Media and Communication
Open Access Journal | ISSN: 2183-2439
Volume 9, Issue 4 (2021)

Algorithmic Systems in the Digital Society

Editors

Sanne Kruikemeier, Sophie Boerman and Nadine Bol

COGITATIO

Media and Communication, 2021, Volume 9, Issue 4
Algorithmic Systems in the Digital Society

Published by Cogitatio Press
Rua Fialho de Almeida 14, 2º Esq.,
1070-129 Lisbon
Portugal

Available online at: www.cogitatiopress.com/mediaandcommunication

Table of Contents

Table of Contents

COGITATIO

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 4, Pages 116–119
https://doi.org/10.17645/mac.v9i4.5005

Editorial

How Algorithmic Systems Changed Communication in a Digital Society

Sanne Kruikemeier [1,*], Sophie C. Boerman [1] and Nadine Bol [2]

[1] Amsterdam School of Communication Research, University of Amsterdam, The Netherlands;
E-Mail: s.kruikemeier@uva.nl (S.K.), s.c.boerman@uva.nl (S.C.B.)
[2] Department of Communication and Cognition, Tilburg Center for Cognition and Communication, Tilburg University, The Netherlands; E-Mail: nadine.bol@tilburguniversity.edu

* Corresponding author

Submitted: 28 October 2021 | Published: 18 November 2021

Abstract
This thematic issue invited submissions that address the opportunities and controversies related to algorithmic influence in a digital society. A total of 11 articles address how the use of algorithms has changed communication in various contexts, and cover topics such as personalized marketing communication, self-tracking for health, political microtargeting, news recommenders, social media algorithms, and urban experiences. The articles also include a wide variety of methods such as surveys, experiments, expert interviews, computational methods, and theoretical work developing frameworks and typologies. They are all united by one central question: How have algorithms and artificial intelligence changed communication, for both senders and receivers? We believe that the collection of topics and methods provide new insights into the different perspectives regarding algorithmic-driven communication—highlighting both the opportunities and challenges—and advance the literature with new findings, frameworks, and typologies.

Keywords
algorithms; automated decision making; communication; digital divides; health trackers; media personalization; online privacy; political microtargeting; recommender systems; transparency

Issue
This editorial is part of the issue "Algorithmic Systems in the Digital Society" edited by Sanne Kruikemeier (University of Amsterdam, The Netherlands), Sophie Boerman (University of Amsterdam, The Netherlands), and Nadine Bol (Tilburg University, The Netherlands).

Algorithms and artificial intelligence (AI) have changed communication delivery modes in society. This is especially noticed by a shift from "mass communication" to increasingly more personalized and automated communication. For instance, by using a vast amount of data, communicators can increasingly personalize (i.e., match messages to characteristics of an individual) and target (i.e., send these matched messages to specific people) their messages. Consequently, algorithms may increasingly be used for automated decision-making. This means that data-driven technologies can be used to make decisions about our life without the interference of humans. This thematic issue addresses the opportunities and challenges related to algorithmic influence in a digital society. A total of 11 articles address how the use of algorithms has changed communication in various contexts, and cover topics such as personalized marketing communication, self-tracking for health, political microtargeting, news recommenders, social media algorithms, and the algorithmic curation of urban experiences. The articles also represent a wide variety of methods such as surveys, experiments, expert interviews, and computational methods, as well as more theory-driven approaches, such as developing frameworks and typologies.

The issue starts with a literature review of academic research on transparency and control in personalized (marketing) communication (Segijn et al., 2021). With its focus on transparency and control, this article addresses an important issue of algorithmic and data-driven communication. Based on their literature review,

the authors conclude that there is little consensus about the definitions of personalization transparency and control. The authors conceptualize personalization transparency and control, and propose that *transparency* involves the degree to which the sender is open about data collection, processing, and sharing, whereas *control* involves the extent to which receivers can start, stop, or maintain data processing. The authors ultimately present the transparency–awareness–control framework which illustrates how the constructs are related and provide concrete propositions to guide future research based upon this framework.

The article by Zarouali et al. (2021) provides further insights into the receiver side of algorithmic communication. The authors present the outcomes of a large survey, which shows that misconceptions about algorithms in the media are highly prevalent among the general population in the Netherlands. Additionally, they show that erroneous representations about media algorithms are more common among older people, lower-educated people, and women, suggesting that algorithms may be expanding digital divides.

Algorithmic-driven processes are also used in health technology, such as in self-tracking applications. Festic et al. (2021) discuss the results of a large representative survey that examines users' risk perceptions and coping strategies to deal with the risks associated with their use of self-tracking applications. They conclude that users' risk awareness is generally low and only a small proportion of the sample applied coping strategies, such as checking the accuracy of self-tracking measurements, to retain autonomy and mitigate the risks of self-tracking. A substantial proportion of the sample indicated to be willing to share their health data with their health insurance if they receive financial advantages for doing so. They further discuss their findings in the light of the privacy calculus by arguing that the expected benefits of using self-tracking technology may outweigh the potential risks.

The fourth contribution by Schäwel et al. (2021) moves to another important topic in which algorithms play a crucial role. Their work focuses on political microtargeting and online privacy. They elaborate on social media users' privacy perceptions and potential regulating behaviors when confronted with political microtargeting. The authors follow the lines of the social media privacy model (Trepte, 2020), and focus on the process of social media privacy as experienced by users when being confronted with political behavioral targeting. Based on their model, they present propositions for future research when analyzing political microtargeting. First, they argue that it is important to consider the complexity of the social media context and individuals' perceptions of it. Second, they argue that it is important to understand users' privacy experiences affecting the outcome of microtargeting. Lastly, they make an important point that it is very important to conduct research regarding microtargeting and privacy along with ethical guidelines.

The following contributions focus on news. First, Jia and Liu (2021) examine whether the source attribution of a news article (human or algorithmic or human-assisted algorithm) affects hostile media perceptions. They found, among other things, that the relative hostile media effect occurs when people read headlines attributed to an algorithmic author. As pointed out by the authors, this indicates that positive perceptions regarding the neutrality of algorithms may not always be true. The next contribution by van der Velden and Loecherbach (2021) focuses also on news consumption and examines the reasons and motivations towards algorithmic versus human gatekeepers. While the focus is different, they found that for surveillance gratifications (keeping up with politics), algorithms are more appreciated. Conversely, when users consume news to pass time, escape from daily worries, or for entertainment, people are less likely to prefer algorithmic news selection. They also found in their study an interesting conditional effect: Users who are more confident in their own abilities are more likely to prefer algorithmic gatekeepers for surveillance gratifications.

The next study focuses on newsroom innovation labs. Cools et al. (2021) examine how algorithmic news recommenders may affect the gatekeeping role of news workers in the newsgathering process and the autonomy of the news workers' role as media agenda setters. The results show that when news workers interact with algorithmic news recommenders, they rely on them to evaluate what is newsworthy, in particular during specific periods, such as an election or a pandemic. They also found that the news workers are fully autonomous, but the algorithmic news recommenders seem to have a positive effect on how certain topics are put on the agenda. Lastly, Wieland et al. (2021) look at news recommenders from the user perspective. The authors report a survey containing an innovative self-programmed recommendation system to study how users evaluate algorithmic news recommendations. They find that users prefer recommendations of the most similar, and not necessarily unexpected, articles, but evaluations also differ depending on personal characteristics.

Another context in which algorithms may play a crucial role is that of social media. Their platforms have received a lot of criticism over the years for reasons related to privacy breaches and manipulative practices. Saurwein and Spencer-Smith (2021) propose a typology of "algorithmic harm" to describe the various harmful or negative effects upon individuals, markets, and society caused in part or in full by the use of algorithms. Their typology includes harms related to algorithmic errors, undesirable or disturbing selections, manipulation by users to achieve algorithmic outputs to harass other users or disrupt public discourse, algorithmic reinforcement of pre-existing harms and inequalities in society, enablement of harmful practices that are opaque and discriminatory, and strengthening of platform power over users, markets, and society. Based on their discussion, they reflect on potential governance strategies to

combat algorithmic harm and reduce platform power by introducing effective ways of external oversight.

