The Invisible Hand of the Unaccountable Algorithm: How Google, Facebook and Other Tech Companies Are Changing Journalism

David R. Brake

INTRODUCTION

That the broad public adoption of the Internet is transforming journalism is well-understood. The way news is gathered, the way it is distributed, sold, and paid for, and the ways it is consumed and redistributed have all changed. Broadly speaking, among those concerned with political economy in journalism studies there have been two main areas of interest about the effects of digitisation on the field. Optimistically, there was the possibility that 'ordinary people' could be better heard because of online commenting features and the use of user-generated content (Vujnovic et al. 2010; Beckett 2008), though this has met some scepticism given the continuing power of large news organizations (McChesney 2014). More recently, it has become apparent that because the Internet has allowed potential advertisers to reach consumers

D.R. Brake (✉)
LSE, London, UK
e-mail: davidbrake@gmail.com

directly without being bundled with editorial products, the business models of conventional journalism across much of the world are being undermined. Internet advertising is going largely to a few tech giants—in 2015, Google received half of all digital advertising in the US and Facebook another 13% (PwC 2016)—and this has clear implications for the ability of media organizations to afford to adequately cover news of public interest.

But not all aspects of the recent changes in the field of journalism that digitisation has helped to foster are equally visible. To the extent that journalism has become digital its processes have also become increasingly mediated by algorithms. While individual implications of the growing importance of algorithms for journalistic activities have already been highlighted (Hermida et al. 2012; Anderson 2013; Poell and Van Dijck 2014; Bakker 2012) I argue it is important to consider the full scope of potential and actual impacts of algorithms to understand how these may be helping to shape the evolution of journalism.

Algorithms in their broadest sense are ubiquitous—as Gillespie puts it, they are "encoded procedures for transforming input data into a desired output, based on specified calculations" (2014, p. 1), and as such do not even necessarily involve computers—a cooking recipe is also an algorithm. For the purposes of this chapter, however, I will focus largely on the "public relevance algorithms" that search engines such as Google and social media companies such as Facebook use "to select what is most relevant from a corpus of data composed of traces of our activities, preferences, and expressions" (2014, p. 2).[1] Sometimes, as with a Google search, this might be the algorithm that determines which sites turn up when a journalist searches for, say "earthquake location". Sometimes an explicit search is not required—Facebook's public relevance algorithm governing users' newsfeeds analyses the 2000 postings that could be visible to the average user every day to determine which 500 will actually be presented on their newsfeed (Backstrom 2013).

As will be outlined in detail in the remainder of this chapter, public relevance algorithms used by search engines and social media platforms influence what stories journalists may find, what they choose to write about and how they do it, and how those stories are themselves found and recirculated.

The critical work of Feenberg on technology and society (Feenberg 1999) provides a useful lens to analyse how these algorithms affect journalism directly and, as importantly, how the way journalists change their

behaviour because of their awareness of the algorithms. In his framework, technologies are initially created with technical features designed to achieve particular goals (e.g. measuring visits to individual stories to establish how popular topics are on a website). Yet through a process he calls "secondary instrumentalisation" (Feenberg 2008), they are adopted by users who use them in their own ways and find their own purposes for them (e.g., rewarding or penalising individual journalists for their stories' popularity might be an unexpected secondary instrumentalisation). Over time, "standard ways of understanding individual devices and classes of devices emerge" (Feenberg 2008, p. 23), creating what he calls technical codes that are both durable and often invisible—in the sense that they are taken for granted. And those codes, in turn, may introduce biases into society—either directly (when multimedia news websites cannot be read by blind users) or indirectly (when journalists change the stories they choose to cover because their stories are being ranked by the number of views they receive). The former he calls "formal bias" and the latter "implementation bias". In describing the role of algorithms in journalism in this chapter through this lens, I hope to illuminate these sometimes-hidden biases to change how these tools are used and, if necessary, encourage the toolmakers to change them.

The Problem of Invisible Algorithmic Gatekeepers

For decades, scholars have recognised the importance of journalistic gatekeepers—the organizations and individuals who decide what will be selected as news. Walter Lippmann noted 90 years ago, "all the reporters in the world working all the hours of the day could not witness all the happenings of the world … the range of subjects these comparatively few men manage to cover would be a miracle indeed, if it were not a standardized routine" (Lippmann 1922, p. 214). Now there is more information available than ever before on the happenings of the world, but when it's online it's usually presented and prioritized through 'standardized routines' (algorithms) before it ever reaches the eyes of journalists—on the one hand—and the public on the other.

The public relevance algorithms Google, Facebook, and others employ are gatekeepers in two directions—they govern much of the information journalists use to research stories. Then, in turn, they govern the size and composition of the audiences their stories receive. They may be more powerful than traditional news editors, but tend to be invisible

and often unaccountable. The companies behind these algorithms tend to claim not to be in the news business at all. Mark Zuckerberg, Facebook's founder, continues to say the site does not act as an editor (D'Onfro 2016)[2] and officials at Google, Apple and Twitter take a similar line (Herbst 2016). They also assert that their public relevance algorithms are neutral, a position which has gone largely unchallenged by the public. As Morozov notes, "Google likes to claim that it is simply an algorithms-powered neutral intermediary that stands between a given user and the collective mind of the Internet. On its corporate website, Google compares the presentation of its search results to democratic elections, with the most-linked sites emerging on top. If the top results lead to sites that are politically incorrect or racist or homophobic, the fault is not Google's but the Internet's." (Morozov 2011).

