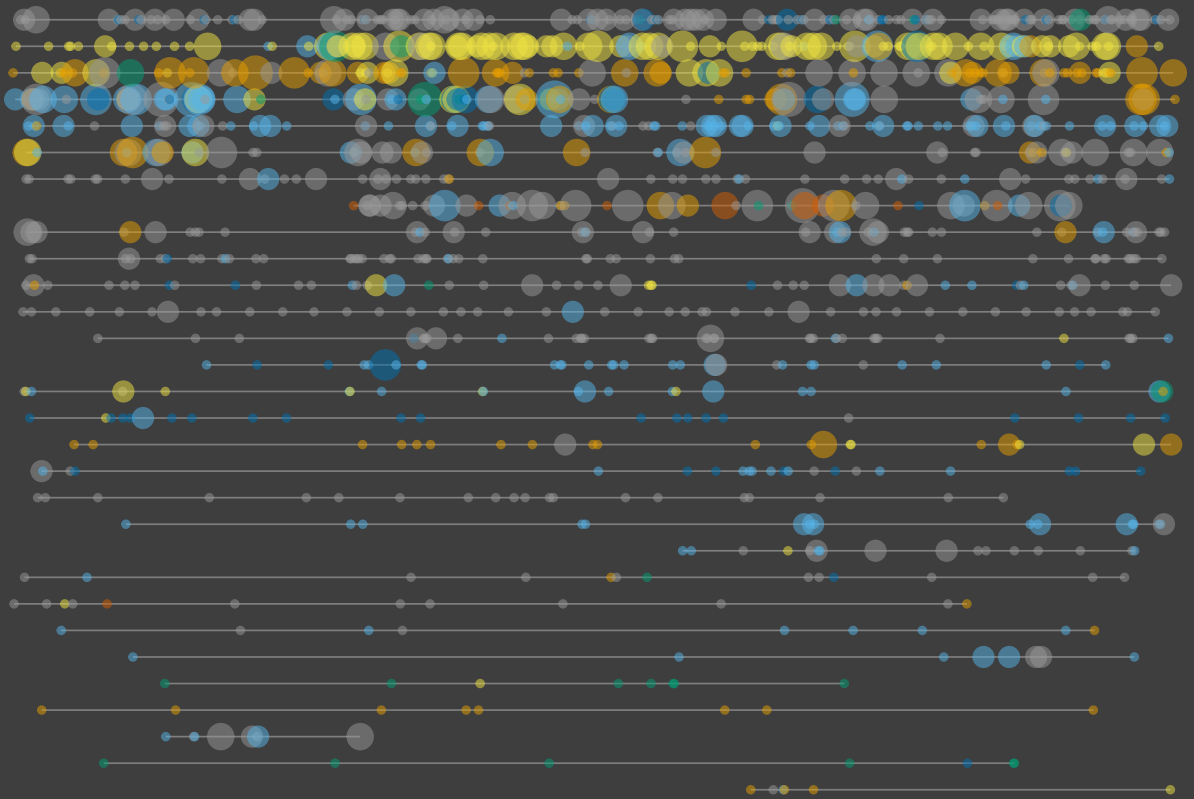


Adam Åbonde Garke

ONLINE NEWS BEHAVIOR



ONLINE NEWS BEHAVIOR

Technological developments in the last decades have fundamentally transformed the news industry and the way citizens interact with media. Personal computers and smartphones abound, and internet has become the most common way citizens access news. Social media platforms and search engines have become both gateways and gatekeepers for finding and accessing content, their algorithms routinely relied upon for distributing stories. Online, news competes for attention with a flood of content.

The move online presents remarkable opportunities for both studying and influencing news behaviors. But despite growing online audiences and the promises of methodological advancements, news companies are struggling to adapt to the new reality they have come to find themselves in. This thesis offers new insights into what online news behaviors of both humans and algorithms look like empirically and illustrate how those can be studied.

The dissertation consists of five original research studies which explores methods and patterns in online news activity. The studies cover several aspects of online news behavior, such as distribution and consumption of news content and the effectiveness of presenting advertising on news websites. The thesis also presents a theoretical position explaining the changes we see in news behaviors and offers a suggested path for news media going forward; focus on function over content when deciding what and how to publish. The thesis has important implications for news professionals, media researchers, and practitioners working with online behaviors.



ADAM ÅBONDE GARKE is a researcher at the Center for Media and Economic Psychology. His research focuses on human and machine behavior, primarily related to digital media.

ISBN 978-91-7731-229-1

DOCTORAL DISSERTATION IN BUSINESS ADMINISTRATION
STOCKHOLM SCHOOL OF ECONOMICS, SWEDEN 2022



Online News Behavior

Adam Åbonde Garke

Akademisk avhandling

som för avläggande av ekonomie doktorsexamen
vid Handelshögskolan i Stockholm
framläggs för offentlig granskning
fredagen den 13 maj 2022 kl 13.15,
sal Torsten, Handelshögskolan,
Sveavägen 65, Stockholm



Online News Behavior

Online News Behavior

Adam Åbonde Garke





Dissertation for the Degree of Doctor of Philosophy, Ph.D.,
in Business Administration
Stockholm School of Economics, 2022

Online News Behavior

© SSE and the author, 2022

ISBN 978-91-7731-229-1 (printed)

ISBN 978-91-7731-230-7 (pdf)

Front cover illustration:

© Adam Åbonde Garke, 2022

Back cover photo:

Personal

Printed by:

BrandFactory, Gothenburg, 2022

Keywords:

News media, online behavior, web tracking data, algorithms, eye tracking,
behavior science

Foreword

This volume is the result of a research project carried out at the Department of Marketing and Strategy at the Stockholm School of Economics (SSE).

The volume is submitted as a doctoral thesis at SSE. In keeping with the policies of SSE, the author has been entirely free to conduct and present his research in the manner of his choosing as an expression of his own ideas.

SSE is grateful for the financial support provided by the Bonnier Family Foundation for Scientific Research and Education, which has made it possible to carry out the project.

Göran Lindqvist

Director of Research
Stockholm School of Economics

Hans Kjellberg

Professor and Head of the
Department of Marketing and Strategy

Acknowledgements

A lot of people have contributed to making this dissertation possible, and my time writing it more interesting and enjoyable. I thank my supervision committee Drs Richard Wahlund, Rickard Sandberg, and Lise Bergman Nordgren, for their time, advice, and guidance. Richard gave me encouragement and freedom to explore different research paths and ideas, and opened many doors allowing me to do so. His never-ending optimism, engagement and genuine curiosity are inspiring. Rickard's advice and support in statistics has been invaluable. He taught me a lot and has given me confidence with data. Lise was a voice of reason in a mainly mentalistic and theory-laden context. She reinforced me to focus on the things I saw as interesting and important in my research, for which I am thankful.

I am very grateful to the Familjen Bonniers Stiftelse för Vetenskaplig Forskning och Utbildning¹ for generous financial support. I thank Sara Öhrvall, whose initiative to create a unique collaboration between industry and academia to study the future of news journalism in the digital media world made much of the research presented in this dissertation possible. I thank the organizations that enabled my research in various ways: Sveriges Television, Sveriges Radio, Dagens Nyheter, Aftonbladet, Svenska Dagbladet, Expressen, Uppsala Nya Tidning, Nepa, VINNOVA, Newstag, and Tobii. I am thankful to Camilla Andersson Dahlin and Henrik Eklund at Newstag for allowing us to use their platform for the experiments in study V (and Camilla for being a co-author on the same paper).

Many friends and colleagues at Stockholm School of Economics have supported me in various ways, from theoretical discussions to laughs over coffee. I have come to view Emelie Fröberg as my informal mentor on many

¹ The Bonnier Family Foundation for Scientific Research and Education (author's translation).

things related to academia. Her support has meant a lot. Thank you also to Patric Andersson, Sara Rosengren, Magnus Söderlund, Udo Zander, Nina Åkestam Wikner, Nurit Nobel, Per Henrik Hedberg, Claire Ingram Bogusz, Mattias Svahn, Henrik Glimstedt, Dante Holmberg, Dan Hjalmarsson, Micael Dahlen, and Erik Modig for being fantastic colleagues. Gustav Almqvist is an inspiration both in thinking and writing and I have enjoyed our many discussions. I am grateful to Jonas Colliander for general advice on academic matters, and acting as mock opponent and giving feedback on how the dissertation could be improved. I thank Agneta Carlin for all those early morning talks in the kitchen, keeping me on track, getting me to Japan, and so much more. Without her the department would crumble.

I also thank Dr David Lazer for welcoming me to his lab, and to spend a semester as visiting scholar at the Network Science Institute at Northeastern University in Boston.

My parents, Kerstin and Peter, have always supported and encouraged me. There is so much I find interesting in this world, and they have most certainly helped nurture that curiosity of mine. I thank my siblings, Anja, Anton, and André, for keeping our home lively while growing up – I now rarely have trouble shutting out the outside world when I need to focus – and challenging my thinking with new perspectives.

Finally, I cannot express enough gratitude to my wife Maria for listening, questioning, explaining, proof-reading, and discussing. Without her patience and sharp mind this thesis would have been much harder to complete. Maria, I love you and look forward to many more stimulating conversations, alone as well as together with our Alma.

Stockholm, March 2022

Adam Åbonde Garke

Table of Contents

1. Introduction.....	1
2. Background.....	5
2.1 <i>Online news consumption</i>	7
2.2 <i>The business of online news</i>	10
2.3 <i>Distribution of online news content</i>	12
3. Methodology.....	15
3.1 <i>What is online news behavior?</i>	15
3.2 <i>How to measure online news behavior</i>	20
3.3 <i>A note on machine behavior</i>	24
3.4 <i>Data and materials</i>	26
3.5 <i>Ethical considerations</i>	28
4. Introducing studies I–V	29
5. Discussion.....	37
5.1 <i>Brief summary of the main findings</i>	37
5.2 <i>Contributions to the literature</i>	38
5.3 <i>Implications for practice</i>	42
6. Limitations	47
7. Future directions	49
8. Conclusion	53
9. References.....	55

Original research studies

I. Explaining the Behaviour of News Consumption 69

II. Understanding Online News Reading: Analyzing Individuals’ Cross-Website Behavior Through Tracking Data..... 83

III. Identifying News Audience Clusters Based on E-Eigenbehaviours: Behavioural Pattern Recognition in Online News Browsing Data. 121

IV. Always-on Eye Tracking Reveals Problems with Financing News Websites Through Advertising..... 147

V. Can Two Wrongs Make a Right? An Empirical Study of Algorithmic Bias in the Delivery of Ads on Facebook And Google..... 165

List of selected figures

Study II, Figure 1	91
There is always someone reading news online. Aggregated average news activity per hour, device, and day type (weekday/weekend).	
Study II, Figure 3	94
Many have regular news reading behaviors. Gauging the regularity in individuals' news consumption through cumulative records.	
Study III, Figure 2	132
Daily news habits of different audience clusters. Using behavioral pattern recognition to segment news audiences.	
Study IV, Figure 1	150
News websites are less effective in showing ads to visitors compared to other online platforms. Viewability funnel for online ads at various webpages.	
Study V, Figure 2	184
Assessing algorithmic bias in the distribution of news videos on Facebook. Observed and expected gender proportions being exposed to news videos.	
Study V, Figure 3	185
Assessing algorithmic bias in the distribution of news videos on Google. Observed and expected gender proportions being exposed to news videos.	

Chapter 1

Introduction

Technological advancements are transforming the news industry² and developments in both these arenas shape our media use behaviors³. In a world that continues to digitize and evolve, what does online news behavior look like and how can we study it?

News media serves an important role in our society according to a vast body of literature in communications research, journalism studies, and political science. Media has the power to influence voting behaviors (Bond et al., 2012; Boulianne, 2020; DellaVigna & Kaplan, 2007) and steer public discussion (King et al., 2017; McCombs & Shaw, 1972). Exposure to news can increase political knowledge and engagement (Boulianne, 2020; Mitchelstein & Boczkowski, 2010; Prior, 2005) and help citizens make informed decisions (Kovach & Rosenstiel, 2014, pp. 13–45). While news remains an important societal institution, the medium where it is presented continues to shift online. Indeed, as print circulation numbers has continued to dwindle (N. Newman et al., 2021; Ohlsson et al., 2020), online news reading becomes ever more prevalent (Mitchelstein & Boczkowski, 2010; Prawesh &

² The American Press Institute defines *news* as “that part of communication that keeps us informed of the changing events, issues, and characters in the world outside” (Dean, 2013).

³ *Media use* refers to “the intended or incidental selection of media types (e.g., TV, video games, the Internet), content (e.g., entertainment, advertising, news), and technologies (e.g., social media)” (Valkenburg et al., 2016, p. 316)

Padmanabhan, 2014). In recent years, the internet has even overtaken TV as the most frequently used source of news (N. Newman et al., 2021).

Despite growing online audiences, news companies are struggling to find viable business models on the internet (Sen, 2019). Some companies have experimented with various variants of paywalls to try and force readers to pay for what they consume online, as it has been difficult to attract enough online subscribers to offset decreasing offline revenues (Oh et al., 2016; Pickard & Williams, 2014; Wahlund et al., 2013). Implementing paywalls, however, also cannibalize on ad revenues due to the limiting effect it has on audience sizes (Lambrecht & Misra, 2016). Advertising, which historically has enabled newspapers to gain political independence by offering a strong revenue stream (Petrova, 2011), may still be one of the most attractive business models online. Yet, while newspapers theoretically should be able to earn more on their online ad space compared to offline, e.g., due to the larger populations possible to reach and the improved targeting opportunities provided by the internet, the competition for ad dollars is also fiercer, which may lead to overall lower revenues (Gentzkow, 2014; Wahlund et al., 2013). The increasing reliance on machines⁴ for operations and as decision support further add to the confusion and complexity in large parts the online news industry (Diakopoulos, 2019; Sen, 2019). All in all, few news companies have managed to become successful in the digital world (Fletcher & Nielsen, 2017b; Küng, 2016). Admittedly, some actors have seen increases in their number of online news subscribers lately, but on a global scale the numbers of paying readers are still at very low levels online (N. Newman et al., 2021). It is clear practitioners need help.

Researchers have tried to provide some respite. There is a growing body of literature on online news as a result. Large efforts from both practitioners and academics have focused on analyzing the types of content that are being published in online news outlets (Anand, 2016; Berger & Milkman, 2012; Toff & Kalogeropoulos, 2020), as well as the effects of consuming online media (Boulianne, 2020; Valkenburg et al., 2016). However, research on both

⁴ Here, machines refer to algorithms and artificial intelligence (AI) systems.

news content and media effects presume there are related media use behaviors. Increasingly, researchers in the social sciences are pointing to the potential of using inductive research approaches (Yarkoni & Westfall, 2017) and larger datasets (Lazer et al., 2009) to further our understanding of human behavior. Alas, much of the published empirical research on media use behaviors is based on either self-report measures (de Vreese & Neijens, 2016; Prior, 2009a) or aggregated behavioral data (Mukerjee et al., 2018; Nelson & Lei, 2018; Taneja, 2016), both of which risk misrepresenting actual media use behaviors. Good empirical data at the level of the individual have, unfortunately, been hard to come by for researchers and more detailed analyses of online news behaviors are called for (Mukerjee et al., 2018). That is the venture point for this thesis.

