that it has an intuitive interpretation, namely that D is the probability that any two randomly selected articles will be from different outlets. We applied this measure to each panellist's news use over the tracking period ($M = 0.55$, $SD = 0.35$).

Our second measure of diversity is Shannon's H (1949), which is expressed as follows:

$$H = -\sum p_i \ln p_i$$

Table 4. Relative outlet slant.

Outlet	Relative slant	Category (more _____ leaning)
breitbart.com	0.94	Right
express.co.uk	0.44	Right
dailymail.co.uk	0.36	Right
thesun.co.uk	0.32	Right
thetimes.co.uk	0.32	Right
telegraph.co.uk	0.30	Right
news.sky.com	0.20	–
msn.com	0.20	–
itv.com/news	0.10	–
metro.co.uk	0.04	–
standard.co.uk	0.04	–
bbc.co.uk/news	0.00	–
uk.news.yahoo.com	−0.04	–
theladbible.com	−0.12	–
channel4.com/news	−0.18	–
mirror.co.uk	−0.26	Left
huffingtonpost.co.uk	−0.26	Left
buzzfeed.com	−0.34	Left
theguardian.com	−0.34	Left
independent.co.uk	−0.38	Left
thecanary.co	−0.70	Left

Here, p is the proportion of articles from an outlet in a repertoire. Shannon's H is less widely used in media and communication research, in part because it cannot be straightforwardly interpreted. However, it is useful here because it is more sensitive to the number of outlets in a repertoire. Whereas Simpson's D will produce a score of 1 for a repertoire that has two articles from two different outlets, and for a repertoire that has four articles from four different outlets (in both examples, it is certain that any two randomly selected articles will be from different outlets), Shannon's H deems the latter to be more diverse even though both are equally balanced in this narrow sense. This means it is better able to characterize the repertoires of infrequent news users. Because repertoires can contain a maximum number of 21 different outlets, there is also a maximum H, which in turn allows us to rescale the scores to range between 0 and 1 inclusive ($M=0.25$, $SD=0.20$).

The second dependent variable estimates the political diversity of people's news repertoires or the balance of people's news use across outlets grouped by their political slant. These groups are defined using audience-based estimates of news outlet slant (Flaxman et al., 2016; Fletcher and Nielsen, 2018b; Gentzkow and Shapiro, 2011). The audience-based approach uses information about the political preferences of an outlet's audience to estimate its relative political slant. We rely on estimates produced by Fletcher, Cornia, and Nielsen (2020). These outlet slant scores were computed by first taking the average (mean) left-right political leaning of a nationally representative UK survey

sample, and then computing the difference between that and the average political leaning of the cross-platform audience for each outlet (see Table 4). Outlets with a slant score of 0 have an audience that very closely resembles the population, whereas outlets with a negative score have a more left-leaning audience, and outlets with a positive score a more right-leaning audience. It is important to note that these are estimates of *relative* slant, and because they refer to differences between audiences and the population, can thus be used to identify outlets that are more left-leaning or right-leaning (and hence, more partisan) than others. However, for the same reasons, this method cannot necessarily be used to identify outlets that are left-leaning or right-leaning in an absolute sense (Guess, 2021).

Outlets with a score of above 0.20 were grouped as more right-leaning, less than –0.20 more left-leaning, and all others classed as having no particular slant. The groups are based on arbitrary cut-offs, but it does mean that broadcasters – which are required to be impartial in the United Kingdom – are placed in the no slant category alongside news aggregators like Yahoo! and MSN, and commuter newspapers like *Metro* and the *Evening Standard*. At the same time, newspapers that often openly endorse certain political parties during elections are placed in the appropriate more left-leaning or more right-leaning group. Both Simpson's D ($M = 0.43$, $SD = 0.33$) and Shannon's H ($M = 0.15$, $SD = 0.12$) can be used to measure the balance of articles from right-leaning outlets, left-leaning outlets and outlets with no particular slant in the same way as for overall diversity. However, it is important to be clear that, in common with previous research (Flaxman et al., 2016), the slant of individual articles is directly inferred from estimates of relative outlet slant. This means that the use of the small number of left-leaning articles that appear in right-leaning outlets (and vice versa), and articles that are apolitical, could be miscoded.