Matamoros-Fernández et al. (2021) zoom in on the specific algorithmic selection processes of YouTube's "up next" feature. By combining computational and qualitative methods, they investigate the type of content displayed by the algorithms underpinning the "up next" feature and discuss to what extent negative claims—such as limiting users' exposure to diverse media content—regarding these algorithms can be empirically proven to be true. This article shows that despite YouTube's diverse algorithmic-driven recommendations, clear "winners" tend to dominate the "up next" selection.

Algorithms also increasingly shape aspects of urban life. As such, the impact of algorithms and their selection processes do not only pertain to the online world, but also impact many offline practices, such as choices of where we sleep, eat, and go. Smets et al. (2021), discuss how the widespread diffusion of digital communication technologies has entered all aspects of urban life and how selection processes shape urban experiences. Based on a literature review, they identify the vast amount of work on algorithmic selection in the online world and use this to construct an analytical lens to study the algorithmic urban experiences. They conclude their article by proposing an integrative framework on algorithmic curation of urban experiences, in which the multiple ways for algorithms to curate urban experiences have been illustrated.

This thematic issue offers a collection of articles that show more refined insight into how algorithms and AI changed communication in different contexts. We believe that the collection of topics, concepts, ideas, methods, findings, and discussed implications provide new insights into the different perspectives regarding algorithmic-driven communication. The articles included in this thematic issue highlight both the opportunities and challenges of algorithmic-driven communication, provide a more nuanced picture of algorithmic impacts by discussing the different boundary conditions in different contexts and advance the literature with new findings, frameworks, and typologies.

Conflict of Interests

The authors declare no conflict of interests.

References

Cools, H., Van Gorp, B., & Opgenhaffen, M. (2021). When algorithms recommend what's new(s): New dynamics of decision-making and autonomy in newsgathering. *Media and Communication*, *9*(4), 198–207.

Festic, N., Latzer, M., & Smirnova, S. (2021). Algorithmic self-tracking for health: User perspectives on risk awareness and coping strategies. *Media and Communication*, *9*(4), 145–157.

Jia, C., & Liu, R. (2021). Algorithmic or human source? Examining relative hostile media effect with a transformer-based framework. *Media and Communication*, *9*(4), 170–181.

Matamoros-Fernández, A., E. Gray, J., Bartolo, L., Burgess, J., & Suzor, N. (2021). What's "up next"? Investigating algorithmic recommendations on YouTube across issues and over time. *Media and Communication*, *9*(4), 234–249.

Saurwein, F., & Spencer-Smith, C. (2021). Automated trouble: The role of algorithmic selection in harms on social media platforms. *Media and Communication*, *9*(4), 222–233.

Schäwel, J., Frener, R., & Trepte, S. (2021). Political microtargeting and online privacy: A theoretical approach to understanding users' privacy behaviors. *Media and Communication*, *9*(4), 158–169.

Segijn, C. M., Strycharz, J., Riegelman, A., & Hennesy, C. (2021). A literature review of personalization transparency and control: Introducing the transparency–awareness–control framework. *Media and Communication*, *9*(4), 120–133.

Smets, A., Ballon, P., & Walravens, N. (2021). Mediated by code: Unpacking algorithmic curation of urban experiences. *Media and Communication*, *9*(4), 250–259.

Trepte, S. (2020). The social media privacy model: Privacy and communication in the light of social media affordances. *Communication Theory.* Advance online publication. https://doi.org/10.1093/ct/qtz035

van der Velden, M., & Loecherbach, F. (2021). Epistemic overconfidence in algorithmic news selection. *Media and Communication*, *9*(4), 182–197.

Wieland, M., von Nordheim, G., & Kleinen-von Königslöw, K. (2021). One recommender fits all? An exploration of user satisfaction with text-based news recommender systems. *Media and Communication*, *9*(4), 208–221.

Zarouali, B., Helberger, N., & de Vreese, C. H. (2021). Investigating algorithmic misconceptions in a media context: Source of a new digital divide? *Media and Communication*, *9*(4), 134–144.

About the Authors

Sanne Kruikemeier (PhD, University of Amsterdam, 2014) is associate professor of political communication and Journalism at the University of Amsterdam and the Amsterdam School of Communication Research. Her research focuses on the consequences and implications of online communication for individuals and society.

Sophie C. Boerman (PhD, University of Amsterdam, 2014) is assistant professor of persuasive communication at the Amsterdam School of Communication Research at the University of Amsterdam. Her research addresses empowerment and resilience in the context of persuasive and/or data-driven communication, with a focus on transparency, literacy and persuasion knowledge, and privacy.

Nadine Bol (PhD, University of Amsterdam) is assistant professor of health communication in the Department of Communication and Cognition at Tilburg University. Her research expertise lies at the intersection of digital technologies, health communication, and vulnerability, centering on how digital health technologies impact vulnerable populations and create (new) digital inequalities.

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 4, Pages 120–133
https://doi.org/10.17645/mac.v9i4.4054

Article

A Literature Review of Personalization Transparency and Control: Introducing the Transparency–Awareness–Control Framework

Claire M. Segijn [1,†], Joanna Strycharz [2,*,†], Amy Riegelman [3] and Cody Hennesy [3]

[1] Hubbard School of Journalism and Mass Communication, University of Minnesota, USA; E-Mail: segijn@umn.edu
[2] Amsterdam School of Communication Research, University of Amsterdam, The Netherlands; E-Mail: j.strycharz@uva.nl
[3] University Libraries, University of Minnesota, USA; E-Mails: aspringe@umn.edu (A.R.), chennesy@umn.edu (C.H.)

* Corresponding author
† These authors contributed equally to this work

Submitted: 28 January 2021 | Accepted: 16 March 2021 | Published: 18 November 2021

Abstract
Through various online activities, individuals produce large amounts of data that are collected by companies for the purpose of providing users with personalized communication. In the light of this mass collection of personal data, the transparency and control paradigm for personalized communication has led to increased attention from legislators and academics. However, in the scientific literature no clear definition of personalization transparency and control exists, which could lead to reliability and validity issues, impeding knowledge accumulation in academic research. In a literature review, we analyzed 31 articles and observed that: 1) no clear definitions of personalization transparency or control exist; 2) they are used interchangeably in the literature; 3) collection, processing, and sharing of data are the three objects of transparency and control; and 4) increased transparency does not automatically increase control because first awareness needs to be raised in the individual. Also, the relationship between awareness and control depends on the ability and the desire to control. This study contributes to the field of algorithmic communication by creating a common understanding of the transparency and control paradigm and thus improves validity of the results. Further, it progresses research on the issue by synthesizing existing studies on the topic, presenting the transparency–awareness–control framework, and formulating propositions to guide future research.

Keywords
awareness; computational advertising; consumer data; control; covert data collection; information disclosure; personalization; privacy; targeting; transparency

Issue
This article is part of the issue "Algorithmic Systems in the Digital Society" edited by Sanne Kruikemeier (University of Amsterdam, The Netherlands), Sophie Boerman (University of Amsterdam, The Netherlands), and Nadine Bol (Tilburg University, The Netherlands).

1. Introduction

Through various online activities, individuals produce large amounts of data that are collected by companies and processed through algorithms for the purpose of providing users with personalized communication (Yun et al., 2020). While personalization is currently applied in many different contexts—e.g., personalized healthcare (Dzau & Ginsburg, 2016) or news recommendations (Thurman et al., 2019)—it very frequently occurs in the form of personalized marketing messages (so-called personalized marketing communication, see Strycharz, van Noort, Helberger, et al., 2019). In this context, personalized communication involves interactions between

companies and consumers, data collection, and processing by companies and delivery of marketing communication (Vesanen & Raulas, 2006).

In the light of this mass collection and advanced processing of personal data through algorithms for the means of personalized communication, disclosures about data collection and processing and individual control of these processes—often called the transparency and control paradigm—have been gaining importance in practice (Deloitte, 2018; Li et al., 2019). For example, recent legal developments, such as the General Data Protection Regulation (GDPR) in the EU in 2016 (enforcement in 2018), and the California Consumer Privacy Act in the US, assign high transparency requirements for companies' data collection and processing practices and strengthen individuals' rights to control their personal data as the main data protection mechanisms (Strycharz et al., 2020; van Ooijen & Vrabec, 2019).