This thesis consists of five original research studies: four empirical papers and one conceptual book chapter. The thesis was initiated as part of a larger project entitled “The future of news journalism in the digital media market” (in short, the Future of Journalism). The Future of Journalism was a unique collaboration between a consortium of large Swedish news publishers, both public service and from the private sector, and academia.⁵ The overarching aim of that project was “to find a path toward a more sustainable journalism” against a backdrop of rapid digitalization in the media market which had disrupted large parts of the industry (Öhrvall, 2015, p. 3). A goal in this thesis is to contribute towards that objective by studying – and hopefully improving our understanding of – different facets of online news behavior.

Indeed, the central question, and common denominator of the research studies in this thesis, is what online news behavior look like at the level of the individual. The inevitable shift online transforms large parts of the news ecosystem (Küng, 2016; Sen, 2019). The intention with this thesis is to try and improve our understanding of what has changed and offer some suggestions for how news behaviors can be studied in the digital media market. I strive to do this, not primarily through theoretical reasoning (although this is

⁵ News publishers: Dagens Nyheter, Svenska Dagbladet, Aftonbladet, Expressen, Uppsala Nya Tidning, Sveriges Radio, and Sveriges Television; Academia: Stockholm School of Economics. The market research company Nepa assisted with data collection.

the focus of study I), but by providing a robust empirical baseline of what different online news behaviors look like ‘in the wild’ (studies II–V).

The purpose of this thesis is to improve our understanding of what online news behavior of both man and machine look like empirically and how those can be studied.

In the next chapter I review the online news literature. The focus is on central parts of the online news industry, specifically how news is consumed, financed, and distributed. Next, I will treat important methodological aspects in online news behavior research, such as what and how to measure media use and how to study algorithms, on which the news industry becomes increasingly reliant (Diakopoulos, 2019). After this, the individual research studies which make up this thesis are presented. Finally, there is a general discussion including main findings, contributions, and implications of the thesis. Limitations and suggested future directions for research are also presented before concluding the thesis.

Chapter 2

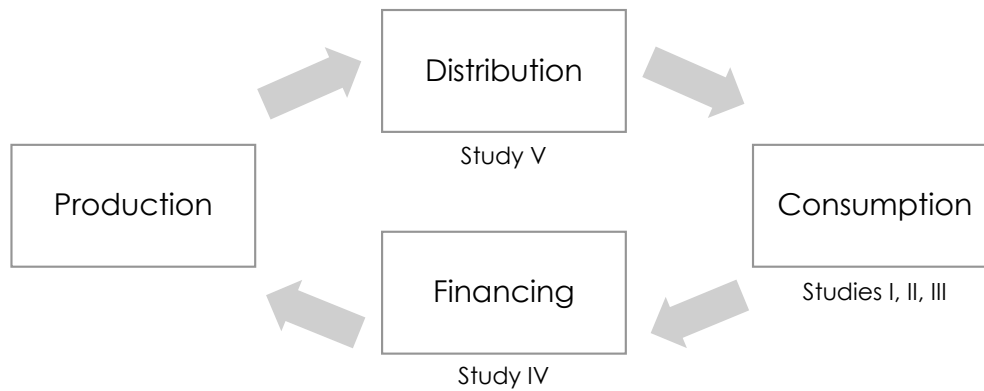
Background

At the heart of the news industry is consumption of the content that is being produced. However, for there to be any content to consume, news media depend on having a viable business model to sustain their operations and an organization for distributing news content to consumers. The research studies included in this thesis address these aspects of consumption, distribution, and financing of online news.⁶ Study I provides a theoretical perspective on why news behaviors are changing, while studies II–V provide empirical evidence on how. An overview of the focus of the research studies is presented in Figure 1.

Many have pointed out that the medium where media is presented and encountered influence both individual human behavior and society at large. McLuhan (1964/2010), for example, famously argued that the introduction of any innovation (such as new media environments) can have large, often unforeseen, consequences both personally and socially. Postman (1985/2006) worried about the future of educated public discussion, as he believed electronic media was habituating citizens to bit-sized visual entertainment, rendering them incapable of focusing on longer news segments or journalistic pieces. Pennypacker (1992) also suggested that more research be aimed at mapping how the press shapes cultural practices, as he argued that live TV, for example, could influence the behavior of politicians.

⁶ This being a thesis in business administration, not in journalism or communication studies, I have not focused the journalistic process, i.e., how news content is produced.

Figure 1. The news industry and the focus of the studies



The figure provides a simplified overview of the online news industry and illustrates the aspect each of the five research papers addresses.

That behaviors are influenced by factors in the environment is well known. Over a century of research in behavioral science bears witness of the powers that the environment exerts over both humans and other species (e.g., Bouton, 2016; Skinner, 1953/1965). Studies in computer science and information systems have also shown how technology, seen as a type of context, can change what users think and do (Fogg, 2003). In the physical world the context is (inevitably) messy. The researcher’s job of identifying the variables that influence behavior thus becomes difficult. But internet, social media, and smartphones have made media omnipresent (Dill-Shackleford, 2014, pp. 3–4; Valkenburg et al., 2016). Increasing amounts of time is spent in the digital world – the “21st century Skinner Box” (Robertson, 2018) – where designers are able to control every aspect of the environment⁷ (R. Epstein & Robertson, 2015; Fogg, 2003; Fogg & Eckles, 2007; Wu, 2017).

⁷ Of course, even though the digital world can be fully controlled, the consumer still finds themselves in a variable physical environment which may influence observed digital behaviors.

The move from printed media to online thus likely changes, as well as present remarkable opportunities for studying, news behaviors. Literature related to the consumption, distribution, and financing of online news will now be discussed in turn.

2.1 Online news consumption

Online news reading has become more prevalent in recent years (Mitchelstein & Boczkowski, 2010; Ohlsson et al., 2020; Prawesh & Padmanabhan, 2014), and recently overtook TV as the most frequently used source of news (N. Newman et al., 2021). Some have suggested that online news acts as a substitute for traditional news media (Jang & Park, 2016) and, overall, researchers have found that online news consumption patterns are similar to those offline: Just as with consumption of printed newspapers, most online news consumption is centered around a few mainstream media outlets (Mukerjee et al., 2018; Nelson, 2020) and online consumers typically visit the home page of their favorite outlet when they look for news (Flaxman et al., 2016; Van der Meer et al., 2020). Researchers have found that other factors – e.g., type of content, reader gender, and interests – influence news consumption patterns more than if the news appear online or in print (d’Haenens et al., 2004). There are, for example, differences between age groups as to what platform individuals use to find and consume online news. A recent report from Sweden found that younger individuals favor social media for online news consumption, while older favor the digital versions of public service news outlets and daily newspapers (Ohlsson et al., 2020).

When media technologies (such as computers or smartphones) are compared against each other, it has been suggested that characteristics of different technologies fosters different behaviors among its users (Molyneux, 2018; Nelson & Lei, 2018). Mobile devices have in many aspects – functionality, availability of content, flexibility of platforms etc. – been seen as inferior compared to computers for accessing information online (Napoli & Obar, 2014). Some have found that mobile users spend less time on news compared with computer users (Dunaway et al., 2018). On the other hand, the portability and accessibility of mobile devices encourage the behavior of ‘second

screening.’ This is the practice of consuming media on two devices simultaneously; for example, watching something on TV, while discussing and commenting on the same event in real-time on social media platforms (Gil de Zúñiga et al., 2015; Vaccari et al., 2015). Like physical features, digital interfaces can also influence users’ interactions with technology. Among mobile users, for example, researchers have found that there is a difference between accessing news via an app and via the browser, with app users engaging more deeply with news content compared to browser users (Dunaway et al., 2018; Nelson & Lei, 2018).

News consumers typically claim they want control over their media diet, in the sense that they want access to all available information (Groot Kormelink & Costera Meijer, 2014), and are skeptical towards selection of content made both by editors and algorithms (Fletcher & Nielsen, 2019). At odds with this, more and more people are turning to social media to consume news. One reason for this migration to social media is that consumers can feel informed and up to date with what is happening in the world without having to visit multiple websites (Hermida et al., 2012; Müller et al., 2016). Social media also allows for incidental exposure to news content, i.e., when the user is primarily focused on other aspects of a platform but is exposed to news as a side-effect (Fletcher & Nielsen, 2018). The feeling of being well-informed, which is amplified by mere quantity of exposure rather than the amount of news actually consumed, promotes replacing other news sources with social media (Müller et al., 2016).

As people get a wider variety of media to choose among, they tend to favor content that is aligned with their interests (Prior, 2005). One study found that the news category *local* was the most popular one, and that readers, in a majority of cases, stay within the same category when they move on to the next news article (Esiyok et al., 2014).

It is common to distinguish between hard and soft news in communication research – roughly corresponding to an index with topics like politics, foreign policy, and business news at one end, and celebrity gossip, drama, and entertainment at the other (e.g., M. A. Baum, 2002; Fletcher & Nielsen, 2019; Prior, 2003). Oftentimes, readers consume much less hard news than they say they do (Flaxman et al., 2016; Tewksbury, 2003), and consume less news overall than they think they do (Prior, 2009a, 2009b).

A natural question to ask in relation to high-choice digital media, is the risk of audience fragmentation across news sources (Prior, 2005). The prevalence of “filter bubbles” where consumers self-select into a narrow set of partisan news outlets has, however, largely been refuted (Arguedas et al., 2022; Flaxman et al., 2016; Fletcher & Nielsen, 2017a; Gentzkow & Shapiro, 2011). Researchers have suggested that a major force working against audience fragmentation at a larger scale is that most news consumers get their news from a few large, legacy outlets. Those that consume smaller amounts of news only visit major sites, while those that consume larger quantities of news – while they may complement their media diet with smaller, more niched outlets – also visit the major outlets (Gentzkow & Shapiro, 2011; Mukerjee et al., 2018). One reason for these behavioral patterns is that the infrastructure of the online news environment (e.g., with few mainstream outlets getting most attention online and social media algorithms curating content for audiences) tend to homogenize news usage across social groups (Taneja et al., 2018). For example, there is evidence that news aggregators, search engines and social media increase users’ exposure to news from more varied sources (Fletcher et al., 2021; Scharkow et al., 2020).

While it seems that online news consumption is becoming more prevalent overall, there has been little discussion in the literature regarding the distribution of consumption behaviors in the datasets used. Most of the published research into online news resorts to analyzing and presenting aggregated summary measures, such as the grand mean. It is generally acknowledged that only reporting summary measures risk masking important features of the data (Anscombe, 1973). Much of human activity is unevenly distributed and skewed (M. E. J. Newman, 2005), like hospital patients’ number of visits to emergency departments (Neelon et al., 2013) or reaction times in psychology experiments (Ratcliff, 1993). As Huberman and colleagues (1998) noted already two decades ago, online browsing is no exception; web surfing seems to be better approximated by a power law distribution rather than a bell curve (i.e., most internet browsing sessions consist of few clicks, while there is a long tail with browsing sessions involving many subsequent clicks). There is reason to believe this would hold also for online news consumption. While a few authors (e.g., Bentley et al., 2019; Möller et al., 2019; Vermeer et al., 2020) do mention that their data on online news behaviors are skewed,

they still focus on reporting and interpreting mean values and further discuss the implications thereof. This way of presenting data effectively amounts to assuming all news consumers behave in the same way, while it is likely that there would be some variation.

It is easy to come up with explanations for things that have already happened (e.g., news consumers turning to social media because it makes them feel informed, Müller et al., 2016); it is much harder to predict what will happen next. In study I, I outline one theoretical perspective which I believe could prove useful in explaining news behaviors – both the ongoing migration from print to online, as well as future changes in behaviors. Additionally, although it is known that legacy media outlets still make up the core of people’s news diets online (Mukerjee et al., 2018; Vermeer et al., 2020), it is unclear how prevalent online news behaviors are in the general population and what characterizes individuals’ news consumption, as well as what the differences between individuals look like. These questions are the focus of studies II and III.

2.2 The business of online news

News media have traditionally been associated with large, fixed costs for production (e.g., journalists salaries and printing) and distribution (e.g., transporting the physical newspapers to kiosks and newsstands) of content (Anand, 2016; Bogart, 1981; Sen, 2019). The revenues have primarily been attributable to two sources: selling newspapers and advertising space (Bogart, 1981; Wahlund et al., 2013; Wu, 2017). Digitalization of the industry have upheaved both sides of this business equation.

As the news industry has been increasingly digitized over the last two and a half decades, much of the fixed costs have been cut (Anand, 2016). For example, as larger parts of the business moves online, there is obviously less need for printing and physical distribution (Lakomaa, 2012). The salaries of journalists must still be accounted for, but simple and repetitive tasks (such as stock reports or the outcome of sports games) can be outsourced to algorithms which decreases the need for a large newsroom staff (Diakopoulos, 2019). As it has become easier and cheaper to produce content, barriers to entry have fallen and the number of news producers have risen, leading to

more fragmented media markets and fiercer fights for the attention of audiences (Küng, 2016).

But just as costs have shrunken online, so have revenues, and most online news outlets are still searching for viable sources of revenue (Sen, 2019). Printed tabloids typically focused on single-copy sales, while printed broadsheets focused on selling subscriptions; both relied on advertising to subsidize the price of newspapers to readers (Bogart, 1981).⁸ Far fewer people have, thus far, been willing to pay for online news. People tend to expect information to be free online, and subscription numbers are much lower online than it used to be for printed newspapers (Andersson et al., 2020; N. Newman et al., 2021). In response to the low rates of people willing to subscribe online, some newspapers are experimenting with various paywalls to force readers to pay (Oh et al., 2016). But erecting paywalls can be detrimental to advertising revenues; as fewer people enter the site, fewer ‘eyeballs’ can be sold (Lambrecht & Misra, 2016). Furthermore, if you convince someone to subscribe to a digital service, one of their expectations besides gaining access to content is often then to get rid of the advertising.

Instead, consumers are, to an even greater extent online than offline, becoming the product being sold to advertisers, rather than the customer to newspapers (Wu, 2017). Thus far, newspapers have not been able to earn as much on advertising on the internet as they did offline, largely due to high competition and unreliable measurement of the effectiveness of online ads (Evans, 2009; Gentzkow, 2014). Yet, online advertising may still prove to be a potential way forward for news websites to become financially viable long-term. The online advertising industry is very large, has been growing fast for years, and shows no signs of slowing down (Choi et al., 2020).