Is what they claim true? It is hard to prove this as the study of these new gatekeepers is difficult. Unlike traditional gatekeepers—newspaper editors, for example—those within search engine companies and social media organizations who code the algorithms that control the flow of news and information online are seldom public figures. The precise way that their algorithms work is shrouded in secrecy. This is because such firms believe these algorithms give them commercial advantages and because companies fear that if the way those algorithms worked was public, various actors, including news organizations, would seek to "game the system" to make their information most prominently displayed to users (Olsen 2003). Tech companies tend to bind their employees with nondisclosure agreements and, as a result, there is a dearth of ethnographic studies that could help researchers or policymakers to understand the values held by the people who program such software, though journalists and writers have occasionally had some access to them (Kirkpatrick 2010; Nunez 2016a).

The designing of algorithms is far from being neutral but influenced by many factors. As Feenberg remarks, "a wide variety of social groups count as actors in technical development. Businessmen, technicians, customers, politicians, bureaucrats are all involved to one degree or another. They meet in the design process where they reveal their influence by proffering or withholding resources, assigning purposes to new devices, fitting them into prevailing technical arrangements to their own benefit, imposing new uses on existing technical means, and so on. The interests and worldview of the actors are expressed in the technologies they participate in designing" (Feenberg 1999, p. 11). In the case of search engines

specifically, Van Couvering found that producers framed quality of search results with reference to two primary schemas—market and "science-technology", and the latter tended to define good search results as those that "satisfy" users, bringing them documents that "answered the user's question or was what he or she wanted" (Van Couvering 2007, p. 876), which, as she points out, can have problematic implications.

> The focus on relevance constrains the articulation of other quality goals. For example, in journalism, objectivity, fairness, diversity, and representation are typical examples of quality goals. In the course of this research, interviewees mentioned many everyday practices in search engine programming that could be considered censorship of search results and have the potential to lead to biases in search. These included blacklisting, or the exclusion of certain sites or site owners; whitelisting, or the automatic inclusion of certain sites or site owners; weighting content according to whether sources were considered to be authoritative or not; and adjusting results based on pressure from executives to respond, for example, to current news events. None of these practices were considered problematic, because all were linked to obtaining greater relevance in search engine results. (Van Couvering 2007, p. 882)

Google tells publishers, for example, "if we find non-news content mixed with news content, we may exclude your entire publication from Google News," but provides no explanation of how its algorithm (or perhaps human intervention) defines or finds non-news content (Google n.d.-b). Clearly, in deciding what weight to put on each of those factors, human beings within Google are making decisions that would hitherto be the function of human editors, and there is no clearly correct unambiguous way to prioritise and operationalize these values.

Complicating matters further, these algorithms by their nature also tend to have different results for different users. In order to try to present the most "relevant" material to individuals, they prioritise the data they provide in part by using information they have about searchers, whether this is given by users themselves (when they are using a social network) or inferred through technical means (locating users via their Internet addresses or guessing at their socio-demographics or interests based on sites they have previously visited). As a result, it is hard for researchers from the outside to guess at the different effects the same algorithm might have for different people. Indeed, because of the complexity of such algorithms and the number of variables they take

into consideration, it may not even be possible for the algorithm's own designers to fully understand how they work in practice—the extreme case of this being algorithms that use "machine learning", where computers program themselves based on patterns in the data they encounter. A recent paper suggests a machine learning algorithms trained using everyday language will 'learn' human biases automatically. For example, the research suggests there's a danger when using a search algorithm of this kind that 'European-American names are associated with pleasantness and African-American names with unpleasantness' (Narayanan 2016). To make matters more complex, tech companies' algorithms are continually tweaked; one industry source says Google changes its algorithms 5 to 600 times a year (Moz.com n.d.).

The effects of such biases introduced by algorithms can be subtle but nonetheless powerful and pervasive. One experiment conducted by Facebook's own researchers concluded that boosting the amount of hard news selected US citizens received raised their voter participation in that year's congressional election from 64 to 67% (Bakshy et al. 2015; O'Neil 2016). Another study suggested that a biased search engine algorithm could "shift the voting preferences of undecided voters by 20% or more" without their knowledge (Epstein and Robertson 2015, p. E4512). If this kind of intervention seems farfetched, Facebook has already been accused of intervening directly to influence a political debate about its free Internet product in India using messages aimed at its users through its service (Bhatia 2016), and at least 61 of its employees sought to ask Mark Zuckerberg through an internal company poll "What responsibility does Facebook have to help prevent President Trump in 2017?" (Nunez 2016a). Google blacked out its home page "doodle" in 2012 to encourage its users to petition the US congress to stop the Stop Online Piracy Act (Zittrain 2014)—this at least was visible, but it would be easy for them to intervene in search results to accomplish other goals.

If algorithms that are biased—inadvertently or deliberately—resulted in biased search results for journalists researching a story, and this in turn caused the story they wrote to be biased, it would be hard for anyone to identify the role of the algorithm and harder still to get the organizations responsible for those algorithms to accept their responsibility. Given the central place that algorithms are occupying in journalistic practice, however, it is crucial to make this problem visible and to start to seek solutions.

In the remainder of this chapter I analyse the three key ways that journalism practices may be affected by algorithms: news values, news sources and news distribution.