The growing importance of the transparency and control paradigm in application of personalized marketing communication is also reflected in academic research. The effects that data collection transparency has on users have been investigated (e.g., Aguirre et al., 2015; Kim et al., 2019) as have the ways in which control over data collection impacts users' perceptions and behavior (e.g., Strycharz, van Noort, Smit, et al., 2019; Zarouali et al., 2018). Individual control over personal data has also been portrayed as a crucial element of privacy (Altman, 1975). However, the literature provides little consensus on how personalization transparency and control should be conceptualized. For example, while Aguirre and et al. (2015) call transparency "overt data collection" and focus on consumer awareness of data collection practices, Kim et al. (2019) write about "ad transparency" in terms of the disclosure of data collection practices. Similarly, control has been conceptualized as abilities users have to control data collection (Joo, 2018), but also as the desires that users have to exercise such control (Strycharz, van Noort, Smit, et al., 2019). Such substantial differences in conceptualizations impact the reliability and validity of the results and impede knowledge accumulation in the field. Therefore, the aim of the current study is to map academic research on personalization transparency and control and provide guidelines for future research on this issue.

To map academic research on transparency and control in personalized marketing communication, we conduct a systematic literature review on personalization transparency and control, provide conceptualizations, and develop a framework to facilitate future research on the topic. The research delivers a substantial contribution to the field of personalized marketing communication by creating a common understanding of the transparency and control paradigm, thus improving the validity of future results. Further, it progresses research by synthesizing existing studies on the topic and presents a framework to guide future research on the topic.

2. Methods

To locate the relevant literature included in this study, electronic searches were conducted in several disciplinary and multidisciplinary databases in July 2020. The primary search strategy was designed by social sciences librarians and conducted in Business Source Premier (EBSCO). It was then translated and conducted in Academic Search Premier (EBSCO), Communication and Mass Media Complete (EBSCO), and Proquest Dissertations & Theses. Due to technical limitations of search interfaces precluding systematic searching, studies from SocArXiv and the Social Science Research Network (SSRN) were obtained via hand searches using relevant keywords and the manual review of search results.

The literature search identified published and unpublished empirical and theoretical studies in databases focused on advertising, marketing, communication sciences, and business that include conceptualizations of the personalization transparency and control paradigm. The full search query and filters for the initial database are available in the Supplementary Files.

The 643 results identified from searching the six databases were exported to EndNote, a citation manager, and were deduplicated, resulting in 589 records. An online systematic review software, Rayyan QCRI, was used to assist the two-part screening process (Ouzzani et al., 2016). Titles and abstracts of the initial 589 studies were independently screened by the first and second authors which resulted in 36 studies. The authors had a percentage agreement of 95.8% and 100% was reached after discussions. An additional three studies were added based on backward citation tracking and one more was added based on the authors' professional contacts, for a total of 40 studies for full-text screening.

The primary two authors screened the full text of 40 studies, narrowing the sample to 31. The PRISMA flow diagram in Figure 1 includes the search, deduplication, screening, and data extraction totals for this study. The authors adhered to the PRISMA statement and checklist for this study to transparently report the procedures (Moher et al., 2009). The coding protocol as well as full coding scheme can be found in the Supplementary Files.

3. Results

The selected articles were qualitatively coded by the first two authors and an overview of the results is presented in Tables 1 and 2. In total, we included 31 studies, eight of which concern transparency, while 25 cover control (two articles mentioned both concepts). We observed that often transparency and control were used interchangeably. Although the concepts are related to each other, we believe that they are different constructs. Additionally, we observed that only a few papers included an explicit definition of transparency or control. From most papers we were able to derive the conceptualization from the text but for others we

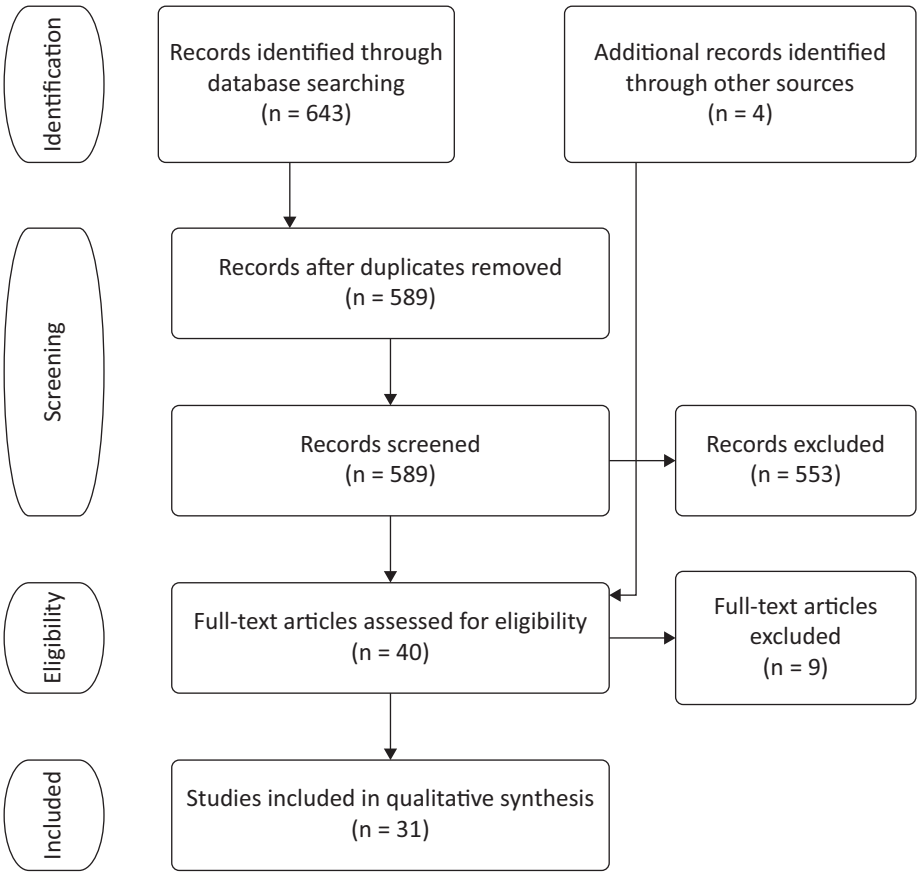


Figure 1. PRISMA flow diagram. Source: Flow diagram adapted from Moher et al. (2009).

were unable to derive any understanding of the use of these concepts.

3.1. Conceptualization of Transparency

Of the eight studies on transparency, six were published and two were not (a Master's thesis and a dissertation). The majority of the papers came from marketing (*n* = 5) and three papers came from a communication science journal (see Table 1). Three papers were published in or after 2018, five were published in or after 2015, and all were published in the 2000s. This indicates that research interest in the issue of transparency has not decreased over time. The majority of the research was conducted in the US (*n* = 4), followed by Europe (Netherlands, Germany, and Sweden).

We observed many different conceptualizations of transparency. This may be explained by the fact that the transparency object differs between studies. We observed three objects of transparency. In the first, transparency concerns data collection practices in general terms, or is related to specific data collection techniques, such as cookies. In the second, a few studies look into transparency of the personalization process (i.e., how data is used for the creation and delivery of personalized messages). In the third, transparency concerns sharing data with third parties.

Looking closer at these conceptualizations, we observed two further differences between how authors use the concept of transparency. First, it is used from the perspective of the data collector or the sender of the personalized message (industry perspective). This refers to the information that is disclosed by the sender to individuals about data collection, the personalization process, or data sharing. In general, transparency from the sender perspective focuses on how such information is disclosed and is closely related to choices about data collection made by the collector. Second, the concept is used in reference to the perspective of the individual whose data is collected and who is the recipient of the personalized message. In this case, it refers not to transparency but to the degree of awareness of data collection, processing, or sharing. Such awareness is also referred to as the overtness of data collection and is closely related to transparency. Looking at the link between these two uses, we argue that transparency (stemming from the industry), in fact, can lead to increased awareness among individuals.