We use the same outlet slant scores for our third dependent variable, which measures the prominence of more partisan outlets in people's news repertoires. Because here we are interested in the degree to which people's repertoires contain outlets that are closer to the ideological extremes rather than the degree to which they are skewed towards a particular ideology, we use the absolute mean slant score of all the visits made to each outlet during the tracking period ($M = 0.20$, $SD = 0.12$; for a similar approach, see Guess, 2021).

Before introducing our independent variables, it is important to make clear that, in common with other studies (e.g. Flaxman et al., 2016), we excluded panellists that accessed fewer than five news articles during the tracking period because data sparsity would produce unreliable measures of diversity. This reduced the number of panellists from 3071 to 2232 in the subsequent regression analysis (descriptive statistics are based on the 3071 news users). In extreme cases, zeros are produced by Simpson's D if a panellist only used one source of news during the entire tracking period, and ones are produced if a panellist used a different source of news every time (usually very infrequent news users that, for example, used BBC News online once and the Guardian online once). Such news use patterns are rarely sustainable in the long term, so 0 and 1 scores are usually an artefact of the finite duration of the tracking period. But more broadly, this limitation reminds us that many of the ways we might characterize individual news repertoires, such as through diversity, do not always make sense for the substantial minority of infrequent news users.

We use five independent variables. They are the number of news visits each panellist made using different modes of access: direct, Facebook, Twitter, search engine and Google News. We determine how a panellist arrived at a news article by looking at the URL of the previously visited page (for a similar approach, see Cardenal et al., 2019; Flaxman et al., 2016). We assume that in almost all cases, people arrive at news articles by clicking on hyperlinks rather than typing long, unwieldy URLs into their browser. Direct access to a news article occurs when the previous URL has the same subdomain but is not a page refresh ($M=56.13$, $SD=110.26$). Direct access therefore includes clicks from a news website's homepage, or from another article on the same site. Search access occurs when the previous URL is associated with an AOL, Bing, Google or Yahoo! search page (e.g. search.aol.co.uk, bing.com/search, google.co.uk/search, uk.search.yahoo.com; $M=3.75$, $SD=6.47$). A check was also applied to see whether the search terms contained in the URL matched words in the article's headline. Facebook, Twitter and Google News access occurs when the previous URL was associated with either Facebook (e.g. facebook.com, l.facebook.com; $M=3.98$, $SD=9.94$), Twitter (e.g. twitter.com, t.co; $M=0.97$, $SD=5.07$) or Google News (e.g. news.google.co.uk, news.google.com; $M=0.36$, $SD=4.57$). Shortened URLs were dealt with by looking at the second URL back (or further) to see where the link originated.

We also employ a number of demographic control variables. They are age group ($1=18$–24, $2=25$–34, $3=35$–44, $4=45$–54, $5=55+$; $M=3.41$, $SD=1.50$), gender (male: 51.5%, female: 48.2%), social grade ($1=E$, $2=D$, $3=C2$, $4=C1$, $5=B$, $6=A$; $M=4.04$, $SD=1.48$), education (no qualifications: 3.3%, GCSE or equivalent: 11.1%, A-level or equivalent: 18.9%, Degree or above: 51.4%, Other: 13.4%). Age group and social grade were treated as continuous, and education and gender were treated as categorical. Some of these variables are self-explanatory, but those unfamiliar with social science research in the United Kingdom may be unaware that National Readership Survey (NRS) social grade is a widely used categorization system for social class. It was developed as part of the NRS over 50 years ago. The categories are entirely based on occupation type, with 'higher managerial, administrative and professional' workers in grade A, and 'semi-skilled and unskilled manual workers' in grade D (grade E comprised 'state pensioners, casual and lowest grade workers, [and the] unemployed with state benefits'). As such, social grade is strongly correlated with income (National Readership Survey, n.d.). As for education, General Certificate of Secondary Educations (GCSEs) are the qualifications earned by most people during compulsory secondary education up to the age of 16. A-levels are studied for by those aged 16–18 and serve as entry qualifications for university. The panellists on average have higher levels of formal education and are from a higher social grade than the national population.