News Values

The first area in which algorithms meet and influence the news is in the selection by journalists of potential stories. While there are many factors influencing this, including direct control (e.g. in authoritarian states), ideology (Shoemaker 2006) and the perceived priorities of journalists' peers (Donsbach 2004), most journalists have to justify their choices at least in part by claiming their work will be of interest to their audiences. Historically, however, while journalists have paid lip service to the idea that their news choices are largely driven by what their audiences want to hear about, studies find they have often written for their peers and editors, not necessarily for the public. As Gans found in his study of US newsrooms, the journalists there "had little knowledge about the actual audience and rejected feedback from it. Although they had a vague image of the audience they paid little attention to it; instead they filmed and wrote for their superiors and for themselves, assuming … what interested them would interest the audience" (1979, p. 229). These are findings echoed by studies across the ocean inside the BBC (Burns 1977). News organizations—particularly large ones—have always had ways to hear from their public, whether through phone calls, letters to the editor or surveys and focus groups, but this information has tended to be used primarily commercially rather than as a tool to inform editorial judgment.

The widespread availability of journalists' email addresses and the common provision of space for reader comment on stories on news websites—at least until recently (Santana 2016)—have undoubtedly increased the amount of exposure journalists have to their audiences, but algorithms also have an important role in the evolution of the journalist-audience relationship. In particular, the reports that algorithms provide counting visitors to stories or measuring conversation around topics may be perceived by journalists and their managers as giving them a quick, easy and unambiguous indication of topics in which their audiences are interested.

A new set of tools—"trending topics" has been provided by search engine and social media companies—most notably Google, Facebook

and Twitter—to help news organizations (and the public) assess what the public is interested in, day by day or even minute by minute. This could be helpful in making journalism more relevant to its audiences, but if audience appeal is given too much sway because it appears easier to define this may further endanger important but less "sexy" investigations that are in the public interest (Nguyen 2013). The outputs of these algorithms might also reinforce the herd mentality that can already exist among journalists who thereby cover topics they might not otherwise choose for fear of missing an angle the competition is exploiting.

But taking Twitter as an example, it is also possible to question whether what its users tweet is in fact representative of what the wider public is interested in—in many countries Twitter users (active ones in particular) are more likely to be younger, urban, highly educated elites who cannot be seen as representing the whole population (Brake 2013; Perrin 2015; Blank 2016). Therefore, it would be bad if journalists see trending topics as a reflection of what the public is interested in and follow too slavishly. It would be even worse if journalists assume what discussed on Twitter actually reflects what interests the public as a whole. And, of course, because social media companies are not transparent about how trending topics are generated, there may be many tacit biases or inaccuracies in what gets highlighted. Indeed, it has emerged that in the case of Facebook, human gatekeepers have been used to "tidy up" the results of Facebook's algorithms and have been accused of introducing a bias against right-wing stories as a result (Nunez 2016a). The subsequent withdrawal of human gatekeeping in favor of a purely algorithmic approach presented problems of its own, with "false stories, misidentified keywords, and celebrity gossip in the place of more serious news." (Wells 2016). One source claims the company's failure to block or downgrade fake news in news feeds in advance of the election of Donald Trump as president of the United States may have been in part due to fear of a backlash from conservatives who benefited from the predominantly right-wing orientation of these sites (Nunez 2016b). In a seeming about-turn, within weeks of the election the company's founder, Mark Zuckerberg, was promising "much more" work would be done to combat fake news by relying on the reports of the public and third parties (all, in all likelihood, managed algorithmically) (Zuckerberg 2016).

Moreover, as Gillespie points out, while exactly what constitutes a "trending" topic is not spelled out for commercial reasons, *trending* by definition tends to focus on issues that have risen to prominence quickly

rather than on topics that may be of high but steady concern, which could tend, he argues, to foster "a public more attuned to the 'new' than to the discussion of persistent problems, to viral memes more than to slow-building political movements" (Gillespie 2012).

Lastly, to the extent that journalists focus on what is trending on social media because it is easy to measure (thanks to those algorithms) they may also end up devoting a disproportionate amount of attention to issues of interest to social media users (who may not be representative of the broader public) and cover issues from their perspective using quotes drawn from social media because they are easy to find.

The addition of trending topics is just the tip of the iceberg when it comes to public relevance algorithms helping to dictate the news agenda. A different set of relevant algorithms is those used to measure the popularity (and profitability) of stories on news websites once they are published. When news is distributed digitally, tracking of individual stories' popularity becomes much easier than it was in print or broadcast media. Using software like Google Analytics or paid-for products like Chartbeat, journalists (and their editors) can know with more precision and in real time how many people are reading each individual story and how they find their way to it (whether it was referred to them on a social network, from a web link or from a search for example). Indeed, using Chartbeat and other software it is possible to learn how much of each story visitors tend to read and which parts they linger on. These tools are not just to be found in Internet news startups but are being increasingly used by major traditional media operations, "Data-informed decision-making previously associated with sites like BuzzFeed, Gawker, and the Huffington Post is increasingly central to editorial processes at organisations like *The Guardian*, *The New York Times*, and *Die Welt* as well as leading public service media like the BBC and various start-ups like Quartz and Ze.tt" (Cherubini and Nielsen 2016, p. 41).

The algorithms in web log analysis software take raw traffic numbers and server data and transform them into information more useful to journalists and marketers by, for example, highlighting search keywords used to find stories on a site, or identifying which journalist's stories are most popular. This can be problematic if journalists then face pressure to produce more stories that are more popular or, in extreme cases, that are on topics that can plausibly contain keywords that attract the most profitable kind of readers.