Online advertising continues to develop and change as the technologies used to employ and measure its effectiveness continue to evolve and improve (Hu et al., 2015). The question is how newspapers will fare as measurement of online ad performance becomes ever better and more precise. On the one

⁸ Interestingly, journalists, whose livelihood to a large extent relies on advertising, have typically been skeptical towards ads in newspapers and only reluctantly accepted it as a “necessary evil” enabling them to carry out their public service (Conill, 2016, p. 912).

hand, ad space on online newspapers' websites is more expensive today than on other online platforms, supposedly because of the more attractive audience (e.g., higher education and income) that it can offer compared to internet in general (Gentzkow, 2014). On the other hand, little is known about how audiences interact with online ads outside the lab, when browsing the internet freely (Ciceri et al., 2020; Dunaway et al., 2018; Wang & Hung, 2019).

One new technology which is likely to reach the broader market in the coming years is webcam eye tracking technology (Papoutsaki et al., 2016). This will enable researchers and practitioners to more accurately judge how individuals interact with the ads they encounter online, on news websites for example, in a natural environment, surfing on their own personal devices, rather than in a limited lab environment. This is the focus of study IV.

2.3 Distribution of online news content

For readers to take part of news content, it must be distributed to the relevant audiences. For printed media, physical distribution was performed by lorries and paperboys; online, algorithms do much of the heavy lifting (Diakopoulos, 2019).

Algorithms have an immense power over users in digital media and can influence consumption behaviors by recommending, curating, and presenting information in more pervasive ways than is possible offline (Z. Epstein et al., 2018; Thorson, 2008). On news websites, algorithms have the potential to steer news consumption by recommending and presenting content to visitors, for example through top lists based on overall popularity among a website's visitors (Prawesh & Padmanabhan, 2014; Thorson, 2008) or through personalization based on a user's previous behavior (Bodó, 2019; Semenov et al., 2022). But the influence of algorithms starts even before a user arrives at the news website. A large part of news websites traffic come from other sources such as social media platforms, news aggregators, and search engines (Flaxman et al., 2016). As the content users see on their screens is sorted and filtered in these other channels, algorithms can influence which users end up at certain news websites in the first place.

Researchers have shown how the ranking of results provided by search engines can have a major impact on what people click on (R. Epstein &

Robertson, 2015). While social media use has been found to be positively related to visiting a higher number of news websites (Scharkow et al., 2020), researchers have also found that the scope of news content users engage in *while still on the social media platform* tend to become narrower the more active the user is (Schmidt et al., 2017). Algorithmic ranking as well as the behavior of individuals plays a part in limiting the news content that individuals are exposed to on social media platforms (Bakshy et al., 2015). It has also been found that small twists to what and how political information is presented in social media can influence real-world voting behaviors (Bond et al., 2012). As Diakopoulos (2019) points out: encouraging people to go out and vote, for example, is not necessarily problematic if everyone is affected in a similar fashion – problems arise if groups of people are specifically targeted and influenced in different ways.

Just like there is a risk that editorial decisions can be biased (Sen & Yildirim, 2015), so can the decisions of algorithms. The number of reports where content distribution algorithms have been found to act in biased or skewed ways, typically to the disadvantage of minorities or based on gender (Datta et al., 2015; Lambrecht & Tucker, 2019) or race (Sweeney, 2013), keeps growing. Related to online news, it has been shown that there is a risk of source diversity bias in the results offered by news aggregators such as Google News, which tends to favor certain news outlets over others when presenting news to individuals (Haim et al., 2018).

Although issues of algorithmic bias (Ali et al., 2019) and accountability (Diakopoulos, 2015), as well as the broader field of machine behavior (Rahwan et al., 2019), has begun to attract researchers' attention lately, there is still much to learn when it comes to algorithms and online news. In study V, we investigate the risk of gender bias in the distribution of online news content by Google's and Facebook's respective distribution algorithms.

Chapter 3

Methodology

Although data are unlikely to replace theory (Mazzocchi, 2015), researchers have pointed to the potential of using larger datasets (Lazer et al., 2009) and inductive approaches (Yarkoni & Westfall, 2017) to further the social sciences. Ideally, and in line with the behavior science research tradition, the data should be longitudinal to enable analyzing variations over time and assessing the stability in behavioral patterns, and both measurement and analysis should be performed at the level of the individual in contrast to population averages (Schork, 2015; Thyer, 2009). I agree and, in short, the stance taken in this thesis is that observational data can be very useful in explaining and predicting online news behavior, both human and machine.

While the studies which constitute this thesis rely on different data and methods (see the respective research study for details regarding the methodologies employed) to answer their specific research questions, they are related as they cover different aspects of the same phenomenon. This chapter begins with a discussion on operationalization and measurement practice in the media use literature. This is followed by a section on how to study algorithms through the lens of ‘machine behavior.’ The chapter concludes by providing an overview of the empirical data used in this thesis and instructions for where to find data and materials necessary to replicate the main findings presented in the papers.

3.1 What is online news behavior?

Broadly defined, everything a person does can qualify as ‘behavior’; the only difference is that some behaviors are overt (visible to other people, such as

picking up their smartphone or saying things out loud), while others are covert (e.g., thoughts and feelings that are not visible to others) (Skinner, 1953/1965; Thyer, 2009). Accordingly, there are many different ways that online news behavior could be operationalized, measured, and analyzed, from granular to very specific (de Vreese & Neijens, 2016).

In research dealing with online behaviors, some of the most common operationalizations are what hyperlinks an individual clicks on to load content in their web browser (Choi et al., 2020) and if they look at a certain portion of the screen which contains some content of interest, as measured through eye tracking (Wang & Hung, 2019) or surveys (Goldfarb & Tucker, 2011). Online behavior data can also be combined or broken down in various ways to create measures of behaviors ranging a single or across multiple websites, and during one and the same or across different browsing episodes (Lalmas et al., 2015). Other common operationalizations of online behavior include volume (e.g., how many news articles are read in sequence and the duration of a browsing episode), engagement (e.g., number of times a person returns to an outlets' website or recording the number of pixels scrolled on the screen as a proxy for reading behavior) (Lalmas et al., 2015), and behavioral paths (how someone move from article to article or outlet to outlet during a browsing episode) (Esiyok et al., 2014; J. Möller et al., 2019; Vermeer et al., 2020).

Various operationalizations are employed in the different research studies in this thesis, depending on their respective research focus. Studies II and III, for example, rely on passively registered web logs containing information on the URLs which were loaded, and at what time, in the participant's browsers. One of the behaviors of interest in these studies was so-called *browsing sessions* – behavior sequences representing continuous episodes of news consumption. We followed industry standard and operationalized a news browsing session as a sequence of subsequent observations (i.e., loaded URLs in a participant's browser) from the same individual with a maximum idle time of

30 minutes between observations (Bucklin & Sismeiro, 2003; Catledge & Pitkow, 1995).⁹

Another example comes from study IV, which employed eye tracking for data collection. One of the behaviors of interest in study IV was whether participants looked at the advertisements they were exposed to on their screens while surfing for online news. This behavior, which we referred to as *ad fixation* in the study, was operationalized as an overlap of the user's gaze with the coordinates of the ad on the screen for at least 80% of gaze points within 100ms (the average sample contained five gaze points within 100ms, based on an eye tracker sampling frequency of 60Hz). For complete description of the behaviors measured and analysed in the research studies, I refer the reader to the full papers towards the end of the thesis.

It is worth pointing out that, although a worthy research topic, *what content a news consumer accesses* does not constitute a behavior by itself. For example, descriptive information about outlets (e.g., The New York Times or The Economist), sections (e.g., business or sports or entertainment), or articles (e.g., a catchy headline or news article written on a specific topic or by a certain journalist) are not records of behaviors. Information that an individual clicked on a hyperlink (leading the person to a certain outlet or news section), that the individual looked at something on their screen (a headline or image or a body of text), or what they thought when they consumed news (as reported through an interview or survey) are however all examples of behaviors.

Furthermore, while not deemed as behavior by itself, surrounding information can serve to enrich our understanding of the recorded behavior. For example, if we know when, where and how certain online news content was accessed, it might be easier to interpret the function of the observed behavior. Such surrounding and contextual information about how online news content is accessed include temporal information (day of week, time of day),

⁹ In fact, even Swedish authorities use this delimiter. If you log in at the Skatteverket (the Swedish Tax Agency) website but remain inactive for some time, you will receive the following message: "The session time has expired. It's been more than 30 minutes since you used this service. You have therefore lost any data that you have not saved." (author's translation).

frequency (e.g., how often), time spent (e.g., time spent on site or duration of browsing sessions), what device (e.g., mobile or computer) and platform (e.g., through the browser or an app) was used, and whether the individual accessed online news content directly or indirectly (e.g., via an intermediary such as social media or a news aggregator or going straight to the news publisher's website) (Bentley et al., 2019).

Zooming in further, a researcher could study the semantic structure of headlines or content and analyse how that influence behaviors like the tendency to click or not (Arapakis et al., 2014; Berger & Milkman, 2012), how individuals are influenced by the source of a news article such as author and type of media (Wahlund et al., 2021), or on a specific topic (not just the broader category of *politics* for example). Zooming out, contextual factors such as physical location (alone in your home by your kitchen table at breakfast) or certain ongoing events in the larger world (the Olympics might make you more likely to take part of news about sports) could also have an influence on a person's online news behavior.

Again, the aspects mentioned above are not behaviors in themselves but rather additional information about the context in which online news behavior occur. Three relevant concepts are useful in this discussion: *discriminative stimuli*, *establishing operations*, and *function of behavior*. A discriminative stimulus is defined as “a stimulus condition which, (1) given the momentary effectiveness of some particular type of reinforcement (2) increases the frequency of a particular type of response (3) because that stimulus condition has been correlated with an increase in the frequency with which that type of response has been followed by that type of reinforcement” (Michael, 1982, p. 149). Headlines, for example, can be viewed as discriminative stimuli which makes clicking a hyperlink more or less likely (W. M. Baum, 2017, p. 98; Michael, 1993) by signalling relevance of their attached news articles to the reader (Dor, 2003). That is why clickbaits work: a headline which promises juicy gossip can increase the probability that a person clicks on it, if the gossip has indeed been juicy in similar situations in the past when the person has encountered clickbait headlines.

An establishing operation is defined as “an environmental event, operation, or stimulus condition that affects an organism by momentarily altering

(a) the reinforcing effectiveness of other events and (b) the frequency of occurrence of that part of the organism's repertoire relevant to those events as consequences” (Michael, 1993, p. 191). An ongoing event in the outside world, such as the Olympics, could be viewed as an establishing operation for an individual if it makes the individual more likely to consume sports news by increasing its effectiveness as reinforcer (colleagues might talk more about sports than regular during lunch when the Olympics are raging, why an individual might want to consume sports news in order to be able to be part of these conversations) (W. M. Baum, 2017, p. 68; Michael, 1982). Similarly, it being Friday afternoon could be viewed as an establishing operation with the opposite effect; it might make the PhD student, who is tired from a long week of reading literature related to their thesis, less likely to engage in online news behaviors on the commute home (the time of the week decreases the effectiveness of consuming news as reinforcer).

The effects of establishing operations, however, can vary both between individuals and for the same individual over time (W. M. Baum, 2017, Chapter 4; N. Möller & Petrini, 2020, pp. 39–41). Building on our previous examples, knowledge that the Olympics are currently being held could make an individual *less* likely to engage in online news behavior, as they fear much of the news reporting will be dominated by this event and they have a history of not being particularly amused by sports news. And it being Friday afternoon could make the wilderness guide (who has been out kayaking in the archipelago for five days) *more* likely to seek out online news content as they have not engaged in such behaviors for several days. Older citizens typically engage more often and spend more time consuming online news compared to younger citizens (Ohlsson et al., 2020). It is thus possible that what would not constitute an establishing operation for a person to engage in online news behavior when they are in their early teens, could very well do so when they are in their thirties, or sixties.

It is not only establishing operations that can vary across individuals and time: the ‘reason’ or function of engaging in online news behaviors can too. The function of maintained behavior is either to increase contact with preferred outcomes (this is called positive reinforcement) or decrease contact with unpreferred ones (this is called negative reinforcement) (W. M. Baum, 2017, p. 68; Bouton, 2016, pp. 51–52; N. Möller & Petrini, 2020, pp. 78–79).

For one individual, the function of the behavior of habitually scrolling the frontpage of an online news outlet's website early in the morning while waiting at the bus stop could be to avoid the need to talk to a tiresome neighbour (negative reinforcement); for another, the function could be to escape from boredom (negative reinforcement) by adding visual stimulation (positive reinforcement). Another example could be an employee killing time before lunch by surfing for online news (negative reinforcement); whereas another employee reads up on current events to gain social recognition by having interesting things to say to their colleagues (positive reinforcement). The above reasoning follows the lens of behavior science, a perspective which is elaborated upon in research study I.

The fact that the function of online news behavior can change over time and vary between individuals, underlines what was mentioned already at the beginning of the chapter: that when studying online news behavior, it is often advantageous with longitudinal data that is also at the level of the individual.

3.2 How to measure online news behavior

A fundamental question for empirical research into news and media behaviors is, naturally, how to accurately measure media use behaviors. This is a notorious task, which has been extensively discussed by scholars (de Vreese & Neijens, 2016; Prior, 2009a, 2009b; Tewksbury, 2003).

A majority of the existing literature relies on self-report measures of one kind or the other (de Vreese & Neijens, 2016; Nelson, 2020; Taneja, 2016).¹⁰ Self-report measures on behaviors can be direct questions about said behaviors (such as, "How often do you do X on a typical day?" or "How many times did you do Y during the past week?"). Alternatively, the researcher can rely on indirect questions intended to gauge the behaviors through various theoretical constructs (e.g., stated intentions of future behaviors, consumer

¹⁰ For example, researchers at the Communication Science Department of the University of Amsterdam have curated a list of measures used in media exposure literature (available at <https://mediaexposuremeasures.org/>); when accessed 2022-02-28 there were 78 measures listed, 72 of those were self-reports.

needs, or interest in certain products or services). Self-reported data are often collected through interviews or surveys and offer many advantages. It is cheap, quick, and relatively simple. There are great examples of recurring reports on media use which offer valuable input to our understanding of aggregated news media use and general trends over time. Examples include the Reuters Institute Digital News Report (N. Newman et al., 2021) and Nordicom Media Barometer (Ohlsson et al., 2020), both of which are yearly reports covering various societal developments in media use.

However, for research concerned with *detailed information about individuals'* media use behaviors, previous studies have found that self-reported measures can be both biased and inaccurate (de Vreese & Neijens, 2016; Prior, 2009a). Measurement problems related to direct questions about media use behaviors are due to a number of reasons. These include deliberately misreporting or misunderstanding the question or scale (Lalmas et al., 2015) and experimenter demand effects (Iyengar, 2011; Zizzo, 2010), but mainly due to recall bias, i.e., because subjects have imperfect memory (Prior, 2009b; Vrijheid et al., 2009). One study on news exposure, for example, found both that individuals severely overestimated their own exposure to network news (on average by a factor of three) and that the bias in self-report measures varied a lot across demographics, which renders comparisons between groups very difficult (Prior, 2009a).