Results

We address our hypotheses using regression analysis. We use beta regression because all dependent variables range from 0 to 1 inclusive. However, because dependent variables with zeros and ones cannot be used in beta regression, in line with the methodological

Table 5. Beta regression models where the dependent variable is diversity.

	Diversity				Political diversity			
	Simpson's *D*		Shannon's *H*		Simpson's *D*		Shannon's *H*	
	exp(*B*)	SE	exp(*B*)	SE	exp(*B*)	SE	exp(*B*)	SE
Age	0.92***	0.02	0.93***	0.02	0.93***	0.02	0.94***	0.02
Gender	0.81***	0.05	0.84***	0.04	0.86**	0.05	0.88**	0.04
Social grade	0.98	0.02	0.99	0.02	0.99	0.02	0.99	0.01
Education								
GCSE	1.24	0.15	1.16	0.13	1.12	0.14	1.07	0.13
A-level	1.17	0.15	1.16	0.13	1.13	0.14	1.10	0.12
Degree	1.14	0.14	1.11	0.12	1.07	0.13	1.04	0.12
Other	1.02	0.15	1.00	0.13	0.98	0.14	0.96	0.12
Direct	0.99***	0.00	0.99***	0.00	0.99***	0.00	0.99***	0.00
Facebook	1.02***	0.00	1.03***	0.00	1.02***	0.00	1.02***	0.00
Twitter	1.02***	0.01	1.02***	0.01	1.02**	0.01	1.01**	0.00
Search engine	1.07***	0.00	1.09***	0.00	1.08***	0.00	1.06***	0.00
Google News	1.02***	0.01	1.03***	0.01	1.02***	0.01	1.02***	0.00
Other news	1.01*	0.00	1.01***	0.00	1.00	0.00	1.00*	0.00
N	2232		2232		2232		2232	
Pseudo *R*^{2}	.12		.15		.12		.12	

GCSE: General Certificate of Secondary Education; *SE*: standard error.
Note.*$p < .05$; **$p < .01$; ***$p < .001$.
Reference category for Gender is Female and No qualifications for Education.

literature, we applied a simple transformation that raises or lowers these values very slightly (Smithson and Verkuilen, 2006).