Editorial focused on potentially profitable readers is not new, but thanks to algorithms that analyse traffic flows and reader demographics, success in courting those readers can appear to be measured more precisely, potentially increasing pressure on individual journalists to cover individual stories in a way that will reach and appeal to wealthier readers—by using certain keywords, for example (Poell and Van Dijck 2014). This phenomenon can be observed at its most extreme at the margins of the journalism industry with organizations like Demand Media, which commissions "how to" articles algorithmically:

> To find out what terms users are searching for, it parses bulk data purchased from search engines, ISPs, and Internet marketing firms (as well as Demand's own traffic logs). Then the algorithm crunches keyword rates to calculate how much advertisers will pay to appear on pages that include those terms… Third, the formula checks to see how many Web pages already include those terms … Armed with those key words, another algorithm, called the Knowledge Engine, dives back into the data to figure out exactly what people want to know about the term. (Roth 2009)

As Nicholas Carr memorably (and prophetically) put it, "If you could get some cheap freelancer to hack together a story on new developments in high-definition televisions, that could really be a bonanza. Manufacturers, retailers and programmers bid a lot for clickthroughs on HDTV-related ads. And the readers attracted to a story on developments in HDTV are likely to be considering some kind of purchase—and thus in the mood to click. Ka-ching, ka-ching" (Carr 2006). This may exacerbate the trend to shift resources away from reporting stories of public interest toward stories that are measurably more profitable.

News Sources

As resources available for reporting become increasingly stretched, and as the pressure to churn out stories ever faster rises, journalists rely more than ever on the web and on social media to help them find stories (or to help them find sources they can then interview) (Hermans et al. 2009). For example, in one study, 81% of the German journalists interviewed said that search engines are important or very important for extensively researching a topic (Machill and Beiler 2009, p. 197). But much of the time that searching process is mediated by public relevance algorithms

(Gillespie 2014). The most obvious of these is Google's pagerank algorithm, although other search engines have similar algorithms.

Google's stated aim is to "organize the world's information and make it universally accessible and useful" (Google n.d.-a), but any organization implicitly makes judgements about what to present to users and how prominently.

Google has historically claimed that the way it orders its results is "objective", though more recently it has acknowledged that objectivity is impossible to achieve and thus the workings of its algorithms amount to an "opinion" (Metz 2010). It is also established that Google sometimes manipulates the results of its algorithms for pro-social ends—to deliver suicide prevention information prominently when people search for the word *suicide*, for example (Cohen 2010). Many scholars have argued that there are important ways in which Google's algorithms may fall short of normative notions of neutrality (Rogers 2004), though of course it remains as difficult to identify neutrality in an algorithm as to identify objectivity in news coverage. Thelwall and Vaughan, for example, suggest that Google overrepresented American websites (Thelwall and Vaughan 2004). Others argue that sites owned by powerful and wealthy organizations can get higher positions in search engine ranking—Hindman, for example, studying the US politics sites that were most linked to online (and hence most likely to show up prominently in search) found "almost all prominent sites are run by long-established interest groups, by government entities, by corporations, or by traditional media outlets" (Hindman et al. 2003, p. 25). Scholars generally do not believe this is due to any deliberate bias on the part of search engines but rather that they are an unintended side effect of the way search engine algorithms are designed.[3] As Van Couvering found from her interviews with search engine producers and engineers, they are largely concerned with providing search results that customers believe are satisfactory rather than worrying about broader concerns of public welfare, fairness or bias (Van Couvering 2007). Thus, the values of the designers of these algorithms journalists use do not necessarily align with what journalists might require in order to best serve the public. For example journalists might wish to be assured that the results they receive when searching for information on the impact of a dam provided the views of as diverse as possible a range of those involved—economists, environmentalists, people living in the area and so on. Search engine

algorithm designers are more likely to want to ensure readers get all the links they expect.

Algorithms are not just used to decide how prominently a source is displayed. They can also be used to attempt to measure its validity. Reveal[4] is an EU project that provides automated tools to suggest whether a tweet or an image is real or fake—certainly of potential value, but one more way in which journalists may be ceding their decision-making to an algorithm—especially if future search engines were to use such tools to exclude material that was thought to be fake.

As information professionals are journalists aware of potential problems and using algorithms critically? Certainly there is strong evidence that the US public as a whole trusts search engines. A 2012 survey found that 91% of search engine users say they always or most of the time find the information they are seeking when they use search engines, 73% say that most or all the information they find as they use search engines is accurate and trustworthy and 66% say search engines are a fair and unbiased source of information (Purcell et al. 2012). By contrast, in a survey of German journalists, 72.9% either did not agree at all or tended not to agree that search engines provide neutral search results (Machill and Beiler 2009, p. 197). They might, therefore, compensate for this (e.g. by searching using several search engines or looking past the first page of results), but it is not clear that they are necessarily exercising the caution they claim to have in everyday practice. After testing journalists' search skills Machill found "in spite of their daily and very extensive research work, journalists do not automatically achieve greater search success [than the general public] with Google" (2009, p. 199). This suggests they may not be as 'digitally literate', or critical, as they claim to be.

Other algorithms used in relation to social media may also introduce biases in the way stories are researched. Twitter itself, for example, and a legion of third party companies like Klout or RightRelevance provide ways to not just search social media by keywords but to highlight postings by those who have the most followers or who are in some other way considered "influential". Influence on the Internet may be a convenient proxy for expertise, but not always an accurate one. To the extent these "influence" tools are or might be used in the future by journalists as a convenient way to seek "experts" instead of using more conventional means of establishing credentials, this may affect the quality and accuracy of stories that are written.