When it comes to indirect questions (i.e., measuring constructs in lieu of behaviors), it has been pointed out that many constructs in the social sciences, albeit theoretically elegant, cannot reliably predict future behaviors – even when they are claimed to do so – and fail to generalize outside the context of the study that conceived them (Yarkoni & Westfall, 2017). The replication crisis (Camerer et al., 2016; Open Science Collaboration, 2015) witness of the fragility of many of the existing ‘truths.’

The fact that reports such as the Digital News Report and Media Barometer can be repeated year after year allows researchers to accurately study developments over time on the aggregate; even though the particulars might be off for a single individual, reliable measures of relative changes for the whole study population can be developed and refined over time (Lalmas et al., 2015). On the other hand, when research is concerned with detailed behavior of individual news consumers, an alternative to self-report measures

is observations of actual behaviors, which have long been favored by behavioral scientists (W. M. Baum, 2017; Bouton, 2016; Sturme, 2008; Thyer, 2009). As a result of the technological advancements in the media industry, there are now reliable methods available for large-scale, automated observations of individual-level digital media behaviors (de Vreese & Neijens, 2016; Ingram Bogusz, 2018; Taneja, 2016).

Although observational methods like web analytics are great for collecting very specific and detailed measurements of digital (overt) behaviors, they are not without limitations. For example, automatic observations are often less in-depth than interviews or questionnaires, which allows the researcher to ask questions about anything, and many measurements are proxies of the things we really wish to learn about (e.g., an individual's interest in a certain news article has to be assumed based on the time they spend on the webpage or the number of times they return) (Lalmas et al., 2015).

Preprocessing is also often needed of the raw data to make it ready for analysis. Large amounts of time can be spent 'tidying' messy datasets (Wickham, 2014) and cleaning it from erroneous observations and noise (Srivastava et al., 2015). For example, if a user slips and clicks on a hyperlink they did not intend to, they might immediately click on the back-button in their browser and then on the correct hyperlink. This would render three observations rather than one: the user (1) going to the wrong page, (2) back to the startpage, and finally (3) to the right page (rather than only going straight to the right page). Strictly speaking, these three observations are not wrong, as they do represent clicking behavior of the participant, but the observations might be considered noise, depending on how the behavior of interest has been operationalized in a study (e.g., if a behavior is operationalized as "the user loads a webpage and stays for at least X seconds").

Another limitation with web log data is that it often contains only one timestamp for when the page was loaded, not another one for when it was closed (Lalmas et al., 2015). The time spent on a webpage must thus be inferred by looking at when the next observation for the same user occurred. However, oftentimes there can be long periods between observations, e.g., if a user is online in the afternoon and then again the next morning, there might be a gap of several hours between observations. In cases like this it is necessary to make assumptions about when a user can still be considered active

or not (Catledge & Pitkow, 1995). One common solution to this problem, that was applied in studies II and III of this thesis, is to group subsequent observations or clicks from one individual into behavioral sequences, or browsing sessions, based on an artificial delimiter such as a certain period without any observed activity (Bentley et al., 2019; Catledge & Pitkow, 1995). The time spent on the last page in each session can then either be dropped or replaced using mean-imputation using the data from earlier observations in the same session.

Notwithstanding these limitations, as the central question of this thesis concerns what detailed online news behavior look like at the level of the individual, observational methods are often preferred over the (possibly) more in-depth but less exact self-report methods (de Vreese & Neijens, 2016; Lalmas et al., 2015; Thyer, 2009). Unfortunately, as Mukerjee and colleagues (2018) have pointed out, the relevant technologies – to conduct, for example, passive registration of web surfing behaviors – and data gathered through such methods are often not directly available to researchers (only to the firms that developed the technologies or own the media platforms). In those cases where observational data have been made accessible for research, it is often aggregated across groups of people, over time (e.g., per month or for the full study period), across different news websites, or abstracted to the main domain (i.e., no specific URLs are available, only information that an individual have visited a certain publisher’s website) (Nelson, 2020; Nelson & Lei, 2018). When detailed, individual-level data are available, it is often confined to a single news publisher’s website (e.g., Wadbring & Bergström, 2021), rendering inferences about general media behaviors difficult (Mukerjee et al., 2018).

By collaborating with various media companies, it has been possible to collect detailed behavioral data for all the empirical studies in this thesis. For study II and III, we collaborated with a consortium of news publishers (see footnote 4 in the Introduction) and the market research company Nepa¹¹ to collect detailed web browser logs over time at the personal devices of about 35 thousand Swedes. To collect the data, participants were asked to install a

¹¹ www.nepa.com

browser plug-in which tracked all their news browsing activity through the course of the study period.

In study IV, we collaborated with the technology company Tobii¹² to collect longitudinal web browser logs for about 1,000 Swedes and managed to match these with continuous eye tracking data for every individual (collected via the participants own webcams). Again, participants were asked to install a browser plug-in which tracked their online news browsing behavior. This time, however, the browser plug-in was also connected to the participants webcams and recorded eye movements in tandem with surfing behaviors.

In study V, we collaborated with the media company Newstag¹³ to set up and conduct two single-case experiments on Facebook's and Google's respective ad distribution platforms to study what audiences the content distribution algorithms chose to deliver content to given various instructions they received from us (see section 3.3 below). For details about the data used, see Table 1 in section 3.4 (Data and materials) and the full papers, which are included in the subsequent section of the dissertation.

3.3 A note on machine behavior

The online news ecosystem is increasingly dependent on algorithms of various complexity to perform a range of functions (Sen, 2019); from sorting and recommending news pieces on news publishers homepages, to producing and distributing journalistic content across the web to audiences through search engines and social media (Diakopoulos, 2019). Even though algorithms are becoming more ubiquitous overall in our digital lives, they are often poorly understood (Rahwan et al., 2019) and have proven hard to study (Z. Epstein et al., 2018). There are several reasons for this, such as propriety of the underlying code as well as the growing complexity of algorithmic systems, which ironically can make them opaque even to their creators (Rahwan et al., 2019).

¹² www.tobii.com

¹³ www.newstag.com

Researchers can, just as with media use behaviors, measure the effects or implications of algorithmic systems indirectly by asking people what they think or believe about the algorithms. But when the research interest lies directly with the decision-making of an algorithmic system, another approach is preferable. By looking at an algorithm as you would an individual, and frame the input and output in terms of stimuli and ensuing behavior, it is possible to make inferences about the inner workings of the algorithm even though this is not directly observable (Z. Epstein et al., 2018). For example, as in study V, we can provide an algorithm with specific stimuli (e.g., an ad for a certain news video and instructions for how to distribute this online) and later observe the results as a form of ‘behavior’ exhibited by the algorithm. In the context of study V, the behavior is the number of people as well as the demographic constitution of the users that the algorithm chose to deliver the news video to.

Through repeated trials, with varying input offered to the algorithm, it is possible to map out which stimuli will yield certain machine behaviors in return. This is similar to experimental studies of the type referred to as single-case research designs that are oftentimes used in empirical behavior science research of humans and other animals (Kazdin, 2011; Thyer, 2009). Such study designs, with repeated measures of individuals’ behaviors over time, provide compelling arguments that a certain behavior really is a function of some other variable which has been altered. In fact, single-case experiments are considered to be the most rigorous and informative when it comes to understanding behavioral phenomena at the level of the individual in evidence-based practice (Schork, 2015; Thyer, 2009). A single demonstration of course offers weak evidence, but as the number of successful demonstrations accumulate (e.g., through replication studies or repeated measurement of individuals’ behaviors), claims can be made about functional relations at a more general level (Sturmey, 2008).

For a more extensive discussion on machine behavior, see study V in this thesis, which consists of two single-case experiments where Google’s and Facebook’s respective ad delivery algorithms act as study subjects.

3.4 Data and materials

The empirical papers of this thesis build on different data to support their conclusions. Table 1 provides a summary of the kind of data used in each of the research studies. Because of nondisclosure agreements, the original data used in this thesis cannot be posted publicly.

For study IV, however, synthetic datasets – where new values have been randomly generated for all variables, thus ensuring anonymity, while exactly mimicking the statistical properties of the original data – and replication code are available on the Open Science Framework, <https://osf.io/dgbse>.

Table 1. Overview of the empirical data used in the studies

	Study II	Study II & III	Study IV	Study V
Observations	276K	3,8M	2.6M	2.4M
Participants	1,000	35,000	1,000	2
Study duration	3 months	5 weeks	1 + 3 months	5 weeks
Data collection period	5 Nov 2017 – 1 Feb 2018	9 Sep – 14 Oct 2016	1 Aug – 1 Sep 2018, 1 Oct 2019 – 1 Jan 2020	18 Jan – 25 Feb 2019
Study setting	Sweden	Sweden	Sweden	US, UK, India, United Arab Emirates
Research design	Longitudinal observation/passive registration	Longitudinal observation/passive registration	Longitudinal observation/passive registration	Single-case research design (repeated experiments)
Type(s) of data	Web browser logs, surveys	Web browser logs, surveys	Web browser logs, eye tracking	Automated observational data
Data collection partner	Nepa	Nepa	Tobii	Newstag

Notes. In study V the focus was on machine rather than human behavior. Thus, while the data included information on millions of individuals that were exposed to our ads on the respective platform, individuals represent 'observations' in our study and the only 'participants' consisted of the ad delivery algorithms of Facebook and Google respectively.

3.5 Ethical considerations

There are ethical considerations to all research involving humans. A formal ethical review of the proposed research is not only important to undergo, but in many cases required under Swedish national law.¹⁴

There are, however, certain conditions which grant an exception to the need to undergo formal ethical review. Study I did not need ethical review, as it was conceptual and did not involve any human participants. Studies II, III and IV were exempt from the ethical review requirement as they were non-interventional (only observational data were collected) and did not involve any sensitive personal information about any of the study participants.¹⁵ Study V was experimental and aimed at influencing the behavior of the study subjects. However, as the study subjects were non-human (content delivery algorithms) ethical review was not required.

Important to note, however, is that while a formal ethical review was not required, all studies were conducted in accordance with the host university's principles for research ethics and informed consent was obtained from all human participants involved in studies II, III and IV. Furthermore, the integrity of the participants was protected in the sense that all data was pseudo-anonymized and we as researchers had no way of identifying individual participants: we did not have access to any personal information, only basic demographic information as well as randomized ID numbers used to connect multiple observations from the same individual.

¹⁴ For more information, see the Swedish Ethical Review Authority: www.etikprovning-myndigheten.se

¹⁵ Nonetheless, because of the comprehensive and detailed nature of the second dataset used in study II, we sought approval from the Swedish Ethical Review Authority prior to collecting the data. The ethical review was approved (application number 2017/747-31/5).

Chapter 4

Introducing studies I–V

This thesis consists of five original research studies: one conceptual book chapter and four empirical papers. Study I, the book chapter, provides a theoretical discussion from a behavior science perspective on why individuals engage with news content. Studies II and III look at individuals' online news consumption behaviors and makes use of extensive web browsing logs. Study IV discusses the business model of financing online news through ads. It relies on longitudinal eye tracking data combined with web browser logs. Study V examines the role of algorithms in the distribution of online news content by means of two experimental studies performed at Google's and Facebook's respective ad delivery platform.

In this chapter, each study is briefly presented and discussed. For an overview of how the studies relate to each other, see Figure 1 in Chapter 2. The full studies are included in the subsequent section of the dissertation.

Study I

Explaining the Behaviour of News Consumption

Single-authored book chapter. Published in *Managing digital transformation* (Åbonde Garke, 2018).

The book chapter is a conceptual work providing a theoretical perspective on what drives online news consumption and suggesting a strategy going forward for how news media can adapt their business model to better suit the digital world. The main point of the book chapter is an argument for news publishers to focus less on what they publish and instead consider what function consuming their content serves for their readers. This argument is grounded in behavioral science and exemplified with successful online news publishers.

In practice, this means that news publishers should think about why someone would read their content and then produce content which suits that need, rather than ‘just’ producing content they think is valuable and important and hoping that someone will read it. This is, perhaps, controversial and I suspect many news professionals will foresee an inevitable race to the bottom if they were to start producing only news that people want to read rather than news they ‘should’ read. However, I think this line of reasoning misses the point of the chapter. It is true that one of the lowest hanging fruits would be to publish clickbaits and gossip, which would attract readers by serving functions such as providing entertainment or escape from boredom. But as evident by news companies such as The Economist, producing content which serves other kinds of functions for readers (such as summarizing information and offer perspective on complex topics) is also possible. Regardless of what function a news company aims at serving, I believe that this strategy is likely to be a successful approach for gaining readers.

Study II

Understanding Online News Reading: Analyzing Individuals' Cross-Website Behavior Through Tracking Data

Single-authored paper. *Working paper*. An earlier version was presented at the *SIR Research Event* (February 2019, Stockholm).

News is increasingly consumed online by many citizens. Yet, while it is known that there are significant differences in how people engage with news, it is still unclear what characterizes individuals' online news reading behavior. Discussions regarding the distribution of consumption behaviors among online news audiences have been surprisingly absent in the literature. The purpose of this paper was to study individual- and session-level news reading behaviors across multiple news websites among a large sample of Swedish news readers.

The study was based on two large longitudinal datasets, collected passively from the personal devices of a total of approximately 36,000 individuals, consisting of detailed behavioral observations of the participants online news browsing histories across multiple news websites. Four research questions were addressed: (i) what the general online news reading patterns were, (ii) how widespread online news habits were, (iii) what kind of news outlets were typically visited and how varied people's online news reading was, (iv) what kind of differences and similarities, if any, in online news activity there were between heavy and light news consumers.

Findings showed that most people consumed rather little online news, both in terms of frequency and volume, but that there was a significant spread with a few highly active individuals (which inflated summary measures, such as the grand mean, of online news reading frequency etc.). There were several similarities among heavy and light news consumers. More than half of the participants, for example, exhibited regular news behaviors, and news sessions were typically brief and confined to a single news outlet, regardless of how active the news consumer was. The paper provides foundational facts about online news behaviors and have implications for both research and practice.

Study III

Identifying News Audience Clusters Based on E-Eigenbehaviours: Behavioural Pattern Recognition in Online News Browsing Data

Co-authored paper.¹⁶ Submitted to *Internet Research*.

There is a large spread in how individuals consume online news, making analysis and presentation of results difficult as the researcher often gets torn between looking at blunt aggregates (e.g., group-wise comparisons of the grand mean) and handling overwhelming idiosyncrasy (e.g., looking at the behavior of every individual separately). The purpose of this paper was twofold. First, to present a method for how researchers on digital behaviors can find such a middle ground between looking at either populations or individuals. Second, describe what online news consumption patterns look like over a typical week.

The study was based on a large dataset of web browser logs for approximately 35,000 Swedish residents over five weeks (same dataset as in study II). The methods we employed – a combination dimension reduction and clustering – have previously been applied in other settings, typically associated with physical behaviors (such as people moving about in a city), but not to digital behaviors. We were able to identify nine distinct news audience clusters, where users within each audience cluster exhibited similar behaviors while there were clear differences between clusters. The clusters differed by what news outlets they typically accessed, how much news content they consumed, and during what time of the day.