Let us first consider H1a and Hb. The results of the regression models where the dependent variable is news repertoire diversity are displayed in Table 5 and visualized in Figure 1. The results show that the more people access news directly, the less diverse their news repertoires, exp(*B*) = .99, $p < .001$. This is the case for both measures of diversity (Simpson's *D* and Shannon's *H*). We report exponentiated coefficients here, which means that a value lower than 1 indicates a negative association. The exponentiated coefficient for direct access is close to 1, as they are for other forms of news access, but it is important to keep in mind that the unit for these variables is one news article access. This means that if someone has a diversity score of 0.500, all else being equal a direct news article access would lower the diversity of their repertoire to 0.497 – but, as is clear from Figure 1, this compounds as people access more and more news directly over the course of a month. Conversely, the more people rely on all different forms of distributed discovery, the more diverse their news repertoires become. There are, however, differences in the strength of the effect on diversity. The effect is strongest for search engine use, meaning that if someone started with a diversity score of 0.5, one search engine access would raise it to 0.52. However, it is important to keep in mind that many panellists did not use search to access news at all during the 1-month tracking period, and the median number
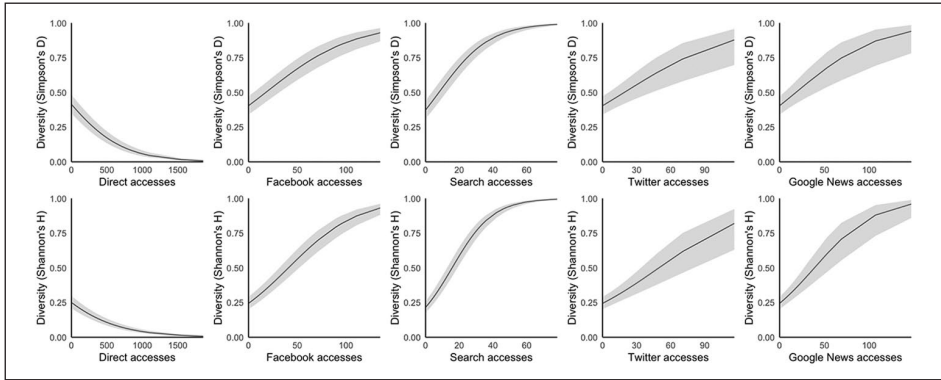
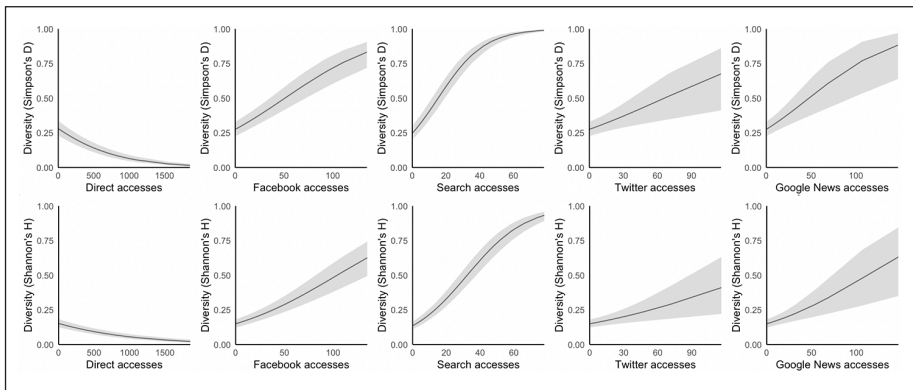**Figure 1.** Diversity plotted against number of each access type.

Figure 2. Political diversity plotted against number of each access type.

of times people used search for news was 1. The exponentiated coefficients for Facebook, Twitter and Google News access are smaller than for search, but significant and positive. We thus find support for H1a and H1b. The more people use direct access, the less diverse their news repertoires become, but the more they use distributed access the more diverse their news repertoires become. The results also show that people in the older age groups have less diverse news repertoires than younger people, and men have less diverse repertoires than women.

Turning to H2a and H2b, the results where the dependent variable is political diversity are displayed in Table 5 and Figure 2. As hypothesized, they show that people who more often access news directly have less politically diverse news repertoires ($\exp(B) = .99$, $p < .001$), whereas people who rely more on all forms of distributed news use have more politically diverse repertoires. In other words, direct access is associated with news repertoires that are more skewed towards either left-leaning outlets, right-leaning outlets, or outlets with no slant, whereas people who use distributed access are more likely to have

Table 6. Beta regression models where the dependent variable is partisan outlet prominence.

	exp(B)	SE
Age	0.95 ***	0.01
Gender	0.71 ***	0.03
Social grade	0.98	0.01
Education		
GCSE	0.93	0.10
A-level	0.95	0.10
Degree	0.91	0.10
Other	0.89	0.10
Direct	0.99***	0.00
Facebook	1.01***	0.00
Twitter	1.01***	0.00
Search engine	1.02***	0.00
Google News	1.01**	0.00
Other news	1.01***	0.00
N	2232	
Pseudo R^2	.15	

GCSE: General Certificate of Secondary Education; *SE*: standard error.
**$p < .01$; ***$p < .001$.
Reference category for Gender is Female and No qualifications for Education.
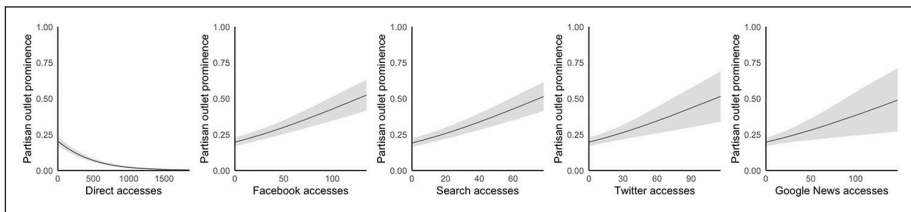

Figure 3. Partisan outlet prominence plotted against number of each access type.

news use balanced across the three categories, and thus more politically diverse repertoires. Hypotheses H2a and H2b are therefore supported.