Based on the different conceptualizations of transparency, we propose to differentiate between transparency and awareness. These constructs are both used in the reviewed literature but with substantially different conceptualizations and from different perspectives (e.g., sender vs. receiver). Building on the conceptualization

provided by Kim et al. (2019) and adjusted according to other reviewed studies related to transparency (see Table 1), we propose the following definition for personalization transparency:

> Personalization transparency: The degree of disclosure of the ways in which firms collect, process, or share (exchange) personal data with the purpose of generating personalized communication.

Next, building on the conceptualization provided by Aguirre et al. (2015) and adjusting it to the reviewed articles on awareness (see Table 1), we propose the following definition for personalization awareness:

> Personalization awareness: The degree to which individuals are cognizant of how and when their personal data are collected, processed, or shared (exchanged) with the purpose of generating personalized communication.

3.2. Conceptualization of Data Control

Of the 25 studies on control, 20 were published and five were unpublished. The majority of the papers were from marketing (*n* = 12), followed by communication (*n* = 7), and law and ethics (*n* = 3; Table 2). Ten papers were published in or after 2018, 15 were published in or after 2015, and all were published in the 2000s. Similar to transparency, we observed that all articles are by different authors. The research was conducted in five different continents with the majority of the data collected in the US (*n* = 8), followed by Europe (*n* = 6), and also including data from South Africa (*n* = 3), South Korea (*n* = 1), and New Zealand (*n* = 1).

We observed different conceptualizations of data control, and many different terms were used to describe it (Table 2). This is not surprising given that most of the work does not build on the other studies under analysis, but rather were developed in parallel around the same time period. Looking more closely at the conceptualizations, we observe three main differences: 1) type of control, 2) concreteness, and 3) object of control.

First, the type of control differs depending on whether the authors talk about actual control (e.g., things that an individual can do) vs. perceived control (e.g., the perceptions of control by an individual). Literature on personalization has shown the importance of separating between reality and perceptions, for example, in (perceived) personalization (De Keyzer et al., 2015; Kramer et al., 2007; Maslowska et al., 2016). Maslowska et al. (2016) found that perceived personalization mediates the relationship between actual personalization and advertising responses. Therefore, we find it important to also distinguish between actual and perceived data control.

Second, the conceptualizations differ in terms of the level of concreteness. For example, some conceptualizations mention specific things that individuals can do to exert control (e.g., opt-out, decline cookies), while others are more abstract (e.g., control without explaining how). In order to increase the applicability of the conceptualizations, we decided to adopt an abstract conceptualization.

Third, similar with the literature on transparency, we observed three objects of control—namely, control over the collection, processing, and sharing of personal data. Therefore, we decide to adopt all three objects into our conceptualization. Based on the conceptualizations of the studies in Table 2, we propose two definitions of data control, one for actual control and one for perceived control:

> Actual data control: The extent to which individuals can start, stop, or maintain what personal data firms collect, process, or share (exchange) with the purpose of generating personalized communication.

> Perceived data control: The extent to which individuals think they can start, stop, or maintain what personal data firms collect, process, or share (exchange) with the purpose of generating personalized communication.

Finally, we found two factors in the studies' conceptualizations that influence the amount of control that individuals have, namely the ability to control (i.e., the skills and knowledge one has to exert control) and the desire to control (i.e., one's motivation to exert control). The inclusion of ability and desire in the conceptualization of control indicates the relevance to control. Because we believe they are distinct concepts we conceptualized ability and desire to control separately. Based on the definition provided by van Ooijen and Vrabec (2019) and other reviewed literature, we propose the following definitions:

> Ability to control: The extent to which an individual has the necessary knowledge and skills to start, stop or maintain firms to collect, process, or share (exchange) personal data with the purpose of generating personalized communication.

> Desire to control: the extent to which an individual has the motivation to start, stop, or maintain firms to collect, process, or share (exchange) personal data with the purpose of generating personalized communication.

3.3. Framework and Future Research Agenda

Based on the conceptualizations of transparency and control, we created the transparency–awareness–control (TAC) framework (Figure 2). Based on the framework, we provide concrete propositions to guide future research (Table 3). The framework differs based on the data collection mode because as Miyazaki (2008) noted, some data collection practices are more covert to consumers by nature (e.g., third-party cookies often

Table 1. Literature on personalization transparency.

Authors (year)	Published	Field	Country of research	Research method	Label	Conceptualization
Aguirre et al. (2015)	Yes	Marketing	The Netherlands	Experiment	Overt/covert data collection practices	"The strategies that firms employ to collect such data differ in the degree to which consumers are aware of how and when their information gets collected" (p. 36).
Awad & Krishnan (2006)	Yes	Marketing	US	Survey	Importance of information transparency	"Implicit in the collection of consumer information" (p. 14).
Boerman et al. (2017)	Yes	Communication science	N/A	Literature review	OBA transparency	"[Consumers want] openness and to be informed about the collection, usage, and sharing of personal data" (p. 367).
Dogruel (2019)	Yes	Communication science	Germany	Experiment	Use transparency	—
Harrysson & Olsson (2019)	No	Marketing	Sweden	Expert interviews	Transparency	—
Kim et al. (2019)	Yes	Marketing	US	Experiment	Ad transparency	"The disclosure of the ways in which firms collect and use consumer personal data to generate behaviorally targeted ads" (pp. 906–907).
Miyazaki (2008)	Yes	Marketing	US	Content analysis	Covertness of cookies	"Another concern regarding cookie placements is the covert nature of their usage. The placement of third-party cookies is often facilitated by the use of 'clear GIFs' that are only one pixel by one pixel in size, which essentially makes them invisible to the consumer" (p. 21).
Stevenson (2016)	No	Communication science	USA	Experiment	Transparency in online advertising personalization processes	"Transparency about some of the ways online ads are personalized for individuals appears" (p. 150).

Table 2. Literature on personalization control.

Authors	Published	Field	Country of research	Research Method/ Approach	Label	Conceptualization
Bamba & Barnes (2007)	Yes	Marketing	UK	Focus groups/ survey	Control over opt-in conditions	—
Beneke et al. (2010)	Yes	Marketing	South Africa	Survey	Consumer control	"the need consumers have to control the terms of the relationship with marketers with regards to what personal information is used, as well as the form and volume of advertising they receive" (p. 85).
Caravella (2007)	No	Marketing	N/A	Experiment	Intrusion control Disclosure control	"To control intrusion (into one's time, assets, or environment)" (p.15). "Strategic self-presentation" (p. 27).
Charters (2002)	Yes	Ethics	N/A	Caste study	Privacy as the right to control	"Privacy is conceived of as a right of an individual to determine to what extent, if at all, information about him or herself will be revealed to others" (p. 247).
Chung (2011)	No	Marketing	US	Survey	User control	"The extent to which consumers can determine the timing, content, and sequence of a transaction" (p. 22).
De Lima & Legge (2014)	Yes	Law	N/A	Comparative	Control (through consent)	"They can choose which cookies can be set on their computer and from whom and therefore the law should achieve its aim to provide individuals with a way to make informed decisions" (p. 68).
Eastin et al. (2016)	Yes	Communication science	US	Survey	Data control	"The degree that mobile users are concerned about their ability to have ownership of their personal information and control access to it" (p. 217).
Gironda & Korgaonkar (2018)	Yes	Marketing	US	Scenario-based survey	Perceived privacy control	"An individual's beliefs in his or her ability to manage the release and dissemination of personal information (Xu et al., 2011, p. 804)" (p. 68).
Harrysson & Olsson (2019)	No	Marketing	Sweden	Interviews	Control	"The ability to affect the dissemination and use of personal information that is collected during, or as a result of, marketing transactions, as well as control over unwanted telephone, mail, or personal intrusions in the consumer's home" (p. 15).

Table 2. (Cont.) Literature on personalization control.