News Redistribution

As people spend ever longer on their digital devices and on social media, the importance of news as a "destination" for the public is diminishing, in favour of what Hermida terms "ambient journalism": "awareness systems that offer diverse means to collect, communicate, share and display news and information in the periphery of a user's awareness" (Hermida 2010). As attention to scheduled news, news "packages" like newspapers and visits to news organizations' home pages decline (Kirkland 2014), news organizations—particularly online ones—are increasingly reliant on search engines and social media to attract readers/viewers to their articles (and to sell those audiences to advertisers). In the US 18% of adults say they "often" get news on social media now—that's only 2% lower than the number who often get it from a newspaper (Pew Research Center's Project for Excellence in Journalism 2016), and a survey of large online publishers in the US and Europe found they were getting about 30% of their traffic from Facebook alone (Lichterman 2016).

But, as noted earlier, search engines and social media are not neutral news distributors, and news organizations are in constant competition to raise the profile of their stories both in search engine results and social media feeds (in much the same way that new sources may struggle to make themselves visible to journalists online as alluded to earlier). Publishers and individual journalists are increasingly focusing not just on ensuring that their articles are accurate and of interest to their intended audience but also that they are treated favourably by search and social media algorithms (Dick 2011).

One of the ways that such algorithms work "as intended" is when they highlight the freshest news, the news that has been linked to by the most respected sources, or (in the case of news distribution through social media) the news that is posted by the people the reader most trusts or wants to hear from. But there are other, less straightforward ways that, for example, Facebook's algorithms may operate—not in the interests of readers but in the interests of Facebook itself. The company has reportedly given priority in its newsfeed to stories that use its live video tool (Valinsky 2016). Not only might this pressure news organisations to provide live video even for stories where it makes little journalistic sense, but it also tends to lock news organisations further into Facebook's ecosystem (since such video is only visible via Facebook). Similarly, users may believe that when using Google News they are receiving news

sources algorithmically selected from all of the potential news sources online. In fact, all news sources Google offers through the "news" part of a search query have been vetted by the company for inclusion as a source (Google n.d.-b), though other alternative news sources may appear in the main Google web search.

Of course, as noted earlier, it is not clear exactly how the algorithms of search sites work nor is it possible to know precisely how a social media site like Facebook prioritises what it presents to each social media user in his or her newsfeed. Nonetheless, there is a large industry that provides advice on how best to promote journalistic content. The benefits of doing this well are seductive—Buzzfeed's ability to have its material shared on social media is one of the reasons it is thought to be a success (Rowan 2014).

The price of falling foul of these algorithms can be devastating. Search engines can punish any sites they suggest have been attempting to manipulate them by excluding them from searches. In 2006 for example, BMW.de (the car manufacturer's German site) was reportedly removed from Google's search results because Google disapproved of ways the company attempted to boost its search engine ranking (BBC News Online 2006). Similarly, Facebook, in response, it said, to user feedback, penalised "clickbait" news items in 2014 (El-Arini 2014). These are headlines that conceal what the story is about in order to force readers to click to find out. One report suggested that "the algorithm change led to a huge drop in traffic for Upworthy [a news organization that employed this tactic] and caused it to change its business model" (Pelegrin 2015).

This is a clear example of the way in which an unaccountable algorithm is effectively intervening in journalistic practices, endangering an attention-getting tool journalists writing headlines have been using for decades. It is not necessary that organisations like Facebook or Google actually use algorithms that distort news priorities to cause concern. Problems can result if enough news publishers believe the algorithms work a certain way and adjust their behaviour accordingly—what Feenberg called an "implementation bias" (2008, p. 11). If editors believe Google tends to push longer stories down its list of search results, they may keep stories short regardless of journalists' need to provide more context for what they write.

Conclusion

Historically, media scholars have tended to be concerned that news agendas were being warped by, for example, ideological bias (whether conscious or unconscious), or proprietorial interference. Google, Facebook and other algorithmic mediators are not traditional media companies. Their proprietors are not typically seen as potentially pursuing political ends using their media power, although researchers are already concerned that they could if they wished (Epstein and Robertson 2015). Assuming that, as they say, they seek to act as neutral information distributors, there remain at least four important concerns. Firstly, algorithms may subtly distort the newsgathering process, affecting what journalists learn in order to craft stories, and their ability to quantify the popularity and profitability of stories may push journalists and editors further towards prioritising audience numbers over public interest when choosing what to cover. Secondly, tech companies' focus on efficiency and customer satisfaction rather than on balance or the public interest may lead them to ignore potential problems in the ways in which their algorithms work for the public and for journalists. Thirdly, because their algorithms are secret and, to some extent, affect every user differently, there is no easy way for researchers, citizens or journalists to perceive and thereby seek to influence their effects. Lastly, media organizations in adjusting themselves to how they believe those algorithms to work may distort their news priorities even if the algorithms themselves are not in fact working in that way.

Journalists and scholars must come to terms with the pervasiveness of the power of the algorithm in influencing a broad range of journalistic activities and priorities. Just because the companies whose algorithms have such an impact may not be using the algorithmic power they have in the service of a political agenda or working in concert, it does not mean that it is any the less important to be aware of these new forces influencing journalistic practice. Gillespie may not be exaggerating when he says "that we are now turning to algorithms to identify what we need to know is as momentous as having relied on credentialed experts, the scientific method, common sense, or the word of God" (Gillespie 2014).