Finally, for H3a and H3b, we change the dependent variable to our measure of how prominent more partisan outlets are in people's news repertoires. The results are displayed in Table 6 and visualized in Figure 3. They show that the more people access news directly, the lower the prominence of more partisan outlets. Conversely, the more they use different forms of distributed access, the more prominent partisan outlets are in people's repertoires. In contrast to diversity, the strength of the effect is similar for all types of distributed access. If someone has a partisan prominence score of 0.250, a news article accessed through Facebook, Twitter, search or Google News is associated with an increase to 0.252. The results, then, support H3a and H3b. In addition, age and gender are again significant variables in the model. Those in the older age groups also have

repertoires featuring more partisan outlets, as do men. However, it is interesting to note that education levels – important for many aspects of news use – do not have a significant association with any of the dependent variables.

As is clear from Table 5, both measures of diversity (Simpson's *D* and Shannon's *H*) produce very similar results – despite differences in how sensitive they are to infrequent news use. However, it is also worth highlighting that the pseudo R^2 for the models ($\leqslant.15$) are quite low, suggesting that other factors aside from access method and basic demographics are important for understanding people's online news repertoires.

As stated earlier, an alternative way of analysing the data would be to examine news sessions, whereby consecutive articles read on the same outlet are grouped together, with the access mode defined by how the panellist arrived at the first article. A summary of this analysis is provided in the Supplemental material. Although the number of direct accesses in the data is substantially reduced, the findings are qualitatively the same as those described in this section.

Discussion

In this study, we analysed tracking data from the United Kingdom and found that the more people use distributed news access through search engines, social media and aggregators, the more diverse their news repertoires. Furthermore, they are also more politically diverse, with news use spread across left- and right-leaning outlets. However, at the same time, as people's news repertoires become more diverse, they also become more likely to also contain partisan outlets.

If we combine the results with what we already know about how attention is distributed across news outlets in the United Kingdom, we can start to paint a more detailed picture. When people access online news directly, half of all visits are to the BBC. However, when we look at the distribution for search, social and aggregator news access, the BBC's share is much smaller and news visits are split more evenly across different outlets (see Table 3). This could be because people who use distributed news access are different from those who prefer direct access in ways that shape their news habits. But it could also be because news recommenders are surfacing links to a more diverse array of news outlets than people would otherwise use. If recommenders are consistently showing people news from a range of outlets, this is likely to result in more diverse repertoires – particularly in a context where direct access is otherwise so oriented towards a single outlet.

Once this happens, the inclusion in people's news repertoires of more partisan outlets becomes somewhat inevitable in the UK context. Because it attracts a large audience made up of equal numbers of centrists, left-leaning and right-leaning people, the BBC has a slant score of 0.00. Although the BBC has arguably been subject to increased criticism in recent years from those on the left and on the right (Cushion, 2018), this score is consistent with the formal requirement of duly impartial news coverage. Given that all other outlets have slant scores further from 0, news repertoires that are more diverse in that they contain more non-BBC visits will almost inevitably feature more partisan outlets.

Our analysis here has treated all news accesses as equal. But it is important to keep in mind that news visits can vary a lot in terms of how much attention and engagement

people have with the content. However, this is not something we observe here, as the average amount of time spent with news does not vary greatly by access mode. The average (median) time spent with a news article accessed directly was 52 seconds, compared to 54 seconds if accessed through social media and 49 seconds if arrived at through search. However, this may be because time spent with news can be challenging to measure from a technical point of view, and may not accurately reflect engagement due to individual-level differences in cognition and digital literacy (Groot Kormelink and Costera Meijer, 2020).