Overall, we found that most people consume very little news content online, with only a small group of highly active readers. As we expected there to be differences in news behaviors between weekdays and weekends, we performed separate analyses based on the type of day. Supporting this notion, we found that 46% of our users changed cluster affiliation between

¹⁶ Joint work with Professor Richard Wahlund, Ph.D., and Associate Professor Rickard Sandberg, Ph.D.

weekdays and weekends. Finally, while the behaviors varied between clusters, results indicated that there was a large overlap when it came to demographic and socio-economic characteristics as well as what device was used to access news (computer or mobile phone). The last findings suggest that media managers might be more successful in understanding their audiences by employing behavioural segmentation of their audiences, rather than dividing up their audiences by demographic subgroups, as is often the case.

Study IV

Always-on Eye Tracking Reveals Problems with Financing News Websites Through Advertising

Co-authored paper.¹⁷ Preprint available at SSRN (Åbonde Garke et al., 2021): <http://doi.org/10.2139/ssrn.3949327>

News media have long relied on ads to subsidize journalism. As news consumption shifts from print to online, the reliance on ads for financing have become even greater as subscription numbers (in general) continues to decline. Meanwhile, online news websites have struggled with their online ad revenue streams, as much of the global online ad revenues flow to other digital platforms such as Google and Facebook. The purpose of this paper was to study how online audiences engage with ads across different websites to evaluate whether ads can be a viable source of financing for online news in the future.

Relying on two longitudinal datasets of web browser logs augmented with continuous eye tracking data, in total comprising about 1,000 individuals, we analyzed how individuals interacted with online ads at different websites while browsing the web freely in a natural context on their personal computers. To account for individual differences in how people engage with ads across different websites, multilevel logistic regression models were estimated with both random intercepts and slopes for users.

¹⁷ Joint work with Assistant Professor Erik Modig, Ph.D., and Martin Söndergaard.

Findings indicate that, *ceteris paribus*, the future for ad financed online news might be grim, as news websites underperforms compared to other online platforms in several respects when it comes to online ads. First, a lower share of the potential ads that were loaded on websites become visible on users' screens for news compared to other websites. Second, of the ads that did become visible, users were less likely to fixate their gaze on ads encountered on news websites compared to other online platforms. Third, of the ads that users did fixate on, there were no differences in the level of attention paid to the ads (as measured by total fixation duration and number of fixations) indicating that the price premiums on ad space currently enjoyed by news websites compared to social media might not be warranted.

Study V

Can Two Wrongs Make a Right? An Empirical Study of Algorithmic Bias in the Delivery of Ads on Facebook and Google

Co-authored paper.¹⁸ Submitted to *Journal of Computer-Mediated Communications*. Presented at *ABAI 10th International Conference* (September 2019, Stockholm, Sweden).

News media, as most online businesses, are increasingly reliant on distribution algorithms to push their content to users' screens as well as attract audiences back to their website. But while algorithms are becoming ever more ubiquitous, an increasing number of studies have pointed to dangers of outsourcing decision-making to algorithms, which have often grown into highly complex and opaque AI systems over time. Such algorithms have, for instance, been found to be biased against various minorities, e.g., by serving different job ads or search results to subpopulations based on gender or race. However, little is known with regards to how content distribution algorithms behave in relation to news content.

¹⁸ Joint work with Professor Richard Wahlund, Ph.D., Camilla Dahlin-Andersson, Ph.D., and Lise Bergman Nordgren, Ph.D.

In two single-case field experiments, one on Facebook and one on Google, we asked the respective platforms' algorithms to distribute various news videos to a mix of both men and women and recorded the demographic constitution of the audiences that the algorithms decided to distribute our content to. To account for potential cultural differences in how the algorithms behaved, both experiments were conducted across four different countries (US, UK, India, and United Arab Emirates). We compared the observed demographic constitutions with the expected demographics for each country based on usage statistics for Facebook and Google, respectively.

Results showed that while certain news videos were strongly targeted towards one gender by the algorithms, there were also videos which were strongly targeted towards the other gender. Surprisingly, we found that the overall distribution across all ads within a country followed the expected gender distribution almost perfectly. In conclusion, it seems as if two wrongs (i.e., biased decisions by the algorithm) might make a right by cancelling each other out. However, this also raises the question of whether there are systematic differences in what content is biased towards men and women. If different subpopulations are exposed to vastly different kinds of news, it could potentially be detrimental to democracy, and is an important topic for future research.

Chapter 5

Discussion

In this chapter, I summarize the main findings, contributions, and implications of the thesis. In the following chapters, I discuss limitations and future research directions before concluding the thesis.

5.1 Brief summary of the main findings

An example of the thinking in the news industry is that “[t]hough it may be interesting or even entertaining, the foremost value of news is as a utility to empower the informed” (Dean, 2013). In contrast, I argue in study I, individuals may consume news for a variety of reasons, and we should analyze which function the behavior of consuming news has for an individual to understand why news behaviors are changing.

Most people consume little news online both in terms of frequency and volume. There is, however, a significant spread with a few highly active individuals in the upper end of the spectrum, as shown in studies II and III. Heavy news consumers are likely to read more news stories and visit a higher number of news outlets in total compared to light users. Interestingly, most news sessions, regardless of how active the news consumer is, are brief and confined to a single news outlet.

Many individuals, more than half of the study population in study II, show habitual tendencies in their online news consumption behaviors. These tendencies are independent of the frequency and volume of their news consumption (i.e., both heavy and light news consumers tend to have news habits).

Online news behaviors are not necessarily dependent on demographic variables, or the device used to access news. Consequently, it might be more fruitful to segment news audiences into clusters based on behavioral data as shown in study III. This kind of analysis is also facilitated by the habitual behaviors that many individuals exhibit.

News websites are currently less effective in delivering ads to audiences, and their users are less prone to engage with the ads they are exposed to, than in search engines and on social media, which was found in study IV. As measurement practices for online advertising improve this risks exacerbating the problem further.

In study V it was found that the content delivery algorithms of Google and Facebook deliver ads for news articles in a balanced way on a general level with regards to gender, mirroring the expected gender distributions in different countries. Individual news stories, however, risk getting skewed exposure towards certain subgroups of a population when distributed by algorithms online.

5.2 Contributions to the literature

As discussed in Chapter 3, there are several problems related to the use of self-report measures when we are really interested in the detailed study of behaviors. These problems include, among other things, experimenter demand effects (Iyengar, 2011; Zizzo, 2010) and recall bias (Prior, 2009b; Vrijheid et al., 2009). When behavioral data on media use have been available, it has often been restricted to artificial lab settings (Dunaway et al., 2018) or aggregated over time or groups of people (Mukerjee et al., 2018). This thesis includes four empirical studies into online news behavior which all rely on detailed behavioral data. Furthermore, the datasets were all collected in field settings, with few restrictions put on the participants behaviors, and over extended periods of time. There are, off course, limitations related to each of these studies, such as the range of news outlets were covered (studies II and III) and the data not covering news behaviors on mobile devices (as in study IV). However, given the sheer size of the datasets used, their longitudinal nature, and that they were collected in a natural environment, I would still argue that this thesis makes an important contribution to the literature

by providing a robust baseline for what online news behaviors really look like (see section 5.1 Brief summary of the main findings above).

The statistically interested reader will notice that this thesis includes rather few of the standard null hypothesis significance tests. This might come as a surprise, given the quantitative nature of the thesis and the fact that I have had very large datasets available to test a myriad of ideas of mine, had I wanted to. But while large samples provide great opportunities, they also create problems related to interpreting statistical significance, such as the tendency to recognize false results as positive ones (Ioannidis, 2005; Lin et al., 2013). The “mindless” application of statistical rituals is likely part of the problem (Gigerenzer, 2004). This is nothing new – already 40 years ago, Meehl (1978, p. 806) charged that “most so-called ‘theories’ in the soft areas of psychology [...] are scientifically unimpressive and technologically worthless.” Meehl’s argument might have been primarily based on philosophical argument rather than empirical evidence, but the replication crisis (Camerer et al., 2016; Open Science Collaboration, 2015) indicates that he was on to something.

One reason for the current state of affairs can be traced to a fundamental misunderstanding of the use of power in statistical tests in the social sciences: if you only specify a directional hypothesis, rather than an exact numerical or range prediction as commonplace in the natural sciences, larger samples and more exact measurement methods will only make it easier to find statistically significant results (Meehl, 1967). Correspondingly, in the studies in this thesis it would have been simple to find many statistically significant results only due to the sheer mass of observations (in the millions) (Lin et al., 2013). One suggestion to deal with the problem of false research findings is by lowering the p -value threshold required to claim new findings from .05 to .005 (Benjamin et al., 2018). Another solution is off course to follow Gigerenzer’s (2004) call to use statistical thinking, with methods and tests appropriate to the task at hand. One important contribution of this thesis is to show that it is possible to produce rigorous quantitative research without relying solely on null hypothesis significance testing.

Closely related to the issue of replicability, power, and significance testing is the tendency in previous research into online news behaviors to analyze and report summary measures, such as comparing means between groups,

even though news behaviors are known to be skewed at the group-level (Bentley et al., 2019; J. Möller et al., 2019; Vermeer et al., 2020). Relying on group averages incurs a risk of reporting inflated effect sizes since much of online news consumption follow long-tailed distributions, as shown and discussed in Study II of this thesis. It is also important to aid statistical thinking and support any claims through other analytical tools, such as visualizations (Cleveland & McGill, 1985; Healy, 2019; Tufte, 1997, Chapter 2), as was done in all studies II–V. This thesis contributes to the literature by stressing the importance for researchers to consider the full range of the data and complement their statistical analysis with other tools such as visualizations of the underlying data when analyzing online news behaviors.

Any single type of data on media use is necessarily limited, and joining different data sources is a powerful way to gain new insights (de Vreese & Neijens, 2016). In study IV, we showed that combining eye-tracking technology with web browser logs can yield valuable new knowledge about online news behaviors, and even more so when employed on participants' personal devices and tracked over extended periods of time. Specifically, we showed that what people click on is quite different from what people look at, and that is quite different from what is shown on the screen. The findings from study IV also contributes to furthering our understanding on viewability and exposure towards online advertising (Exposito-Ventura et al., 2020).

Borrowing methods from other disciplines can be fruitful. In study III, we presented and tested a method for performing behavioral segmentation and analyses on online news data by employing a combination of dimension reduction and unsupervised clustering. This methodological approach was originally developed in computer science (Eagle & Pentland, 2009) and has been successfully applied to behavioral data of physical movements (Calabrese et al., 2010; Jiang et al., 2012). The method, however, is particularly apt at dealing with rich and large online behavior data, and to my knowledge we are the first to show how this methodology can be adapted for online behaviors.

There are two major benefits with the approach employed in study III. First, using eigendecomposition to reduce the dimensionality of the data allows us to find a middle ground between looking at blunt population or

group averages and the overwhelming idiosyncrasy that exists among the behavior of individuals (Eagle & Pentland, 2009). In practice, this allows managers to identify distinct groups of readers which exhibit similar behavioral patterns over time and tailor products or services to these audience clusters. Second, the method is capable of handling very large volumes of data. We employed the method to a big dataset containing online news behaviors from more than 30,000 individuals over five weeks. As availability of data in research dealing with online behaviors soars (Lin et al., 2013) this method, which was originally developed as a method for dimension reduction, will come even more into its own right. The method would, for example, be capable of handling data covering longer timespans, a larger number of news outlets, or greater detail in the behaviors operationalized and measured.

In study V, we showed that single-case experimental designs, which are common within behavior science but rare in other fields, can successfully be applied to study machines by designing an experimental study to evaluate the behavior of algorithms, in this case Facebook's and Google's content delivery systems. In standard statistical applications and experiments, data typically entail (relatively) little information on many individuals (e.g., gathered through a cross-sectional study), samples are randomly drawn from some population (suggesting external validity of potential findings), and the researcher relies on inferential statistics to assess differences between or within groups (Perone & Hursh, 2013; Sturmey, 2008; Thyer, 2009). A single-case experiment, in contrast, works over time and at the level of the individual, by introducing the various treatments to the same individual in succession and record if and how this has any effect (Kazdin, 2011; Perone & Hursh, 2013). The single-case research design is a very potent scientific method, capable of identifying functional relationships at the level of the individual (Schork, 2015; Thyer, 2009).

In study V, the single-case research design framework allowed us to audit the 'behavior' of Facebook's and Google's content distribution algorithms. Our findings nuanced recent studies on machine behavior, arguing that algorithmic decisions can become biased or discriminatory as a side-effect of optimizing for ad effectiveness (Ali et al., 2019; Lambrecht & Tucker, 2019), by showing that in the case of online news it seems as if the algorithms under study were not biased with regards to gender on a general level. Furthermore,

one of the most interesting findings in study V was that the two algorithms disagreed on many counts of what audience was ‘suitable’ to receive certain news articles. For example, while one platform presented a certain news article to mostly women, the other presented it mostly to men. Findings like this underscore the need for further investigations into machine behavior, as we cannot *a priori* know how an AI system will respond to our requests, or how such systems’ decision-making will differ from each other.

An early version of the methodology applied in study V was presented at a poster session at the *N=1 Stockholm Symposium* conference arranged by Karolinska Institutet in October 2018, which resulted in positive discussions with other conference attendees. The completed study was later presented at the *ABAI 10th International Conference* in September 2019. On both occasions, conference attendees seemed genuinely surprised to see their thinking and methods applied to subject matters outside their own disciplinary silo. I believe that behavior science, as discussed in study I, could provide a theoretical explanation for why news consumers migrate from traditional news media to the internet and social media. Applying the lens of behavior science could also point one way forward for how to retain (or grow) news audiences. With this thesis, I contribute towards the call for researchers to extend and spread applied behavioral science beyond the typical realms of autism and developmental disabilities (Normand & Kohn, 2013) by showing how single-case research designs can be applied (1) outside behavior science, and (2) to study machines.

5.3 Implications for practice

One goal of this thesis was to contribute towards the industry’s need to “find a path toward a more sustainable journalism” (Öhrvall, 2015). So, what is the way forward?

One of the major, and in my view most surprising, findings in this thesis is just how little online news content most people consume. While more news outlets are putting their online content behind paywalls (N. Newman et al., 2021), there are also indications that as more people start paying for other forms of online media their willingness to pay for online news could also increase (Fletcher & Nielsen, 2020). But people who read very little news

content online are, most likely, not going to be interested in paying for a subscription as they are happy enough with reading headlines and/or the free articles they can access without having to pay-to-read. The small group that does consume more news typically do so from a range of different news outlets, even though they get the bulk of their news from their favourite news outlet. This group may also not be interested in subscribing, as a subscription to a single newspaper might be too limiting in terms of content and multiple subscriptions too expensive. One can envisage a future, though, where there is a group of readers who would be interested in a bundled subscription like we see in other media categories such as music (e.g., Spotify) or movies (e.g., Netflix); pay a single monthly fee and gain access to a large and varied body of content from different producers. There are also examples of the opposite: companies like Substack¹⁹ offers extreme ‘unbundling’ by allowing individuals to subscribe to and support the work of individual journalists and writers whose work and style they like. Through such services, individuals can sign up for many micro subscriptions on topics from writers they are really interested in, instead of paying for a whole newspaper with broader but (perhaps) less relevant content.