Authors	Published	Field	Country of research	Research Method/ Approach	Label	Conceptualization
Humbani & Jordaan (2015)	Yes	Communication science	South Africa	Survey	Perceived control	—
Johnson et al. (2020)	Yes	Marketing	N/A	Case study	Consumer choice	"AdChoices enables consumer choice through a website that allows consumers to opt out of behaviorally targeted advertising. Consumers who opt out still see ads, just not ads that are targeted based on their previous browsing behavior" (p. 33).
Joo (2018)	No	Information and media	US	Experiment	User control over the information sharing	"The ability of the user to control the information stream" (p. 27).
Ketelaar & van Balen (2018)	Yes	Communication science	The Netherlands	Survey	Control	"Control over what their devices are tracking, and how this tracking is performed, and they are receiving it" (p. 175).
Leenes & Kosta (2015)	Yes	Law	N/A	Comparative	User control	Control as choice regarding cookies
Midha (2012)	Yes	Communication/ psychology	US	Survey	Consumer Privacy empowerment	"A psychological construct related to the individual's perception of the extent to which he/she can control the distribution and use of his/her personally identifying information" (p. 200).
Milne & Rohm (2000)	Yes	Marketing	US	Survey	Control of information disclosure	"Disclosure of information pertains to the capturing and storing of consumer information in databases" (p. 239).
Miltgen & Smith (2019)	Yes	Information systems	France	Experiment	Privacy protective behavior	"Individuals' data disclosure decisions, such as withholding, and the phenomenon of individuals providing falsified data in privacy-related decisions" (p. 697).
Mpinganjira & Maduku (2019)	Yes	Marketing	South Africa	Survey	Perceived privacy control	"The ability of individuals to manage privacy tools and settings on their mobile phones in order to enhance their personal privacy" (p. 468).

Table 2. (Cont.) Literature on personalization control.

Authors	Published	Field	Country of research	Research Method/ Approach	Label	Conceptualization
Richard & Meuli (2013)	Yes	Marketing	New Zealand	Survey	Perceived behaviour control	"Individual's perception of the ease or difficulty of performing a specific task or action (Ajzen, 1991)" (p. 703).
Song et al. (2016)	Yes	Marketing	South Korea	Experiment	Consumers' control	"The features that grant consumers both access to their personal information and authority to determine how such information can be used for personalized services" (p. 93).
Stevenson (2016)	No	Communication	US	Experiment	Consumer data control	"The degree to which respondents believed they had the ability to control how marketers used their personal information to target them with ads online" (p. 283).
Strycharz, van Noort, Smit, et al. (2019)	Yes	Communication science	The Netherlands	Experiment	Protection (motivation)	"Desire to adjust the settings offered by advertising platforms so that they do not receive personalized ads (which also means that their data is not processed for this purpose)" (p. 4).
Tucker (2014)	Yes	Marketing	US	Field experiment	Perceived control over privacy	—
van Ooijen & Vrabec (2019)	Yes	Communication science/law	N/A	Comparative	Individual control	"The extent to which an individual is consciously aware of a situation and has the conscious intention and the ability to start, stop, or maintain a situation" (p. 93).
Zarouali et al. (2018)	Yes	Communication science	Belgium	Experiment	Privacy control salience	"The extent to which an individual is consciously aware of a situation and has the conscious intention and the ability to start, stop, or maintain a situation" (p. 3).
					Perceived control	"Perceived control can be defined as the degree to which an individual views an event as within their control. In this study, it refers to whether consumers feel they have control over managing their privacy settings on a SNS" (p. 4).

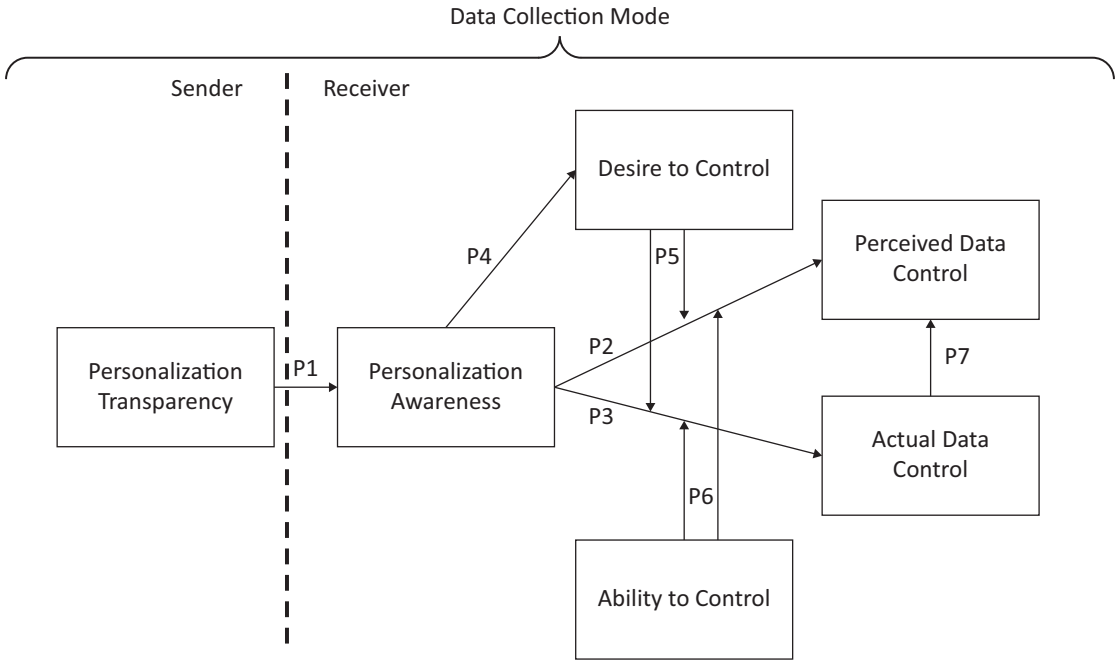


Figure 2. Transparency–awareness–control framework.

facilitated by the use of "clear GIFs" that are virtually invisible to individuals) and thus require transparency for raising awareness, while others have a less covert nature and require action from the individual (e.g., the proactive sharing of personal data such as an email address). We will explain the framework by means of four examples derived from the reviewed research. Furthermore, we will give different examples for the three objects of transparency and control (i.e., collection, processing, and sharing of data).

3.3.1. Example 1: High Transparency, High control

High transparency involves disclosure of data collection, processing, or sharing by the sender. Examples of high transparency from the reviewed literature regarding data collection include: detailed explanations of how personal data is collected and how long it will be stored (Song et al., 2016) as well as disclosures of covert data collection methods such as cookies, providing information on what they are and what data they collect (Miyazaki,

Table 3. TAC framework propositions.

P	Proposition
1	Transparency about data collection, processing, or sharing is a condition for individual awareness of such practices. The higher the degree of transparency, the higher the awareness.
2	Personalization awareness is a condition for perceived control over data for personalization. The higher the degree of awareness, the more likely that individuals have perceptions of control. Individuals need to be aware of data collection, processing, and sharing to perceive control.
3	Personalization awareness among individuals is a condition for having actual control over data for personalization. The higher the degree of awareness, the more likely that individuals will have control. When individuals are not aware of their data being collected, processed, or shared, it is not possible for them to control these actions.
4	Personalization awareness is a condition for having the desire to control data collection and personalization processes. Only individuals who are aware that their data is collected, processed, or shared can have the desire to control these processes.
5	The relationship between personalization awareness and (perceived and actual) control depends on the desire to control. Only with sufficient levels of desire to control, aware individuals will be able to exert some control.
6	The relationship between personalization awareness and (perceived and actual) control depends on the ability to control. Only with sufficient levels of ability (skills and knowledge) to control, aware individuals will be able to exert some control.
7	Higher actual control is more likely to lead to more perceived control.

2008). Examples of personalization processes include a high level of disclosure on data used to personalize the message (e.g., types of behavioral data or location data used and functions such as "Why am I seeing this ad?" offered by senders; see Dogruel, 2019; Kim et al., 2019). Regarding sharing, disclosures involve information about specific sources of data (e.g., sources of behavioral and location data used for advertising; Dogruel, 2019) and information on third-parties with whom the data will be shared (as required, for example, by the GDPR).