This study focused on the United Kingdom, but it is far from certain that the rise of distributed access has had the same consequences in other media systems. Although studies may consistently fail to find evidence for people's worst fears about echo chambers and filter bubbles, the specific dynamics are likely to vary. In a country like the United States, where partisan self-selection guides the media choices of those with strong political views, an increase in diversity from distributed news access might result in comparatively less partisan news repertoires if recommenders are surfacing news from more centrist sources. In countries that have comparably strong public service media – such as those in Northern Europe (e.g. Denmark, Finland, Sweden and Norway; Brüggemann et al., 2014) – an increase in diversity may have only a small effect on the prominence of more partisan outlets because mainstream outlets deviate a little from the centre (Newman et al., 2017). That said, there may be other consequences of an increased reliance on distributed access, such as a greater likelihood of exposure to alternative media, or even misinformation and disinformation. Either way, researchers should be wary of extrapolating out from single-country studies to the rest of the globe.

Nonetheless, the findings – particularly those on political diversity – still add to the growing body of evidence challenging the existence of echo chambers and filter bubbles (at least when it comes to news), given they suggest any imbalance in people's news repertoires is primarily a result of self-selection rather than algorithmic selection. But because this diversity seems to be accompanied by the use of more partisan outlets, we should be wary of the simple view that diversity is good and homogeneity is bad. We should remain open to the possibility that both cross-cutting exposure and diversity can have negative consequences. Indeed, it may be that exposure to conflicting partisan views, rather than over-exposure to like-minded views, will offer a better explanation for the negative outcomes – like polarization – that are sometimes associated with distributed media use (Bail et al., 2018). Similarly, although consuming news from a variety of outlets may offer some benefits, some may simply be more comfortable with a world where most people only access news from impartial sources like the BBC – where differing views are often recognized, but presented in a certain way.

How we evaluate distributed news use will, in the end, depend on normative decisions about what democratic role – if any – societies want platforms to play. The findings we present here describe recommender systems that are imperfectly aligned with any one of the aforementioned ideal types identified by Helberger (2019). Diversity is clearly a core component of a 'deliberative' recommender, and we present evidence here that – at least if compared to what people choose to do if left to their own devices – distributed news access can increase diversity. But at the same time, we do not find evidence for other 'deliberative' requirements, such as the 'prominence of public service media content' or

a 'preference for rational tone' and 'consensus seeking' (Helberger, 2019: 1008). In fact, we sometimes find the opposite.

On the basis of the evidence presented here, and evidence from other studies, it seems that many news recommender systems embody at least some of the elements from all four of Helberger's types. For example, we see opportunities for active user curation and the surfacing of important topics characteristic of 'liberal' recommenders, the opportunities for participation and active citizenship associated with 'participatory' recommenders, and – as already discussed – the diversity and cross-cutting exposure that embody 'deliberative' and 'critical' systems. Indeed, news recommenders that are a 'hybrid' of ideal types may offer a more accurate description of systems we encounter in the real world. If so, future research might ask to what extent we can measure and evaluate individual dimensions, like diversity and cross-cutting exposure, to ultimately understand whether hybrid recommenders are able to play any kind of meaningful democratic role.

Funding

ORCID iD

Richard Fletcher 🆔 https://orcid.org/0000-0002-5496-2112

Supplemental material

Supplemental material for this article is available online.

Notes

1. We define outlets using the subdomains and directories listed in Table 1. For broadcasters (e.g. BBC, ITV and Sky), we use the '/news/' directory to avoid counting non-news content. BBC also includes the '/sport/' directory so that sports news is included (sport is already included in the '/news/' directory for other outlets). Similarly, for Yahoo! we count visits to the news subdomain.
2. Prior to the analysis, the dataset was cleaned to remove many of the non-article visits. For example, visits were removed if the title or headline contained fewer than five words (to remove generic section pages, crosswords, banner clicks, etc.), or if they were visits to miscellaneous pages (e.g. 403 Forbidden, 404 Not Found, etc.). We also removed visits of shorter than 5 seconds as this is unlikely to signify meaningful engagement. However, it is important to note that cleaning large datasets using rules like these is an imperfect process, and that a small number of non-article visits will erroneously remain in the data.