As Napoli and Caplan argue (2016), we must also encourage technology companies to recognise their own responsibilities (and, if necessary, use national governments to accomplish this[5]). Once they acknowledge that their algorithms affect journalism at almost every level,

these companies must develop an ethic that draws on the best ideals of journalism—at least insofar as their activities and algorithms are exercising an editorial role or influencing the editorial roles of others. They have the expertise to understand at least in part how their algorithms might be modified to provide, for example, better ideological balance or to ensure previously unheard voices are heard in the online public sphere, and should work in partnership with journalists, scholars and policymakers to achieve these ends.

Lastly, where necessary, these technology companies should be made accountable to third parties for the effects of their algorithms, whether these be scholarly ethics committees or legislative bodies. This implies some disclosure of the ways in which their algorithms work, even if only in a controlled fashion (just as we struggle to find ways to regulate spy agencies while not revealing to the world all of their secrets).[6] This task will not be easy and it is not clear what the right choices will be, but it is vital to start asking the right questions.

Notes

1. Although it is outside of the scope of this chapter to go into detail, it is also worth noting that other algorithms are increasingly being used alongside or instead of journalists to produce simple news stories—for example, summaries of sporting events or basic analysis of financial results. The cofounder and CTO of Narrative Science predicts that in 15 years, more than 90% of news will be written by algorithm (Levy 2012). Bakker (2012) has described this shift and outlined some of the potential consequences.
2. Although two vice presidents at Facebook announced in October a policy to "begin allowing more items that people find newsworthy, significant, or important to the public interest" (Kaplan and Osofsky 2016), which would seem to imply a journalistic organization somewhere in Facebook.
3. Although in the case of search engines in authoritarian regimes, biases may be more deliberately introduced, for example where China's Baidu search engine controls the visibility of Chinese Internet events, most likely at the behest of the Chinese government (Jiang 2012).
4. http://revealproject.eu.
5. In the European Union, the passage in May 2016 of the General Data Protection Regulation may be an important first step. This forces software companies to explain to users the algorithmic logic behind decisions made about them and allows for fines of up to 4% of a company's global revenues for noncompliance (Goodman and Flaxman 2016).

6. As this chapter was being completed, Google, Facebook, Amazon, IBM and Microsoft announced a "Partnership on AI to benefit people and society" (http://www.partnershiponai.org), whose tenets include making AI research "actively engaged with and accountable to a broad range of stakeholders" and making AI's operations "understandable and interpretable by people". If they are prepared to govern their public relevance algorithms using this framework it would be an important step forward.

References

Anderson, C. W. (2013). Towards a sociology of computational and algorithmic journalism. *New Media & Society, 15*(7), 1005–1021. doi:10.1177/1461444812465137.

Backstrom, L. (2013, August 6). *News Feed FYI: A window into News Feed.* Facebook. Available from: https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed [January 1].

Bakker, P. (2012). Aggregation, content farms and huffinization: The rise of low-pay and no-pay journalism. *Journalism Practice, 6*(5–6), 627–637. doi:http://dx.doi.org/10.1080/17512786.2012.667266.

Bakshy, E., Messing, S., & Adamic, L. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science, 348*(6239), 1130–1132. doi:http://dx.doi.org/10.1126/science.aaa1160.

BBC News Online. (2006, February 6). BMW given Google "death penalty". BBC News Online. Available from: http://news.bbc.co.uk/2/hi/4685750.stm. [May 30].

Beckett, C. (2008). *SuperMedia: Saving journalism so it can save the world.* Malden, MA: Blackwell.

Bhatia, R. (2016, May 12). The inside story of Facebook's biggest setback. *The Guardian.* Available from: https://www.theguardian.com/technology/2016/may/12/facebook-free-basics-india-zuckerberg.

Blank, G. (2016). The digital divide among Twitter users and its implications for social research. *Social Science Computer Review.* doi:http://dx.doi.org/10.1177/0894439316671698.

Brake, D. R. (2013). Journalists, user generated content and digital divides. In J. Gordon, G. Stewart, & P. Rowinski (Eds.), *Br(e)aking the news: Journalism, politics and new media* (pp. 253–270). Oxford: Peter Lang.

Burns, T. (1977). *The BBC: Public institution and private world.* London: Macmillan.

Carr, N. (2006, March 10). *The clickthrough's tyrannical efficiency.* Available from: http://www.roughtype.com/archives/2006/03/the_tyranny_of_1.php.

Cherubini, F., & Nielsen, R. K. (2016). *Editorial analytics: How news media are developing and using audience data and metrics*. Reuters Institute for the Study of Journalism. Available from: http://digitalnewsreport.org/publications/2016/editorial-analytics-2016/.

Cohen, N. (2010, April 5). 'Suicide' query prompts Google to offer hotline. *New York Times*. Available from: http://www.nytimes.com/2010/04/05/technology/05google.html.

Dick, M. (2011). Search engine optimisation in UK news production. *Journalism Practice, 5*(4), 462–477. Available from: http://dx.doi.org/10.1080/17512786.2010.551020.

D'Onfro, J. (2016, August 29). *Facebook is telling the world it's not a media company, but it might be too late*. Business Insider. Available from: http://www.businessinsider.com/mark-zuckerberg-on-facebook-being-a-media-company-2016-8.