High actual control involves the possibility for individuals to act and is usually preceded with high transparency. The reviewed literature includes opt-out functions from data sharing with websites and apps (Joo, 2018). From the individual perspective, such control can also involve providing false information to the data collector (Miltgen & Smith, 2019). Regarding the personalization process, it includes privacy control menus that allow individuals to opt-out from processing for personalization (meaning not seeing personalized ads; see Strycharz, van Noort, Smit, et al., 2019; Zarouali et al., 2018). Finally, regarding data sharing, the literature proposes privacy settings that allow individuals to opt-out from third parties accessing their personal information (Tucker, 2014).

3.3.2. Example 2: High Transparency, Low Control

While high transparency may contribute to higher awareness among individuals, it does not automatically imply higher control. In cases of high transparency and low actual control, the same transparency mechanisms are in place as described above, but they either do not come with the possibility for action by the user (or have very limited options) to stop data collection (Zarouali et al., 2018), or they do not have opt-out signs in the app or web interface that would allow the user to impact the processing for personalization (Joo, 2018).

While it is not common to display disclosures but provide no opt-out/privacy control features (actual control), providing such features does not imply high *perceived* control (Zarouali et al., 2018). Perceived control may be impeded by lack of awareness, no desire to control personalization processes, or lack of ability to exercise control. An example of high transparency and low perceived control is data collection through cookies. Such data collection has to be disclosed on websites, but this disclosure does not foster the perception of control among individuals (Miyazaki, 2008).

3.3.3. Example 3: Low Transparency, Low Control

Low transparency regarding data collection involves not specifying what or how data are collected (Miyazaki, 2008). Regarding processing, it involves not disclosing what data have been used for personalization (Dogruel, 2019; Kim et al., 2019) and regarding sharing, how data have been obtained from other parties or if they will be shared with third parties. As Miyazaki (2008) argues, covert data collection techniques such as the use of third-party cookies facilitated by pixel-sized images on websites are practically invisible to individuals. Such techniques have been called non-obvious by the Federal Trade Commission (2000). For these non-obvious data collection techniques, with no transparency, individual awareness is difficult to achieve. As a result, individuals are not able to exercise control over such practices. Therefore, low transparency about non-obvious practices is often the object of regulations (such as the e-Privacy directive that obligates transparency about cookies in the EU).

3.3.4. Example 4: Low Transparency, High Control

This category does not exist as transparency is a condition for control. When it is not transparent how data are collected, processed, or shared, individuals are not aware of these practices (e.g., the use of third-party cookies are not disclosed on a website), and therefore they are not able to stop such practices.

4. Conclusions

The growing importance of the transparency and control paradigm for personalized communication has led to increased attention from legislators and academics. This calls for clear definitions of the concepts involved to increase validity and facilitate future research, which was the aim of this study.

By means of a systematic literature review, we analyzed 31 articles on personalization transparency and control. The concept of transparency has been around for a longer time because it has been relevant to other communication strategies that are more covert, such as native advertising (Wojdynski & Evans, 2020). However, control seems to be a phenomenon specific to communication strategies that rely on personal data that has been receiving increasing attention in the recent years. In our literature review we specifically focused on the conceptualization of transparency and control for personalized communication. This led us to four conclusions.

First, the literature review confirmed that there is no common definition of either transparency or control, which highlights the need for a shared understanding of these concepts. While studies included in the review have different focuses of research (different types of advertising including online advertising in general, online behavioral advertising, and mobile advertising) and different control mechanisms related to advertising (e.g., mobile phone settings, privacy protection, advertising opt-out mechanism), they all investigate different aspects of transparency and control related to personalized marketing communication. Hence, based on the reviewed literature, we formulated definitions of both personalization transparency and control. The different focuses of the studies included could have contributed

to the diversity of conceptualizations found in the literature. However, even while focusing on one specific object and the papers that study that object (e.g., transparency about data collection for advertising through tracking cookies), we observe differences. Moreover, we find that many studies did not include any definitions of their terms. Our study, therefore, contributes to the literature by synthesizing different definitions, analyzing them, and proposing one definition to help research on this topic in the field of personalized marketing communication move forward. In addition, we made a distinction between actual and perceived control, which is important because previous research on personalization shows that they are different concepts. Future research could examine whether they have different predictive powers.

Second, we observed that the concepts of transparency and control were often used interchangeably in the literature. Although we believe that these concepts are related (see Figure 2), we argue that they are separate constructs. We also observed that other concepts were often entangled with understandings of transparency and control: Awareness, for example, was often integrated in the transparency conceptualization. However, we argue that transparency is about information disclosure from the sender side, while awareness concerns the extent to which individuals are conscious of the practices from the receiver's side. This is an important distinction for future research to take into consideration. Also, we found that ability and desire to control were integrated into definitions of control (Figure 2).

Third, we observed that the objects of transparency and control differed between conceptualizations. We found three objects of transparency and control, namely collection, processing, and sharing of data. We believe it is important to acknowledge the different objects because what information is disclosed or what individuals can do to exert control differs for each object.

Finally, we introduced the TAC framework to visualize the relationship between the concepts discussed, providing concrete propositions to guide future research (Table 3). Note that although we argue that transparency and control are positively related, it does not mean that more transparency automatically leads to more control. As shown in the TAC framework, transparency provided by the sender first needs to increase awareness in the receiver before it could lead to more control. In addition, we argue that the ability and desire to control are boundary conditions for the relationship between awareness and control. Future research should empirically test the propositions of the framework to validate the claims. In addition to this theoretical contribution, the TAC framework has important implications for privacy regulations, since transparency regarding data collection and processing practices is a core issue in current regulatory approaches. In fact, both the GDPR and the California Consumer Privacy Act, which aim to strengthen individuals' rights regarding control over their personal data, portray transparency as the main data protection mechanisms in online data collection processes by requiring companies to be more transparent about their data collection practices (Strycharz et al., 2020; van Ooijen & Vrabec, 2019).

Furthermore, the TAC framework, while developed specifically in the context of personalized marketing communication, can be applied and tested to other areas of personalization research. Personal data collection and algorithmic processing that enable personalization can also be used in health communication (e.g., personalized healthcare; Dzau & Ginsburg, 2016), political communication (e.g., political microtargeting; Zuiderveen Borgesius et al., 2018) or journalism (e.g., news recommendations; Thurman et al., 2019) and lead to the same questions about transparency, individual awareness, and control. The TAC framework can therefore be used to further explore consumer empowerment in these areas.

In sum, this study provided definitions of personalization transparency and control for the use of personalized communication, as well as for the related concepts of awareness, ability, and desire to control. While the concepts are not new to the literature, the increasing use and importance of data for personalized marketing communication, computational advertising, and other forms of algorithmic communication make them important concepts of interest. Increased comprehension of the transparency and control paradigm gives us a chance to better understand how data collection practices work, what effects they have on individuals, and what implications this may have for industry practices and privacy regulations.

Conflict of Interests

The authors declare no conflict of interests.

Supplementary Material

Supplementary material for this article is available online here: https://osf.io/f3ndw/?view_only=c66708fdf27741d0894b8f2620be49a7.

References

Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing*, *91*(1), 34–49.

Altman, I. (1975). *The environment and social behavior*. Brooks/Cole.

Awad, N. F., & Krishnan, M. S. (2006). The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS quarterly*, *30*(1), 13–28.

Bamba, F., & Barnes, S. J. (2007). SMS advertising, permission and the consumer: A study. *Business Process*

Management Journal, *13*(6), 815–829.

Beneke, J., Cumming, G., Stevens, A., & Versfeld, M. (2010). Influences on attitude toward mobile text message advertisements: An investigation of South African youth. *International Journal of Mobile Marketing*, *5*(1), 77–97.

Boerman, S. C., Kruikemeier, S., & Zuiderveen Borgesius, F. J. (2017). Online behavioral advertising: A literature review and research agenda. *Journal of Advertising*, *46*(3), 363–376.

Caravella, M. N. (2007). *Privacy, strategic information disclosure and new customer acquisition: Implications for customer relationship management* [Unpublished doctoral dissertation]. Harvard University.

Charters, D. (2002). Electronic monitoring and privacy issues in business-marketing: The ethics of the DoubleClick experience. *Journal of Business Ethics*, *35*, 234–254.