References

Bail CA, Argyle LP, Brown TW, et al. (2018) Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences in the United States of America* 115(37): 9216–9221.

Bakshy E, Messing S and Adamic LA (2015) Exposure to ideologically diverse news and opinion on Facebook. *Science* 348: 1130–1132.

Barberá P (2020) Social media echo chambers, and political polarization. In: Persily N and Tucker JA (eds) *Social Media and Democracy: The State of the Field and Prospects for Reform*. Cambridge: Cambridge University Press, pp. 34–55.

Barberá P, Jost JT, Nagler J, et al. (2015) Tweeting from left to right: is online political communication more than an echo chamber. *Psychological Science* 26(10): 1531–1542.

Boczkowski PJ, Mitchelstein E and Matassi M (2018) 'News comes across when i'm in a moment of leisure': understanding the practices of incidental news consumption on social media. *new media & society* 20(10): 3523–3539.

Brüggemann M, Engesser S, Büchel F, et al. (2014) Hallin and Mancini revisited: four empirical types of Western media systems. *Journal of Communication* 64(6): 1037–1065.

Bruns A (2019) *Are Filter Bubbles Real?* Cambridge: Polity Press.

Cardenal AS, Aguilar-Paredes C, Galais C, et al. (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.

Carpenter S (2010) A study of content diversity in online citizen journalism and online newspaper articles. *new media & society* 12(7): 1064–1084.

Cushion S (2018) Journalism under (Ideological) threat: safeguarding and enhancing public service media into the 21st century. *Journalism* 20(1): 69–71.

Dunaway J, Searles K, Sui M, et al. (2018) News attention in a mobile era. *Journal of Computer-Mediated Communication* 23(2): 107–124.

Flaxman S, Goel S and Rao JM (2016) Filter bubbles echo chambers, and online news consumption. *Public Opinion Quarterly* 80: 298–320.

Fletcher R and Nielsen RK (2017) Are news audiences increasingly fragmented? A cross-national comparative analysis of cross-platform news audience fragmentation and duplication. *Journal of Communication* 67(4): 476–498.

Fletcher R and Nielsen RK (2018a) Are people incidentally exposed to news on social media? A comparative analysis. *new media & society* 20(7): 2450–2468.

Fletcher R and Nielsen RK (2018b) Automated serendipity: the effect of using search engines on the diversity and balance of news repertoires. *Digital Journalism* 8(6): 976–989.

Fletcher R, Cornia A and Nielsen RK (2020) How polarized are online and offline news audiences? A comparative analysis of twelve countries. *The International Journal of Press/politics* 25(2): 169–165.

Gentzkow M and Shapiro JM (2011) Ideological segregation online and offline. *The Quarterly Journal of Economics* 126: 1799–1839.

Groot Kormelink T and Costera Meijer I (2020) A user perspective on time spent: temporal experiences of everyday news use. *Journalism Studies* 21(2): 271–286.

Guess AM (2015) Measure for measure: an experimental test of online political media exposure. *Political Analysis* 23(1): 59–75.

Guess AM (2021) (Almost) Everything in moderation: new evidence on Americans' online media diets. *American Journal of Political Science* 76: 1–16.

Hallin DC and Mancini P (2004) *Comparing Media Systems: Three Models of Media and Politics*. Cambridge: Cambridge University Press.

Hasell A (2020) Shared emotion: the social amplification of Partisan news on Twitter. *Digital Journalism*. Epub ahead of print 2 November 2020. DOI: 10.1080/21670811.2020.1831937.

Helberger N (2019) On the democratic role of news recommenders. *Digital Journalism* 7(8): 903–1012.

Kalogeropoulos A, Fletcher R and Nielsen RK (2019) News brand attribution in distributed environments: do people know where they get their news? *new media & society* 21(3): 583–601.

Kotkov D, Wang S and Veijalainen J (2016) A survey of serendipity in recommender systems. *Knowledge-Based Systems* 111: 180–192.