Donsbach, W. (2004). Psychology of news decisions. *Journalism, 5*(2), 131–157. doi:http://dx.doi.org/10.1177/146488490452002.

El-Arini, K. (2014, August 25). *News Feed FYI: Click-baiting*. Facebook. Available from: http://newsroom.fb.com/news/2014/08/news-feed-fyi-click-baiting/ [May 30].

Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences, 112*(33), E4512–E4521. Available from: http://www.pnas.org/content/112/33/E4512.abstract.

Feenberg, A. (1999). *Questioning technology*. London: Routledge.

Feenberg, A. (2008). From critical theory of technology to the rational critique of rationality. *Social Epistemology, 22*(1), 5–28. doi:http://dx.doi.org/10.1080/02691720701773247.

Gans, H. J. (1979). *Deciding what's news: A study of CBS evening news, NBC nightly news, Newsweek and Time*. London: Constable.

Gillespie, T. (2012). Can an algorithm be wrong? *Limn*, 2. Available from: http://limn.it/can-an-algorithm-be-wrong.

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society*. Cambridge, MA: MIT Press.

Goodman, B., & Flaxman, S. (2016). *European Union regulations on algorithmic decision-making and a 'right to explanation'*. New York: ICML Workshop on Human Interpretability in Machine Learning. Available from: https://arxiv.org/abs/1606.08813.

Google. (n.d.-a). *About Google*. Available from: http://www.google.com/about.

Google. (n.d.-b). *Getting into Google News*. Available from: https://support.google.com/news/publisher/answer/40787?hl=en. [May 31].

Herbst, J. (2016). The algorithm is an Editor. *The Wall Street Journal*. Available from: http://www.wsj.com/articles/the-algorithm-is-an-editor-1460585346.

Hermans, L., Vergeer, M., & D'Haenens, L. (2009). Internet in the daily life of journalists: Explaining the use of the internet by work-related characteristics and professional opinions. *Journal of Computer-Mediated Communication, 15*(1), 138–157. doi:http://dx.doi.org/10.1111/j.1083-6101.2009.01497.x.

Hermida, A. (2010). From TV to Twitter: How ambient news became ambient journalism. *M/C Journal*, *13*(2). Available from: http://www.journal.media-culture.org.au/index.php/mcjournal/article/view/220.

Hermida, A., Fletcher, F., Korell, D., & Logan, D. (2012). Share, like, recommend. *Journalism Studies, 13*(5–6), 815–824. doi:10.1080/1461670X.2012.664430.

Hindman, M., Tsioutsiouliklis, K., & Johnson, J. A. (2003). 'Googlearchy': How a few heavily-linked sites dominate politics on the web. In *Annual Meeting of the Midwest Political Science Association*, Chicago. Available from: https://www.researchgate.net/publication/228485022_Googlearchy_How_a_Few_Heavily-Linked_Sites_Dominate_Politics_on_the_Web.

Jiang, M. (2012). The business and politics of search engines: A comparative study of Baidu and Google's search results of internet events in China. *SSRN eLibrary*. Available from: http://ssrn.com/paper=2027436.

Kaplan, J., & Osofsky, J. (2016, October 21). *Input from community and partners on our community standards*. Facebook. Available from: http://newsroom.fb.com/news/2016/10/input-from-community-and-partners-on-our-community-standards/ [October 21].

Kirkland, S. (2014, May 19). *3 takeaways from the "death of the homepage" and The New York Times innovation report*. Poynter. Available from: http://www.poynter.org/2014/3-takeaways-from-the-death-of-the-homepage-and-the-new-york-times-innovation-report/252632 [May 28].

Kirkpatrick, D. (2010). *The Facebook effect: The inside story of the company that is connecting the world* (1st Simon & Schuster hardcover ed.). New York: Simon & Schuster.

Levy, S. (2012). Can an algorithm write a better news story than a human reporter? *Wired*. Available from: http://www.wired.com/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter.

Lichterman, J. (2016). *Survey of large publishers: 30 percent of our website visits come from Facebook*. Available from: http://www.niemanlab.org/2016/09/survey-of-large-publishers-30-percent-of-our-website-visits-come-from-facebook. [September 30].

Lippmann, W. (1922). *Public opinion*. London: Allen & Unwin.

Machill, M., & Beiler, M. (2009). The importance of the internet for journalistic research. *Journalism Studies, 10*(2), 178–203. doi:10.1080/14616700802337768.

McChesney, R. W. (2014). *Digital disconnect: How capitalism is turning the internet against democracy*. New York: The New Press.

Metz, C. (2010, December 16). *Google drops nuke on 'objective' search engine utopia*. Available from: http://www.theregister.co.uk/2010/12/16/google_algorithms_are_google_opinions. [May 23].

Morozov, E. (2011, August 4). Don't be evil. *The New Republic*. Available from: https://newrepublic.com/article/91916/google-schmidt-obama-gates-technocrats.

Moz.com. (n.d.). *Google algorithm change history*. Available from: https://moz.com/google-algorithm-change. [September 28].

Napoli, P. M., & Caplan, R. (2016). *When media companies insist they're not media companies and why it matters for communications policy*. Telecommunications Policy Research Conference, Arlington, VA. doi:10.2139/ssrn.2750148

Narayanan, A. (2016). *Language necessarily contains human biases, and so will machines trained on language corpora*. Available from: https://freedom-to-tinker.com/blog/randomwalker/language-necessarily-contains-human-biases-and-so-will-machines-trained-on-language-corpora/ [August 26].