Although individual citizens typically read little news, there was a significant and steady flow of news consumption activity for most news outlets throughout the day at an aggregated level (study III). There is clearly a large online news audience in total, although the attention from most individuals might be limited. This supports previous research which have argued that journalism still has the potential to influence society at large (King et al., 2017), and suggests that advertising could be a viable way forward as a source of financing, as ad revenue is primarily based on the number of times ads are served to users (Ferguson, 1983; Hu et al., 2015). Unfortunately, in study IV, we found that news readers interact less with ads on news websites compared to social media which suggests that news media need to reevaluate their current advertising practices.

One factor urging news media to speed up this reevaluation process is developments within webcamera technology. Webcameras on computers

¹⁹ www.substack.com

and front-facing cameras on mobile devices are getting ever more powerful, and modern smartphones and tablets can even use facial recognition to allow its owner to do various things. Apple, for example, lets users "unlock [their] iPhone or iPad, authenticate purchases, sign in to apps, and more — all with just a glance" using their Face ID technology (Apple Inc., 2021). Enabling eye tracking on such a camera only requires integrating a simple JavaScript library on any webpage, such as the open source project WebGazer.js developed at Brown University (Papoutsaki et al., 2016). If eye tracking via webcams become widely used on personal computers and smartphones in the future, the measurement of online ad effectiveness will improve. This could have severe consequences for the financing of news websites, as advertisers might choose to allocate a larger part of their marketing budget on other more effective platforms.

A reason why social media is more effective in serving advertising to its users than news outlets could be that the function of interacting with ads and the function of interacting with content is more similar on social media (e.g., purpose of both is killing time) than on a news website (e.g., purpose of a news visit could be to read up on current events, which news content allows the user to do but ads gets in the way of), as discussed in study V. The most common type of targeting in online advertising today is to serve personalized ads to users based on their browsing history, but this has been criticized as it invades the privacy and integrity of users (Evans, 2009). Another targeting method is called contextualized advertising, where ads are matched with the digital context rather than behavioral history of the user (Goldfarb, 2014).

Previous research have found that placing advertisements in thematically incongruent print media could increase its effectiveness by making readers process the ads more (Dahlén et al., 2008). Research into online advertising, on the other hand, have found that users avoid looking at ads that are made distinctive relative to the rest of the website, a phenomenon referred to as "banner blindness" (Benway, 1998; Cho & Cheon, 2004), and that ads which resemble the editorial content are better remembered (Hervet et al., 2011). This speaks in favor of employing contextualized advertising over personalized advertising at online news websites.

Contextual advertising can be compared with native advertising, i.e., sponsored news articles. Native advertising closes the gap between what is

perceived by the news reader as ads and news articles, by designing ads that are very similar in form to the editorial content (Aribarg & Schwartz, 2020; Conill, 2016) and is generally seen as an effective form of advertising from a practitioner point of view (Harms et al., 2017). Native advertising, however, raises questions about ethics (Carlson, 2015) as well as the credibility of news outlets (Aribarg & Schwartz, 2020), as readers often have a hard time distinguishing sponsored content (i.e., native ads) from actual news content (Wojdowski, 2016). One way forward to try and increase the effectiveness and profitability of advertising on online news websites could be to focus more on contextual advertising instead of personalized advertising, in the hope that it will reduce the gap between consuming news content and ads, but without entering the grey area that is native advertising.

Closely related to the question of how ads should be served on news websites is the question of how news websites should attract audiences, as discussed in study V. When online news publishers rely on algorithms to attract audiences, it is important to follow up online campaigns and study what audiences the content delivery algorithms decide to distribute various news stories to. Because, while algorithms might not deliver content in biased or skewed ways on a general level (e.g., it does not underserve certain content to various demographics), there could be systematic discrepancies on a topic-level. Even if, say, women and men visit a certain news publisher to an equal amount, there could be detrimental effects to the democratic process if men were primarily exposed to news within one category (e.g., opinion pieces), while women to another (e.g., business articles) as asymmetries in knowledge and information could arise over time.

Ad delivery algorithms only do what they are told; try to maximize some target as set by the advertiser. This creates tension, though, as the algorithms by optimizing for some goal can lead to unintended consequences, such as gender bias (Ali et al., 2019; Datta et al., 2015). It is an inherent characteristic of these platforms that advertisers cannot explicitly tell an ad delivery algorithm to only target one gender, as this would allow advertisers to intentionally discriminate towards groups of people (Lambrecht & Tucker, 2019). However, previous research has found that content delivery algorithms rely heavily on automated image classification – more than what text, hyperlinks or headlines are used – when deciding what audiences to distribute ads to

(Ali et al., 2019). This provides one potential work-around for media managers, should they detect that their ads start reaching a skewed audience in unintended or discriminatory ways: by experimenting with different images on the ads it is possible to influence to what users the ad will be delivered to.

Chapter 6

Limitations

There are limitations to this thesis. Studies II, III and IV focused on Swedish news consumers. Sweden has a high level of internet penetration (The Economist Intelligence Unit, 2021) and ranks high in terms of democracy (The Economist Intelligence Unit, 2020) compared to global averages. Sweden also has a quite unique news media environment, where public service broadcasters enjoy a specially strong and respected position in the media market (N. Newman et al., 2021). These factors could potentially lead to different online news behaviors in Sweden compared to elsewhere, why it is important to acknowledge that it is not known how the findings presented in this thesis applies outside a Swedish context. Findings from study V, which focused on machine behavior, should be more widely applicable, as the study was conducted across four countries with diverse cultural and news landscapes (US, UK, India, United Arab Emirates).

While the samples in studies II, III and IV were substantial in size (with thousands of participants) and matched the Swedish population to a reasonable degree with respect to age and gender, the respondents were not randomly drawn from the general population. Rather, the participants were recruited from existing market research panels to which they have self-selected to participate. As such, there is a risk that the behaviors observed in the presented research might not be representative for the Swedish population. The question of size and representativeness of the sample is not relevant to study V, which followed a single-case research design (the study subjects were the ad delivery algorithms of Google and Facebook themselves).

Furthermore, the data which studies II, III and IV rely on were collected on one device per participant: either a mobile device (smartphone or tablet) or computer, whichever device they used when they signed up to join the market research panel. However, in many countries it is now common to use two or more devices to access news on a regular basis (N. Newman et al., 2021). Thus, while the results presented here are likely to have captured a significant part of the news behaviors of the participants, we do not necessarily know what their full media diets look like. This is not a relevant parameter in the case of study V, as the algorithms themselves are not found on any specific device but are an integrated part of Google's and Facebook's wider ecosystems.

Cultural practices continue to evolve over time (often driven, at least in part, by technological developments). As such, the findings presented in this thesis are not universal laws, but rather indicative of the current media landscape as behaviors will continue to be shaped by the environment. A few big technological leaps have occurred in a relatively short period of time which have all had revolutionary impact on both the news industry and our media behaviors: the internet, mobile devices (first laptops, later tablets and smartphones), and social media. New innovative technologies are almost certain to emerge in the near future as well, but what these are and how they will play out we will have to wait and see.

Chapter 7

Future directions

Studies II, III and IV are based on passively registered behavioral data, i.e., not experiments where I try to influence or modify behaviors. While the current work provides a robust baseline for what online news behavior looks like, a natural next step would be to set up experimental studies to better understand the functional relations outlined on an overarching conceptual level in study I. There are some experimental studies on related topics, such as how variations in informational messages on social media can change user behaviors (Bond et al., 2012) or how the publication of articles on certain topics can influence what people discuss online (King et al., 2017). To the best of my knowledge, however, there are no studies that apply the rigorous methods from behavioral science (e.g., single-case research designs) to the topic of online news behavior.

Although comprehensive, most of the data in this thesis relies on web browser behaviors from mobile devices and computers. On mobile devices users can also choose to engage with news content directly through news apps (Stroud et al., 2020). There have been reports that mobile app usage is complementary to visiting mobile news websites (Xu et al., 2014) and that mobile news consumers behave differently depending on whether they use an app or the browser, typically with app users engaging more deeply with news content than browser users (Dunaway et al., 2018). Apps have the possibility to prompt users to engage in online news behaviors through push notifications (Stroud et al., 2020), a form of discriminative stimuli not avail-

able to news websites accessed through a browser. Previous research investigating news access through apps or browsers is, however, mostly limited to surveys or narrowly confined lab settings. Further field research relying on observational data comparing web browser with app behaviors on mobile devices is warranted.

Eye tracking technology adds a valuable dimension to other kinds of data on media behaviors (Ciceri et al., 2020). Eye tracking studies in field settings are rare in the literature though, and the data employed in study IV is consequently quite unique. Just as media use has rapidly moved from offline to online, there is another move underway where mobile is to a large extent becoming the most prominent platform (Dunaway et al., 2018). More field studies relying on eye tracking would be valuable overall, and eye tracking research on mobile devices, in particular, is likely to yield fascinating results.

Many news companies are losing subscribers and struggling for profitability (Sen, 2019). In the digital world, advertising might be one solution for news companies going forward. Unfortunately, as shown in study IV, news websites are currently both worse at making ads visible on their visitors' screens and visitors are less likely to engage with the ads that do show up compared to social media and search engine websites. The ineffectiveness of traditional ad banners (Benway, 1998; Ciceri et al., 2020), integrity issues related to personalized ads (Evans, 2009) and ethical issues related to native advertising (Carlson, 2015) points the way towards another type of advertising: contextual ads. Future research should explore how webpage structure and ad placements on news websites could improve, so that a larger proportion of ads are shown to visitors. Future studies should also investigate whether a shift from personalized to contextualized ads could help news websites improve their ad effectiveness, and hence ad revenues and long-term financing, by lowering the discrepancy between interacting with ads and content.

The nascent field of machine behavior has only recently gained attention in mainstream research and there are many white spots left to be studied (Rahwan et al., 2019). There are a growing number of reported findings where algorithmic behavior have resulted in biased, unfair or even discriminatory outcomes (Buolamwini & Gebu, 2018; Sweeney, 2013). Apart from

study V in this thesis, which focused on issues related to attracting news audiences to news websites, most studies on machine behavior so far have focused on problems with the results pages of search engines (e.g., Datta et al., 2015; R. Epstein & Robertson, 2015) or with advertising on social media (e.g., Ali et al., 2019; Lambrecht & Tucker, 2019). As the news industry employs more algorithms in their daily operations (Diakopoulos, 2019) it is important that the call for researchers with knowledge in experimental methods and hypothesis testing to engage in the study of machine behavior should be answered (Z. Epstein et al., 2018).

Finally, the definition of the term *news* that has been employed throughout this dissertation (see chapter 1. Introduction) is broad and inclusive, but also rather vague. As shown in Wahlund et al. (2021) there is no clear consensus in the literature for how to define the terms of *news* and *journalism*. There are many aspects to consider – such as the company producing the news content, who is writing the news, the brand of the outlet where the news pieces are being published, the device the consumer use to access news (for example smartphone or computer), the platform (e.g., news aggregator or social media), or time and place where news is consumed – which could all influence what the consumer views as ‘news content’. It seems odd that no clear definition exists in the literature and remains an important task going forward.

Chapter 8

Conclusion

This thesis opened with the question what online news behavior look like and how we can study it. In the first research study I offered a conceptual view on online news behavior and why we see the change in behaviors that we do; in the four empirical papers that followed, which employed diverse research methods, different aspects of online news behavior were studied. I currently believe that journalism will continue to have an important societal role to fill in the future, although it is unlikely that online newspapers will be able to rely on subscriptions to survive. Relying on advertising, however, might be.

When it comes to what online news behavior look like, I found that on an aggregated level there is a steady flow of news activity throughout the day, while on the individual level news consumption is a sparse activity for most people. Still, about two thirds of all individuals showed habitual online news behaviors. We further found that there is a large difference between the ad performance on news websites compared to social media platforms, highlighting the urgent need for news publishers to oversee their current reliance on ads for financing; either by improving the effectiveness of advertising in news media, or by turning to other sources of revenue. There has been alarms about biased algorithms discriminating various groups of people online. We did not find obvious flaws in how the ‘digital paperboys’ distribute online news content, but it is important to continuously monitor the algorithms as their behavior continue to evolve and change over time.

Regarding how online news behaviors can be studied, I showed that borrowing methods from other fields such as computer science and behavior science can yield valuable and interesting insights when analysing complex behavioral patterns in large samples. Combining different data sources, like web browsing logs with eye tracking, also add substantial explanatory value when studying online news behaviors. Furthermore, as elaborated upon in study I, I believe that the changes in how people consume news can largely be explained through the lens offered by behavior science. This would be an interesting venue to continue exploring in future research.