Loecherbach F, Moeller J, Trilling D, et al. (2020) The unified framework of media diversity: a systematic literature review. *Digital Journalism* 8(5): 605–642.

McDonald DG and Dimmick J (2003) The conceptualization and measurement of diversity. *Communication Research* 30(1): 60–79.

Moore M and Tambini D (Eds) (2018) *Digital Dominance: The Power of Google Amazon Facebook, and Apple*. Oxford: Oxford University Press.

Müller P and Schulz A (2019) Alternative media for a populist audience? Exploring political and media use predictors of exposure to Breitbart, Sputnik, and Co. *Information Communication & Society* 24: 277–293.

Napoli PM (1999) Deconstructing the diversity principle. *Journal of Communication* 49(4): 7–34.

National Readership Survey (n.d.) Social Grade. Available at: http://www.nrs.co.uk/nrs-print/lifestyle-and-classification-data/social-grade/ (accessed 21 August 2019).

Newman N and Kalogeropoulos A (2017) *Mapping Online News Discovery for Computer Users in the UK*. Oxford: Reuters Institute for the Study of Journalism.

Newman N, Fletcher R, Kalogeropoulos A, et al. (2017) *Reuters Institute Digital News Report 2017*. Oxford: Reuters Institute for the Study of Journalism.

Newman N, Fletcher R, Kalogeropoulos A, et al. (2018) *Reuters Institute Digital News Report 2018*. Oxford: Reuters Institute for the Study of Journalism.

Newman N, Fletcher R, Kalogeropoulos A, et al. (2019) *Reuters Institute Digital News Report 2019*. Oxford: Reuters Institute for the Study of Journalism.

Pariser E (2011) *Filter Bubbles: What the Internet Is Hiding from You*. London: Penguin Books.

Pasquale F (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

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

Scharkow M, Mangold F, Stier S, et al. (2020) How social network sites and other online intermediaries increase exposure to news. *Proceedings of the National Academy of Sciences in the United States of America* 117(6): 2761–2763.

Shannon CE and Weaver W (1949) *The Mathematical Theory of Communication*. Champaign, IL: The University of Illinois Press.

Simpson EH (1949) Measurement of diversity. *Nature* 163: 688.

Smithson M and Verkuilen J (2006) A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* 11(1): 54–71.

Sunstein CR (2017) *#Republic: Divided Democracy in the Age of Social Media*. Princeton, NJ: Princeton University Press.

Tan Y and Weaver DH (2013) Agenda diversity and agenda setting from 1956 to 2004: what are the trends over time? *Journalism Studies* 14(6): 773–789.

Taneja H (2016) Using commercial audience measurement data in academic research. *Communication Methods and Measures* 10(2–3): 176–178.

Tewksbury D, Weaver AJ and Maddex BD (2001) Accidentally informed: incidental news exposure on the world wide web. *Journalism & Mass Communication Quarterly* 78: 533–554.

Valeriani A and Vaccari C (2016) Accidental exposure to politics on social media as online participation equalizer in Germany Italy, and the United Kingdom. *new media & society* 18: 1857–1874.

Webster JG and Ksiazek TB (2012) The dynamics of audience fragmentation: public attention in an age of digital media. *Journal of Communication* 62(1): 39–56.

Weeks BE, Lane DS, Kim DH, et al. (2017) Incidental exposure selective exposure, and political information sharing: integrating online exposure patterns and expression on social media. *Journal of Computer-Mediated Communication* 22(6): 363–379.

Zuiderveen Borgesius FJ, Trilling D, Möller J, et al. (2016) Should we worry about filter bubbles? *Internet Policy Review* 5(1): 1–16.

Author biographies

Richard Fletcher is a senior research fellow at the Reuters Institute for the Study of Journalism and leads the Institute's research team.

Antonis Kalogeropoulos is a lecturer in Communication and Media at the University of Liverpool and a research associate of the Reuters Institute for the Study of Journalism.

Rasmus Kleis Nielsen is the director of the Reuters Institute for the Study of Journalism and professor of Political Communication at the University of Oxford.