Nguyen, A. (2013). Online news audiences: The challenges of web metrics. In K. Fowler-Watt & S. Allan (Eds.), *Journalism: New Challenges*. Bournemouth: Centre for Journalism and Communication Research, University of Bournemouth. Available from: http://microsites.bournemouth.ac.uk/cjcr/publications/journalism-new-challenges/.

Nunez, M. (2016a). *Facebook employees asked Mark Zuckerberg if they should try to stop a Donald Trump presidency*. Gizmodo. Available from: http://gizmodo.com/facebook-employees-asked-mark-zuckerberg-if-they-should-1771012990 [September 24].

Nunez, M. (2016b, November 18). *Facebook's fight against fake news was undercut by fear of conservative backlash*. Available from: https://gizmodo.com/facebooks-fight-against-fake-news-was-undercut-by-fear-1788808204 [November 18].

Olsen, S. (2003, August 27). Project searches for open-source niche. *CNet*. Available from: http://www.cnet.com/news/project-searches-for-open-source-niche/ [August 27].

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy* (1st ed.). New York: Crown.

Pelegrin, W. (2015). *Upworthy unleashed clickbait on the internet, but now it wants to take it back*. Digital Trends. Available from: http://www.digitaltrends.com/web/upworthy-cofounder-admits-clickbait-is-bad [May 30].

Perrin, A. J. (2015). *Social media usage: 2005–2015*. Pew Internet & American Life Project. Available from: http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015.

Pew Research Center's Project for Excellence in Journalism. (2016). *The state of the news media 2016: Digital news audience fact sheet*. Pew Research Center's

Project for Excellence in Journalism. Available from: http://www.journalism.org/2016/06/15/digital-news-audience-fact-sheet.

Poell, T., & Van Dijck, J. (2014). Social media and journalistic independence. In J. Bennett & N. Strange (Eds.), *Media independence: Working with freedom or working for free?* (pp. 182–201). London: Routledge.

Purcell, K., Brenner, J., & Rainie, L. (2012). *Search engine use 2012.* Pew Internet & American Life Project. Available from: http://www.pewinternet.org/2012/03/09/search-engine-use-2012.

PwC. (2016). *IAB internet advertising revenue report.* Available from: http://www.iab.com/insights/iab-internet-advertising-revenue-report-conducted-by-pricewaterhousecoopers-pwc-2.

Rogers, R. (2004). *Information politics on the web.* Cambridge, MA: MIT Press.

Roth, D. (2009). The answer factory: Fast, disposable, and profitable as hell. *Wired, 17*(11). Available from: http://www.wired.com/magazine/2009/10/ff_demandmedia/all/1.

Rowan, D. (2014, February). How BuzzFeed mastered social sharing to become a media giant for a new era. *Wired UK.* Available from: http://www.wired.co.uk/article/buzzfeed.

Santana, A. D. (2016). Controlling the conversation. *Journalism Studies, 17*(2), 141–158. doi:10.1080/1461670X.2014.972076.

Shoemaker, P. J. (2006). News and newsworthiness: A commentary. *Communications, 31*(1), 105–111. doi:10.1515/COMMUN.2006.007.

Thelwall, M., & Vaughan, L. (2004). Search engine coverage bias: Evidence and possible causes. *Information Processing & Management, 40*(4), 693–707. Available from: http://www.scit.wlv.ac.uk/~cm1993/papers/search_engine_bias_preprint.pdf.

Valinsky, J. (2016, March 1). Facebook tweaks News Feed algorithm to give preference to live videos. *Newsday.* Available from: http://digiday.com/platforms/facebook-tweaks-news-feed-algorithm-give-preference-live-videos. [May 30].

Van Couvering, E. (2007). Is relevance relevant? Market, science, and war: Discourses of search engine quality. *Journal of Computer-Mediated Communication, 12*(3), 866–887. doi:10.1111/j.1083-6101.2007.00354.x.

Vujnovic, M., Singer, J. B., Paulussen, S., Heinonen, A., Reich, Z., Quandt, T., et al. (2010). Exploring the political-economic factors of participatory journalism. *Journalism Practice, 4*(3), 285–296. doi:10.1080/17512781003640588.

Wells, G. (2016, September 6). Facebook's 'trending' feature exhibits flaws under new algorithm. *Wall Street Journal.* Available from: http://www.wsj.com/articles/facebooks-trending-feature-exhibits-flaws-under-new-algorithm-1473176652.

Zittrain, J. (2014, June 1). Facebook could decide an election without anyone ever finding out. *The New Republic*. Available from: https://newrepublic.com/article/117878/information-fiduciary-solution-facebook-digital-gerrymandering.

Zuckerberg, M. (2016). *A lot of you have asked what we're doing about misinformation, so I wanted to give an update*. Available from: https://www.facebook.com/zuck/posts/10103269806149061 [November 20].

Author Biography

David R. Brake is a journalist and independent scholar, trained at the LSE. He lived and worked in the UK for 27 years but post-Brexit lives in Newfoundland, Canada. His research interests include how the advent of online and social media is changing the landscape of journalism, the political economy of journalism, opportunities for creative self-expression and the digital divide, and how people imagine their audiences on social media. His most recent book is "Sharing Our Lives Online: Risks and Exposure in Social Media", and a 2013 chapter on the use of social media by journalists can be found in "Br(e)aking The News: Journalism, Politics and New Media". More information and links to his writing are available at http://davidbrake.org/.