References

- Åbonde Garke, A. (2018). Explaining the behaviour of news consumption. In P. Andersson, S. Movin, M. Mähring, R. Teigland, & K. Wennberg (Eds.), *Managing digital transformation* (pp. 203–215). SSE Institute for Research, Stockholm School of Economics.
- Åbonde Garke, A., Söndergaard, M., & Modig, E. (2021). Always-On Eye Tracking Reveals Problems with Financing News Websites Through Advertising. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3949327>
- Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., & Rieke, A. (2019). Discrimination through optimization: How Facebook’s ad delivery can lead to skewed outcomes. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–30. <https://doi.org/10.1145/3359301>
- Anand, B. N. (2016). *The content trap: A strategist’s guide to digital change* (First Edition). Random House.
- Andersson, J., Bäck, J., & Ernbrandt, T. (2020). *Svenskarna och internet 2020*. Internetstiftelsen. <https://www.svenskarnaochinternet.se>
- Anscombe, F. J. (1973). Graphs in Statistical Analysis. *The American Statistician*, 27(1), 17–21. <https://doi.org/10.1080/00031305.1973.10478966>
- Apple Inc. (2021, January 22). *Use Face ID on your iPhone or iPad Pro*. Apple Support. <https://support.apple.com/en-us/HT208109>
- Arapakis, I., Lalmas, M., Cambazoglu, B. B., Marcos, M.-C., & Jose, J. M. (2014). User engagement in online News: Under the scope of sentiment, interest, affect, and gaze: User Engagement in Online News: Under the Scope of Sentiment, Interest, Affect, and Gaze. *Journal of the Association for Information Science and Technology*, 65(10), 1988–2005. <https://doi.org/10.1002/asi.23096>
- Arguedas, A. R., Robertson, C. T., Fletcher, R., & Nielsen, R. K. (2022). *Echo Chambers, Filter Bubbles, and Polarisation: A Literature Review*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-01/Echo_Chambers_Filter_Bubbles_and_Polarisation_A_Literature_Review.pdf
- Aribarg, A., & Schwartz, E. M. (2020). Native Advertising in Online News: Trade-Offs Among Clicks, Brand Recognition, and Website Trustworthiness. *Journal of Marketing Research*, 57(1), 20–34. <https://doi.org/10.1177/0022243719879711>

- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132. <https://doi.org/10.1126/science.aaa1160>
- Baum, M. A. (2002). Sex, Lies, and War: How Soft News Brings Foreign Policy to the Inattentive Public. *American Political Science Review*, *96*(1), 91–109. <https://doi.org/10.1017/S0003055402004252>
- Baum, W. M. (2017). *Understanding behaviorism: Behavior, culture, and evolution* (Third Edition). Wiley.
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., ... Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*(1), 6–10. <https://doi.org/10.1038/s41562-017-0189-z>
- Bentley, F., Quehl, K., Wirfs-Brock, J., & Bica, M. (2019). Understanding online news behaviors. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–11. <https://doi.org/10.1145/3290605.3300820>
- Benway, J. P. (1998). Banner Blindness: The Irony of Attention Grabbing on the World Wide Web. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(5), 463–467. <https://doi.org/10.1177/154193129804200504>
- Berger, J., & Milkman, K. L. (2012). What Makes Online Content Viral? *Journal of Marketing Research*, *49*(2), 192–205. <https://doi.org/10.1509/jmr.10.0353>
- Bodó, B. (2019). Selling News to Audiences – A Qualitative Inquiry into the Emerging Logics of Algorithmic News Personalization in European Quality News Media. *Digital Journalism*, *7*(8), 1054–1075. <https://doi.org/10.1080/21670811.2019.1624185>
- Bogart, L. (1981). *Press and public: Who reads what, when, where, and why in American newspapers*. L. Erlbaum Associates.
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*(7415), 295–298. <https://doi.org/10.1038/nature11421>
- Boulianne, S. (2020). Twenty Years of Digital Media Effects on Civic and Political Participation. *Communication Research*, *47*(7), 947–966. <https://doi.org/10.1177/0093650218808186>
- Bouton, M. E. (2016). *Learning and behavior: A contemporary synthesis* (Second edition). Sinauer Associates, Inc., Publishers.
- Bucklin, R. E., & Sismeiro, C. (2003). A Model of Web Site Browsing Behavior Estimated on Clickstream Data. *Journal of Marketing Research*, *40*(3), 249–267. JSTOR.
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Conference on Fairness, Accountability and Transparency*, 77–91. <http://proceedings.mlr.press/v81/buolamwini18a.html>

- Calabrese, F., Reades, J., & Ratti, C. (2010). Eigenplaces: Segmenting space through digital signatures. *IEEE Pervasive Computing*, 9(1), 78–84. <https://doi.org/10.1109/MPRV.2009.62>
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmeld, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433–1436. <https://doi.org/10.1126/science.aaf0918>
- Carlson, M. (2015). When news sites go native: Redefining the advertising–editorial divide in response to native advertising. *Journalism*, 16(7), 849–865. <https://doi.org/10.1177/1464884914545441>
- Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the World-Wide web. *Computer Networks and ISDN Systems*, 27(6), 1065–1073. [https://doi.org/10.1016/0169-7552\(95\)00043-7](https://doi.org/10.1016/0169-7552(95)00043-7)
- Cho, C.-H., & Cheon, H. J. (2004). Why Do People Avoid Advertising on the Internet? *Journal of Advertising*, 33(4), 89–97. <https://doi.org/10.1080/00913367.2004.10639175>
- Choi, H., Mela, C. F., Balseiro, S. R., & Leary, A. (2020). Online Display Advertising Markets: A Literature Review and Future Directions. *Information Systems Research*, 31(2), 556–575. <https://doi.org/10.1287/isre.2019.0902>
- Ciceri, A., Russo, V., Songa, G., Gabrielli, G., & Clement, J. (2020). A Neuroscientific Method for Assessing Effectiveness of Digital vs. Print Ads: Using Biometric Techniques to Measure Cross-Media Ad Experience and Recall. *Journal of Advertising Research*, 60(1), 71–86. <https://doi.org/10.2501/JAR-2019-015>
- Cleveland, W. S., & McGill, R. (1985). Graphical Perception and Graphical Methods for Analyzing Scientific Data. *Science*, 229(4716), 828–833. JSTOR.
- Conill, R. F. (2016). Camouflaging Church as State. *Journalism Studies*, 17(7), 904–914. <https://doi.org/10.1080/1461670X.2016.1165138>
- d’Haenens, L., Jankowski, N., & Heuvelman, A. (2004). News in Online and Print Newspapers: Differences in Reader Consumption and Recall. *New Media & Society*, 6(3), 363–382. <https://doi.org/10.1177/1461444804042520>
- Dahlén, M., Rosengren, S., Törn, F., & Öhman, N. (2008). Could Placing ADS Wrong be Right?: Advertising Effects of Thematic Incongruence. *Journal of Advertising*, 37(3), 57–67. <https://doi.org/10.2753/JOA0091-3367370305>
- Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, 2015(1), 92–112. <https://doi.org/10.1515/popets-2015-0007>
- de Vreese, C. H., & Neijens, P. (2016). Measuring media exposure in a changing communications environment. *Communication Methods and Measures*, 10(2–3), 69–80. <https://doi.org/10.1080/19312458.2016.1150441>

- Dean, W. (2013, October 9). *What is the purpose of journalism?* American Press Institute. <https://www.americanpressinstitute.org/journalism-essentials/what-is-journalism/purpose-journalism/>
- DellaVigna, S., & Kaplan, E. (2007). The Fox News Effect: Media Bias and Voting*. *The Quarterly Journal of Economics*, 122(3), 1187–1234. <https://doi.org/10.1162/qjec.122.3.1187>
- Diakopoulos, N. (2015). Algorithmic Accountability: Journalistic investigation of computational power structures. *Digital Journalism*, 3(3), 398–415. <https://doi.org/10.1080/21670811.2014.976411>
- Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.
- Dill-Shackleford, K. (Ed.). (2014). *The Oxford handbook of media psychology* (1. issue as an Oxford University Press paperback). Oxford-Univ.-Pr.
- Dor, D. (2003). On newspaper headlines as relevance optimizers. *Journal of Pragmatics*, 35(5), 695–721. [https://doi.org/10.1016/S0378-2166\(02\)00134-0](https://doi.org/10.1016/S0378-2166(02)00134-0)
- Dunaway, J., Searles, K., Sui, M., & Paul, N. (2018). News attention in a mobile era. *Journal of Computer-Mediated Communication*, 23(2), 107–124. <https://doi.org/10.1093/jcmc/zmy004>
- Eagle, N., & Pentland, A. S. (2009). Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7), 1057–1066. <https://doi.org/10.1007/s00265-009-0739-0>
- 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. <https://doi.org/10.1073/pnas.1419828112>
- Epstein, Z., Payne, B. H., Shen, J. H., Dubey, A., Felbo, B., Groh, M., Obradovich, N., Cebrian, M., & Rahwan, I. (2018). Closing the AI Knowledge Gap. *ArXiv:1803.07233v1 [Cs]*. <http://arxiv.org/abs/1803.07233v1>
- Esiyok, C., Kille, B., Jain, B.-J., Hopfgartner, F., & Albayrak, S. (2014). Users' reading habits in online news portals. *Proceedings of the 5th Information Interaction in Context Symposium on - IIX '14*, 263–266. <https://doi.org/10.1145/2637002.2637038>
- Evans, D. S. (2009). The online advertising industry: Economics, evolution, and privacy. *Journal of Economic Perspectives*, 23(3), 37–60. <https://doi.org/10.1257/jep.23.3.37>
- Exposito-Ventura, M., Ruiperez-Valiente, J. A., & Forne, J. (2020). Analyzing and Testing Viewability Methods in an Advertising Network. *IEEE Access*, 8, 118751–118761. <https://doi.org/10.1109/ACCESS.2020.3005478>
- Ferguson, J. M. (1983). Daily Newspaper Advertising Rates, Local Media Cross-Ownership, Newspaper Chains, and Media Competition. *The Journal of Law & Economics*, 26(3), 635–654.
- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298–320. <https://doi.org/10.1093/poq/nfw006>

- Fletcher, R., Kalogeropoulos, A., & Nielsen, R. K. (2021). More diverse, more politically varied: How social media, search engines and aggregators shape news repertoires in the United Kingdom. *New Media & Society*, 146144482110273. <https://doi.org/10.1177/14614448211027393>
- Fletcher, R., & Nielsen, R. K. (2017a). Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication: Are News Audiences Increasingly Fragmented? *Journal of Communication*, 67(4), 476–498. <https://doi.org/10.1111/jcom.12315>
- Fletcher, R., & Nielsen, R. K. (2017b). Paying for Online News: A comparative analysis of six countries. *Digital Journalism*, 5(9), 1173–1191. <https://doi.org/10.1080/21670811.2016.1246373>
- Fletcher, R., & Nielsen, R. K. (2018). Are people incidentally exposed to news on social media? A comparative analysis. *New Media & Society*, 20(7), 2450–2468. <https://doi.org/10.1177/1461444817724170>
- Fletcher, R., & Nielsen, R. K. (2019). Generalised scepticism: How people navigate news on social media. *Information, Communication & Society*, 22(12), 1751–1769. <https://doi.org/10.1080/1369118X.2018.1450887>
- Fletcher, R., & Nielsen, R. K. (2020). Are Netflix and Spotify Subscribers More Likely to Pay for Online News? Comparative Analysis of Data From Six Countries. *International Journal of Communication*, 14(0), 19.
- Fogg, B. J. (2003). *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan Kaufmann Publishers.
- Fogg, B. J., & Eckles, D. (Eds.). (2007). *Mobile Persuasion: 20 Perspectives on the Future of Behavior Change*. Stanford Captology Media.
- Gentzkow, M. (2014). Trading Dollars for Dollars: The Price of Attention Online and Offline. *The American Economic Review*, 104(5), 481–488.
- Gentzkow, M., & Shapiro, J. M. (2011). Ideological Segregation Online and Offline. *The Quarterly Journal of Economics*, 126(4), 1799–1839. <https://doi.org/10.1093/qje/qjr044>
- Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, 33(5), 587–606. <https://doi.org/10.1016/j.socec.2004.09.033>
- Gil de Zúñiga, H., Garcia-Perdomo, V., & McGregor, S. C. (2015). What Is Second Screening? Exploring Motivations of Second Screen Use and Its Effect on Online Political Participation: Second Screen Use & Online Political Participation. *Journal of Communication*, 65(5), 793–815. <https://doi.org/10.1111/jcom.12174>
- Goldfarb, A. (2014). What is Different About Online Advertising? *Review of Industrial Organization*, 44(2), 115–129. <https://doi.org/10.1007/s11151-013-9399-3>
- Goldfarb, A., & Tucker, C. (2011). Online Display Advertising: Targeting and Obtrusiveness. *Marketing Science*, 30(3), 389–404. <https://doi.org/10.1287/mksc.1100.0583>

- Groot Kormelink, T., & Costera Meijer, I. (2014). Tailor-Made News: Meeting the demands of news users on mobile and social media. *Journalism Studies*, *15*(5), 632–641. <https://doi.org/10.1080/1461670X.2014.894367>
- Haim, M., Graefe, A., & Brosius, H.-B. (2018). Burst of the Filter Bubble? *Digital Journalism*, *6*(3), 330–343. <https://doi.org/10.1080/21670811.2017.1338145>
- Harms, B., Bijmolt, T. H. A., & Hoekstra, J. C. (2017). Digital Native Advertising: Practitioner Perspectives and a Research Agenda. *Journal of Interactive Advertising*, *17*(2), 80–91. <https://doi.org/10.1080/15252019.2017.1357513>
- Healy, K. (2019). *Data Visualization: A practical introduction*. Princeton University Press. <http://socviz.co/>
- Hermida, A., Fletcher, F., Korell, D., & Logan, D. (2012). Share, Like, Recommend: Decoding the social media news consumer. *Journalism Studies*, *13*(5–6), 815–824. <https://doi.org/10.1080/1461670X.2012.664430>
- Hervet, G., Guérard, K., Tremblay, S., & Chtourou, M. S. (2011). Is banner blindness genuine? Eye tracking internet text advertising. *Applied Cognitive Psychology*, *25*(5), 708–716. <https://doi.org/10.1002/acp.1742>
- Hu, Y. (Jeffrey), Shin, J., & Tang, Z. (2015). Incentive Problems in Performance-Based Online Advertising Pricing: Cost per Click vs. Cost per Action. *Management Science*, *62*(7), 2022–2038. <https://doi.org/10.1287/mnsc.2015.2223>
- Huberman, B. A., Pirolli, P. L. T., Pitkow, J. E., & Lukose, R. M. (1998). Strong Regularities in World Wide Web Surfing. *Science*, *280*(5360), 95–97. <https://doi.org/10.1126/science.280.5360.95>
- Ingram Bogusz, C. (2018). Digital Trace Data: Which data should we collect and what should we do once we have it? In P. Andersson, S. Movin, M. Mähring, R. Teigland, & K. Wennberg (Eds.), *Managing digital transformation* (pp. 115–132). SSE Institute for Research, Stockholm School of Economics.
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, *2*(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Iyengar, S. (2011). Laboratory Experiments in Political Science. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge Handbook of Experimental Political Science* (pp. 73–88). Cambridge University Press. <https://doi.org/10.1017/CBO9780511921452>
- Jang, S., & Park, M. (2016). Do New Media Substitute for Old Media?: A Panel Analysis of Daily Media Use. *Journal of Media Economics*, *29*(2), 73–91. <https://doi.org/10.1080/08997764.2016.1170021>
- Jiang, S., Ferreira, J., & González, M. C. (2012). Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery*, *25*(3), 478–510. <https://doi.org/10.1007/s10618-012-0264-z>
- Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed). Oxford University Press.