Chung, T.-L. (2011). *Consumers' adoption of mobile coupon: A value-based adoption model* [Unpublished doctoral dissertation]. Purdue University.

De Keyzer, F., Dens, N., & De Pelsmacker, P. (2015). Is this for me? How consumers respond to personalized advertising on social network sites. *Journal of Interactive Advertising*, *15*(2), 124–134.

De Lima, D., & Legge, A. (2014). The European Union's approach to online behavioural advertising: Protecting individuals or restricting business? *Computer Law & Security Review*, *30*(1), 67–74.

Deloitte. (2018). *A new era for privacy GDPR six months on*. https://www2.deloitte.com/uk/en/pages/risk/articles/gdpr-six-months-on.html

Dogruel, L. (2019). Too much information!? Examining the impact of different levels of transparency on consumers' evaluations of targeted advertising. *Communication Research Reports*, *36*(5), 383–392.

Dzau, V. J., & Ginsburg, G. S. (2016). Realizing the full potential of precision medicine in health and health care. *Jama*, *316*(16), 1659–1660.

Eastin, M. S., Brinson, N. H., Doorey, A., & Wilcox, G. (2016). Living in a big data world: Predicting mobile commerce activity through privacy concerns. *Computers in Human Behavior*, *58*, 214–220.

Federal Trade Commission. (2000). *Privacy online: Fair information practices in the electronic marketplace—A Federal Trade Commission report to congress*. https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-fair-information-practices-electronic-marketplace-federal-trade-commission-report/privacy2000text.pdf

Gironda, J. T., & Korgaonkar, P. K. (2018). iSpy? Tailored versus invasive ads and consumers' perceptions of personalized advertising. *Electronic Commerce Research and Applications*, *29*, 64-77.

Harrysson, A., & Olsson, J. (2019). *Personalization paradox: The wish to be remembered and the right to be forgotten* [Unpublished Master's thesis]. Uppsala University.

Humbani, M., & Jordaan, Y. (2015). The role played by gender, household income and age in factors contributing to consumers' attitudes towards short message service advertisements. *Communicare*, *34*(1), 27–48.

Johnson, G. A., Driver, S. K., & Du, S. (2020). Consumer privacy choice in online advertisement: Who opts out and at what cost to industry? *Marketing Science*, *39*(1), 33–51.

Joo, E. (2018). S*ponsor visibility, customization, and user control in the era of interactive technology: Effects on causal attribution of sponsor's motives, sponsor attitudes, and credibility in the context of sponsored mobile health-related apps* [Unpublished doctoral dissertation]. Michigan State University.

Ketelaar, P. E., & van Balen, M. (2018). The smartphone as your follower: The role of smartphone literacy in the relation between privacy concerns, attitude and behaviour towards phone=embedded tracking. *Computers in Human Behavior*, *78*, 174–182.

Kim, T., Barasz, K., & John, L. K. (2019). Why am I seeing this ad? The effect of ad transparency on ad effectiveness. *Journal of Consumer Research*, *45*(5), 906–932.

Kramer, T., Spolter-Weisfeld, S., & Thakkar, M. (2007). The effect of cultural orientation on consumer responses to personalization. *Marketing Science*, *26*(2), 246–258.

Leenes, R., & Kosta, E. (2015). Taming the cookie monster with Dutch law: A tale of regulatory failure. *Computer Law & Security Review*, *31*(3), 317–335.

Li, H., Yu, L., & He, W. (2019). The impact of GDPR on global technology development. *Journal of Global Information Technology Management*, *22*(1), 1–6.

Maslowska, E., Smit, E. G., & Van den Putte, B. (2016). It is all in the name: A study of consumers' responses to personalized communication. *Journal of Interactive Advertising*, *16*(1), 74–85.

Midha, V. (2012). Impact of consumer empowerment on online trust: An examination across genders. *Decision Support Systems*, *54*(1), 198–205.

Milne, G. R., & Rohm, A. J. (2000). Consumer privacy and name removal across direct marketing channels: Exploring opt-in and opt-out alternatives. *Journal of Public Policy & Marketing*, *19*(2), 238–249.

Miltgen, C. L., & Smith, H. J. (2019). Falsifying and withholding: Exploring individuals' contextual privacy-related decision-making. *Information & Management*, *56*(5), 696–717.

Miyazaki, A. D. (2008). Online privacy and the disclosure of cookie use: Effects on consumer trust and anticipated patronage. *Journal of Public Policy & Marketing*, *27*(1), 19–33.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS med*, *6*(7), Article e1000097.

Mpinganjira, M., & Maduku, D. K. (2019). Ethics of mobile behavioral advertising: Antecedents and outcomes

of perceived ethical value of advertised brands. *Journal of Business Research*, *95*, 464–478.

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan: A web and mobile app for systematic reviews. *Systematic Reviews*, *5*, Article 210.

Richard, J. E., & Meuli, P. G. (2013). Exploring and modelling digital natives' intention to use permission-based location-aware mobile advertising. *Journal of Marketing Management*, *29*(5/6), 698–719.

Song, J. H., Kim, H. Y., Kim, S., Lee, S. W., & Lee, J. H. (2016). Effects of personalized e-mail messages on privacy risk: Moderating roles of control and intimacy. *Marketing Letters*, *27*(1), 89–101.

Stevenson, D. M. (2016). *Data, trust, and transparency in personalized advertising* [Unpublished doctoral dissertation]. University of Michigan.

Strycharz, J., Ausloos, J., & Helberger, N. (2020). Data protection or data frustration? Individual perceptions and attitudes towards the GDPR. *European Data Protection Law Review*, *6*(3), 407–421.

Strycharz, J., van Noort, G., Helberger, N., & Smit, E. (2019). Contrasting perspectives: Practitioner's viewpoint on personalised marketing communication. *European Journal of Marketing*, *53*(4), 635–660.

Strycharz, J., van Noort, G., Smit, E., & Helberger, N. (2019). Protective behavior against personalized ads: Motivation to turn personalization off. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *13*(2), Article 1.

Thurman, N., Möller, J., Helberger, N., & Trilling, D. (2019). My friends, editors, algorithms, and I: Examining audience attitudes to news selection. *Digital Journalism*, *7*(4), 447–469.

Tucker, C. E. (2014). Social networks, personalized advertising, and privacy controls. *Journal of Marketing Research*, *51*(5), 546–562.

van Ooijen, I., & Vrabec, H. U. (2019). Does the GDPR enhance consumers' control over personal data? An Analysis from a behavioural perspective. *Journal of Consumer Policy*, *42*(1), 91–107.

Vesanen, J., & Raulas, M. (2006). Building bridges for personalization: a process model for marketing. *Journal of Interactive Marketing*, *20*(1), 5–20.

Wojdynski, B. W., & Evans, N. J. (2020). The covert advertising recognition and effects (CARE) model: Processes of persuasion in native advertising and other masked formats. *International Journal of Advertising*, *39*(1), 4–31.

Yun, J. T., Segijn, C. M., Pearson, S., Malthouse, E. C., Konstan, J. A., & Shankar, V. (2020). Challenges and future direction of computational advertising measurement systems. *Journal of Advertising*, *49*(4), 446–458.

Zarouali, B., Poels, K., Ponnet, K., & Walrave, M. (2018). Everything under control? Privacy control salience influences both critical processing and perceived persuasiveness of targeted advertising among adolescents. *Cyberpsychology: Journal of Psychosocial Research and Cyberspace*, *12*(1), Article 5.

Zuiderveen Borgesius, F., Möller, J., Kruikemeier, S., Ó Fathaigh, R., Irion, K., Dobber, T., Bodó, B., & de Vreese, C. H. (2018). Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review*, *14*(1), 82–96.

About the Authors

Claire M. Segijn (PhD) is an assistant professor of advertising at the Hubbard School of Journalism and Mass Communication, University of Minnesota, Twin Cities. Her research interests are in information processing and effects of the usage of multiple media at the same time (e.g., multiscreening, synced advertising).

Joanna Strycharz (PhD) is an assistant professor of persuasive communication at the Amsterdam School of Communication Research, University of Amsterdam. Her research focuses on personalized advertising and its impact on consumers, their privacy, and consumer empowerment.