- King, G., Schneer, B., & White, A. (2017). How the news media activate public expression and influence national agendas. *Science*, *358*(6364), 776–780. <https://doi.org/10.1126/science.aao1100>
- Kovach, B., & Rosenstiel, T. (2014). *The elements of journalism: What newspeople should know and the public should expect* (3rd ed.). Three Rivers Press.
- Küng, L. (2016). *Innovators in digital news*. I.B. Tauris & Co. Ltd in association with the Reuters Institute for the Study of Journalism, University of Oxford.
- Lakomaa, E. (2012). *Sedelpressen: Dagens industri under 30 år*. Hydra.
- Lalmas, M., O'Brien, H., & Yom-Tov, E. (2015). *Measuring user engagement*. Morgan & Claypool Publishers. <http://ieeexplore.ieee.org/servlet/opac?bknumber=6980476>
- Lambrecht, A., & Misra, K. (2016). Fee or Free: When Should Firms Charge for Online Content? *Management Science*, *63*(4), 1150–1165. <https://doi.org/10.1287/mnsc.2015.2383>
- Lambrecht, A., & Tucker, C. (2019). Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science*. <https://doi.org/10.1287/mnsc.2018.3093>
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Computational Social Science. *Science*, *323*(5915), 721–723. <https://doi.org/10.1126/science.1167742>
- Lin, M., Lucas, H. C., & Shmueli, G. (2013). Research Commentary—Too Big to Fail: Large Samples and the p-Value Problem. *Information Systems Research*, *24*(4), 906–917. <https://doi.org/10.1287/isre.2013.0480>
- Mazzocchi, F. (2015). Could Big Data be the end of theory in science?: A few remarks on the epistemology of data-driven science. *EMBO Reports*, *16*(10), 1250–1255. <https://doi.org/10.15252/embr.201541001>
- McCombs, M. E., & Shaw, D. L. (1972). The Agenda-Setting Function of Mass Media. *The Public Opinion Quarterly*, *36*(2), 176–187.
- McLuhan, M. (2010). *Understanding Media: The extensions of man* (Reprinted). Routledge. (Original work published 1964)
- Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*(2), 103–115.
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*(4), 806–834. <https://doi.org/10.1037/0022-006X.46.4.806>
- Michael, J. (1982). Distinguishing Between Discriminative and Motivational Functions of Stimuli. *Journal of the Experimental Analysis of Behavior*, *37*(1), 149–155. <https://doi.org/10.1901/jeab.1982.37-149>
- Michael, J. (1993). Establishing Operations. *The Behavior Analyst*, *16*(2), 191–206. <https://doi.org/10.1007/BF03392623>

- Mitchelstein, E., & Boczkowski, P. J. (2010). Online news consumption research: An assessment of past work and an agenda for the future. *New Media & Society*, 12(7), 1085–1102. <https://doi.org/10.1177/1461444809350193>
- Möller, J., van de Velde, R. N., Merten, L., & Puschmann, C. (2019). Explaining online news engagement based on browsing behavior: Creatures of habit? *Social Science Computer Review*, 089443931982801. <https://doi.org/10.1177/0894439319828012>
- Möller, N., & Petrini, I. (2020). XBT: *Existentiell beteendeterapi: Livet, döden och Skinner*.
- Molyneux, L. (2018). Mobile News Consumption: A habit of snacking. *Digital Journalism*, 6(5), 634–650. <https://doi.org/10.1080/21670811.2017.1334567>
- Mukerjee, S., Majó-Vázquez, S., & González-Bailón, S. (2018). Networks of Audience Overlap in the Consumption of Digital News. *Journal of Communication*, 68(1), 26–50. <https://doi.org/10.1093/joc/jqx007>
- Müller, P., Schneiders, P., & Schäfer, S. (2016). Appetizer or main dish? Explaining the use of Facebook news posts as a substitute for other news sources. *Computers in Human Behavior*, 65, 431–441. <https://doi.org/10.1016/j.chb.2016.09.003>
- Napoli, P. M., & Obar, J. A. (2014). The Emerging Mobile Internet Underclass: A Critique of Mobile Internet Access. *The Information Society*, 30(5), 323–334. <https://doi.org/10.1080/01972243.2014.944726>
- Neelon, B., Ghosh, P., & Loeb, P. F. (2013). A spatial Poisson hurdle model for exploring geographic variation in emergency department visits: *Spatial Hurdle Model for Exploring Geographic Variation*. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 389–413. <https://doi.org/10.1111/j.1467-985X.2012.01039.x>
- Nelson, J. L. (2020). The persistence of the popular in mobile news consumption. *Digital Journalism*, 8(1), 87–102. <https://doi.org/10.1080/21670811.2019.1612766>
- Nelson, J. L., & Lei, R. F. (2018). The effect of digital platforms on news audience behavior. *Digital Journalism*, 6(5), 619–633. <https://doi.org/10.1080/21670811.2017.1394202>
- Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323–351. <https://doi.org/10.1080/00107510500052444>
- Newman, N., Fletcher, R., Schulz, A., Andi, S., Robertson, C. T., & Nielsen, R. K. (2021). *Reuters Institute Digital News Report 2021* (SSRN Scholarly Paper ID 3873260). Reuters Institute for the Study of Journalism. <https://papers.ssrn.com/abstract=3873260>
- Normand, M. P., & Kohn, C. S. (2013). Don't Wag the Dog: Extending the Reach of Applied Behavior Analysis. *The Behavior Analyst*, 36(1), 109–122.
- Oh, H., Animesh, A., & Pinsonneault, A. (2016). Free Versus For-a-Fee: The Impact of a Paywall on the Pattern and Effectiveness of Word-of-Mouth via Social Media. *MIS Quarterly*, 40(1), 31–56. <https://doi.org/10.25300/MISQ/2016/40.1.02>
- Ohlsson, J., Hellingwerf, K., Facht, U., & Bucht, C. (2020). *Mediebarometern 2019*. Nordicom, University of Gothenburg.
- Öhrvall, S. (2015). *The Future of News Journalism in the Digital Media Market*. Vinnova.

- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. <https://doi.org/10.1126/science.aac4716>
- Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable Webcam Eye Tracking Using User Interactions. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence - IJCAI 2016*. <https://par.nsf.gov/biblio/10024076-webgazer-scalable-webcam-eye-tracking-using-user-interactions>
- Pennypacker, H. S. (1992). Is behavior analysis undergoing selection by consequences? *American Psychologist*, *47*(11), 1491–1498. <https://doi.org/10.1037/0003-066X.47.11.1491>
- Perone, M., & Hursh, D. E. (2013). Single-case experimental designs. In G. J. Madden, W. V. Dube, T. D. Hackenberg, G. P. Hanley, & K. A. Lattal (Eds.), *APA handbook of behavior analysis: Vol. Vol. 1: Methods and principles* (pp. 107–126). American Psychological Association. <https://doi.org/10.1037/13937-005>
- Petrova, M. (2011). Newspapers and Parties: How Advertising Revenues Created an Independent Press. *The American Political Science Review*, *105*(4), 790–808.
- Pickard, V., & Williams, A. T. (2014). Salvation Or Folly?: The promises and perils of digital paywalls. *Digital Journalism*, *2*(2), 195–213. <https://doi.org/10.1080/21670811.2013.865967>
- Postman, N. (2006). *Amusing ourselves to death: Public discourse in the age of show business* (20th anniversary ed). Penguin Books. (Original work published 1985)
- Prawesh, S., & Padmanabhan, B. (2014). The “Most Popular News” Recommender: Count Amplification and Manipulation Resistance. *Information Systems Research*, *25*(3), 569–589. <https://doi.org/10.1287/isre.2014.0529>
- Prior, M. (2003). Any good news in soft news? The impact of soft news preference on political knowledge. *Political Communication*, *20*(2), 149–171. <https://doi.org/10.1080/10584600390211172>
- Prior, M. (2005). News vs. Entertainment: How Increasing Media Choice Widens Gaps in Political Knowledge and Turnout. *American Journal of Political Science*, *49*(3), 577–592. <https://doi.org/10.1111/j.1540-5907.2005.00143.x>
- Prior, M. (2009a). The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly*, *73*(1), 130–143. <https://doi.org/10.1093/poq/nfp002>
- Prior, M. (2009b). Improving Media Effects Research through Better Measurement of News Exposure. *The Journal of Politics*, *71*(3), 893–908. <https://doi.org/10.1017/S0022381609090781>

- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A. 'Sandy,' ... Wellman, M. (2019). Machine behaviour. *Nature*, *568*(7753), 477–486. <https://doi.org/10.1038/s41586-019-1138-y>
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532. <https://doi.org/10.1037/0033-2909.114.3.510>
- Robertson, R. E. (2018, January 15). The 21st Century Skinner Box. *Behavioral Scientist*, *Special issue: Connected State of Mind*. <https://behavioralscientist.org/21st-century-skinner-box/>
- Scharkow, M., Mangold, F., Stier, S., & Breuer, J. (2020). How social network sites and other online intermediaries increase exposure to news. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1918279117>
- Schmidt, A. L., Zollo, F., Del Vicario, M., Bessi, A., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2017). Anatomy of news consumption on Facebook. *Proceedings of the National Academy of Sciences*, *114*(12), 3035–3039. <https://doi.org/10.1073/pnas.1617052114>
- Schork, N. J. (2015). Personalized medicine: Time for one-person trials. *Nature*, *520*(7549), 609–611. <https://doi.org/10.1038/520609a>
- Semenov, A., Rysz, M., Pandey, G., & Xu, G. (2022). Diversity in news recommendations using contextual bandits. *Expert Systems with Applications*, *195*, 116478. <https://doi.org/10.1016/j.eswa.2021.116478>
- Sen, A. (2019). Profits and Profitability of News Media. In T. P. Vos & F. Hanush (Eds.), *The International Encyclopedia of Journalism Studies* (pp. 1–6). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118841570.iejs0070>
- Sen, A., & Yildirim, P. (2015). Clicks Bias in Editorial Decisions: How Does Popularity Shape Online News Coverage? *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2619440>
- Skinner, B. F. (1965). *Science and human behavior* (First Free Press Paperback edition). The Free Press. (Original work published 1953)
- Srivastava, M., Garg, R., & Mishra, P. K. (2015). Analysis of Data Extraction and Data Cleaning in Web Usage Mining. *Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015)*, 1–6. <https://doi.org/10.1145/2743065.2743078>
- Stroud, N. J., Peacock, C., & Curry, A. L. (2020). The Effects of Mobile Push Notifications on News Consumption and Learning. *Digital Journalism*, *8*(1), 32–48. <https://doi.org/10.1080/21670811.2019.1655462>
- Sturmey, P. (2008). Radical Behaviorism. In *Behavioral Case Formulation and Intervention: A Functional Analytic Approach*. (pp. 21–35). John Wiley & Sons. <https://doi.org/10.1002/9780470773192>

- Sweeney, L. (2013). Discrimination in Online Ad Delivery. *Queue*, 11(3), 10. <https://doi.org/10.1145/2460276.2460278>
- Taneja, H. (2016). Using Commercial Audience Measurement Data in Academic Research. *Communication Methods and Measures*, 10(2–3), 176–178. <https://doi.org/10.1080/19312458.2016.1150971>
- Taneja, H., Wu, A. X., & Edgerly, S. (2018). Rethinking the generational gap in online news use: An infrastructural perspective. *New Media & Society*, 20(5), 1792–1812. <https://doi.org/10.1177/1461444817707348>
- Tewksbury, D. (2003). What Do Americans Really Want to Know? Tracking the Behavior of News Readers on the Internet. *Journal of Communication*, 53(4), 694–710. <https://doi.org/10.1111/j.1460-2466.2003.tb02918.x>
- The Economist Intelligence Unit. (2020). *Democracy Index 2020: In sickness and in health?* www.eiu.com
- The Economist Intelligence Unit. (2021). *The Inclusive Internet Index* [Dataset]. <https://theinclusiveinternet.eiu.com/>
- Thorson, E. (2008). Changing Patterns Of News Consumption And Participation: News Recommendation Engines. *Information, Communication & Society*, 11(4), 473–489. <https://doi.org/10.1080/13691180801999027>
- Thyer, B. A. (2009). Epistemology: A behavior analytic perspective. *Humana. Mente*, 3(11), 45–63.
- Toff, B., & Kalogeropoulos, A. (2020). All the News That's Fit to Ignore. *Public Opinion Quarterly*, 84(S1), 366–390. <https://doi.org/10.1093/poq/nfaa016>
- Tufte, E. R. (1997). *Visual explanations: Images and quantities, evidence and narrative*. Graphics Press.
- Vaccari, C., Chadwick, A., & O'Loughlin, B. (2015). Dual Screening the Political: Media Events, Social Media, and Citizen Engagement: Dual Screening the Political. *Journal of Communication*, 65(6), 1041–1061. <https://doi.org/10.1111/jcom.12187>
- Valkenburg, P. M., Peter, J., & Walther, J. B. (2016). Media Effects: Theory and Research. *Annual Review of Psychology*, 67(1), 315–338. <https://doi.org/10.1146/annurev-psych-122414-033608>
- Van der Meer, T. G. L. A., Hameleers, M., & Kroon, A. C. (2020). Crafting Our Own Biased Media Diets: The Effects of Confirmation, Source, and Negativity Bias on Selective Attendance to Online News. *Mass Communication and Society*, 23(6), 937–967. <https://doi.org/10.1080/15205436.2020.1782432>
- Vermeer, S., Trilling, D., Kruikemeier, S., & de Vreese, C. (2020). Online News User Journeys: The Role of Social Media, News Websites, and Topics. *Digital Journalism*, 1–28. <https://doi.org/10.1080/21670811.2020.1767509>

- Vrijheid, M., Armstrong, B. K., Bédard, D., Brown, J., Deltour, I., Iavarone, I., Krewski, D., Lagorio, S., Moore, S., Richardson, L., Giles, G. G., McBride, M., Parent, M.-E., Siemiatycki, J., & Cardis, E. (2009). Recall bias in the assessment of exposure to mobile phones. *Journal of Exposure Science & Environmental Epidemiology*, *19*(4), 369–381. <https://doi.org/10.1038/jes.2008.27>
- Wadbring, I., & Bergström, L. (2021). Audiences behind the Paywall: News Navigation among Established versus Newly Added Subscribers. *Digital Journalism*, *9*(3), 319–335. <https://doi.org/10.1080/21670811.2021.1878919>
- Wahlund, R., Rademaker, C., Nilsson, C. P., & Svahn, M. (2013). Mediernas roll i marknadskommunikationen. In G. Nygren & I. Wadbring (Eds.), *På väg mot mediavärlden 2020: Journalistik, teknik, marknad* (5th ed., pp. 159–182). Studentlitteratur.
- Wahlund, R., Svahn, M., & Åbonde Garke, Adam. (2021, June 10). *Consumers' evaluations of journalistic and private news in different digital media – An experimental study*. IAREP-SABE. https://sabeconomics.org/iarep-sabe-2021/wordpress/wp-content/uploads/cfdb7_uploads/1622657687-file-abstract-ID240_Consumers-evaluation-The-experimental-study-IAREP-SABE.pdf
- Wang, C.-C., & Hung, J. C. (2019). Comparative analysis of advertising attention to Facebook social network: Evidence from eye-movement data. *Computers in Human Behavior*, *100*, 192–208. <https://doi.org/10.1016/j.chb.2018.08.007>
- Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, *59*(1), 1–23. <https://doi.org/10.18637/jss.v059.i10>
- Wojdyski, B. W. (2016). The Deceptiveness of Sponsored News Articles: How Readers Recognize and Perceive Native Advertising. *American Behavioral Scientist*, *60*(12), 1475–1491. <https://doi.org/10.1177/0002764216660140>
- Wu, T. (2017). *The attention merchants: The epic scramble to get inside our heads* (First Vintage Books Edition). Vintage Books, a division of Penguin Random House LLC.
- Xu, J., Forman, C., Kim, J. B., & Van Ittersum, K. (2014). News Media Channels: Complements or Substitutes? Evidence from Mobile Phone Usage. *Journal of Marketing*, *78*(4), 97–112. <https://doi.org/10.1509/jm.13.0198>
- Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. <https://doi.org/10.1177/1745691617693393>
- Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98. <https://doi.org/10.1007/s10683-009-9230-z>