Amy Riegelman is a social sciences librarian at the University of Minnesota, Twin Cities, where her work includes reference and instruction responsibilities as well as co-chairing a systematic review service. Amy is also a member of the Campbell Collaboration's Information Retrieval Methods Group.

Cody Hennesy is the journalism and digital media librarian at the University of Minnesota, Twin Cities, where he develops services and support for text and data mining research and the computational social sciences.

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 4, Pages 134–144
https://doi.org/10.17645/mac.v9i3.4090

Article

Investigating Algorithmic Misconceptions in a Media Context: Source of a New Digital Divide?

Brahim Zarouali [1,*], Natali Helberger [2] and Claes H. de Vreese [1]

[1] Amsterdam School of Communication Research, University of Amsterdam, The Netherlands;
E-Mails: b.zarouali@uva.nl (B.Z.), c.h.devreese@uva.nl (C.H.d.V.)
[2] Institute for Information Law, University of Amsterdam, The Netherlands; E-Mail: n.helberger@uva.nl

* Corresponding author

Submitted: 29 January 2021 | Accepted: 19 April 2021 | Published: 18 November 2021

Abstract
Algorithms are widely used in our data-driven media landscape. Many misconceptions have arisen about how these algorithms work and what they can do. In this study, we conducted a large representative survey (*N* = 2,106) in the Netherlands to explore algorithmic misconceptions. Results showed that a significant part of the general population holds (multiple) misconceptions about algorithms in the media. We found that erroneous beliefs about algorithms are more common among (1) older people (vs. younger people), (2) lower-educated people (vs. higher-educated), and (3) women (vs. men). In addition, it was found that people who had no specific sources to inform themselves about algorithms, and those relying on their friends/family for information, were more likely to have algorithmic misconceptions. Conversely, media channels, school, and having one's own (online) experiences were found to be sources associated with having fewer algorithmic misconceptions. Theoretical implications are formulated in the context of algorithmic awareness and the digital divide. Finally, societal implications are discussed, such as the need for algorithmic literacy initiatives.

Keywords
algorithms; algorithmic awareness; digital divide; misconceptions; technology

Issue
This article is part of the issue "Algorithmic Systems in the Digital Society" edited by Sanne Kruikemeier (University of Amsterdam, The Netherlands), Sophie Boerman (University of Amsterdam, The Netherlands), and Nadine Bol (Tilburg University, The Netherlands).

1. Introduction

In our data-driven media landscape, algorithms play an increasingly important role in how online users use, navigate, and consume online information and communication (Beer, 2017; Lee, 2018; Ricci, 2015). For instance, recommendation algorithms allow online platforms and legacy media alike to make personalized recommendations based on people's profiles; content moderation algorithms are used to determine the ranking of the contents that are being shown to us; automated filtering algorithms allow us to detect instances of misinformation, harmful, or unlawful content; etc. Given their widespread use and impact on people's media and information consumption, having a proper sense of what algorithms are and are capable of doing is a necessary condition for digital citizenship. However, recent studies have indicated that a significant part of the population has limited knowledge about the algorithms used in online platforms (e.g., Facebook, Google, etc.), as well as misconceptions about how they work (Cotter & Reisdorf, 2020; e.g., Eslami et al., 2015; Rader & Gray, 2015).

Misconceptions refer to incorrect ideas formed as a result of unfounded concerns and erroneous beliefs. These ideas may be due to the increased hype about the promises of algorithms and machine learning, which has fueled a variety of false assumptions (de Saint Laurent, 2018). These misconceptions can raise some

serious issues, such as a highly polluted public debate about algorithms (with many loud voices that contribute to a dystopian view of the future) and missing out on the full potential of algorithms for societal good (de Saint Laurent, 2018; Elish & boyd, 2018; First, 2018; Frank et al., 2017). Maybe even more seriously, misconceptions about the workings and consequences of algorithms can contribute to major societal problems, such as the spread of misinformation and deep fakes, data-driven manipulation and re-enforcing stereotypes, and inequalities and discrimination (Eubanks, 2017; Mohamed et al., 2020).

As an integral part of education, misconceptions must be addressed to avoid anxiety, fatalism, and distress about technological developments. Critical to this effort is knowing the extent to which these misconceptions have infiltrated our society, as well as from where they originate. Drawing on the theoretical tenets of algorithmic awareness, we present findings from a large representative survey (N = 2,106) in the Netherlands in which we explore the prevalence of various misconceptions about algorithms and their distribution among demographic groups, as well as mapping out the main information sources related to these misconceptions. In a concluding discussion section, we address the societal implications of the findings, as well as the theoretical contributions to the literature of algorithmic awareness, algorithmic accountability, and the digital divide. Finally, we address how to overcome these misconceptions and empower people to become informed citizens in the age of information technologies.

2. Literature Review

2.1. Algorithmic Awareness

Algorithms can be described as codified procedures for transforming vast amounts of input data into the desired output, based on specified calculations (Gillespie, 2014). From a technical perspective, algorithms are very complicated entities and are part of a larger and equally complex socio-technical infrastructure (Kitchin, 2017). In addition, they serve as an important competitive advantage for many companies (e.g., big tech platforms), which explains why such companies are very reluctant to expose their algorithmic codes to the outside world (Pasquale, 2015). Given this technical complexity and increased lack of transparency, it is very hard to be exactly aware of what algorithms are doing (Cotter & Reisdorf, 2020). Adding to the lack of transparency is the fact that many algorithmic applications strive to provide a seamless user experience, optimizing for invisibility and normalization over time. However, despite these constraints, people can still develop—to some extent—a conceptual awareness about algorithms and their effects (Cotter & Reisdorf, 2020; Eslami et al., 2015; Zarouali et al., 2021). In an online media context, conceptual awareness would mean that users know that there is a dynamic system in place that can personalize and customize the information that they see or hear, based on a corpus of data composed of digital traces (Hargittai et al., 2020; Zarouali et al., 2021).

To date, only a limited body of research has focused on people's algorithmic awareness (Hargittai et al., 2020). These studies focus on specific mediated contexts, such as algorithmic curation in social media newsfeeds (Eslami et al., 2015; Rader & Gray, 2015; Zarouali et al., 2021), online search (Cotter & Reisdorf, 2020), and news platforms (Powers, 2017). Although the results of these studies are not entirely univocal, we can conclude that the findings so far show that people are characterized by a lack of awareness of algorithmic content curation (e.g., Eslami et al., 2015; Powers, 2017). In addition to this, studies have also shown that there is a strong variation in algorithmic awareness among certain parts of the population (Hargittai et al., 2020; Rader & Gray, 2015). This has been referred to as the "algorithmic knowledge gap," which might contribute to a new digital divide, and thus merits further investigation (Cotter & Reisdorf, 2020). Therefore, this study aims to provide a more solid empirical ground by focusing on the prevalence of algorithmic (mis)conceptions, and discussing these findings in the light of digital divides.

The importance of providing more solid empirical insights into algorithmic awareness and algorithmic misconceptions is also important from the perspective of algorithmic accountability and the construction of digital citizenship. Social media users, as digital citizens, have an important role in critically scrutinizing algorithms and the services they are enabling, but also in challenging or resisting algorithms that conflict with users' rights and interests (Hintze et al., 2019). Algorithmic awareness becomes a precondition for algorithmic accountability. Meijer and Grimmelikhuijsen (2021) describe algorithmic accountability as "the justification of the organizational usage of an algorithm and explanations for its outcomes to an accountability forum that can ask questions, pass judgement, and impose consequences" (p. 60). In order to be able to ask the necessary questions and hold controllers of algorithms accountable, users need to possess what Koene et al. (2019) call "algorithmic literacy," along with the ability to act and exercise agency. It is not difficult to see how algorithmic misconceptions and misleading imaginaries inhibit the ability of users to exercise critical citizenship and thereby hold algorithmic power to account. This also explains why so many public policy measures are directed at increasing algorithmic awareness through transparency and media literacy initiatives (see European Commission, 2020a; highlighting the importance of algorithmic awareness to enhancing the ability of individuals to be aware of their rights and know how to act upon them, see Council of Europe, 2020, p. 8). Empirical insights into algorithmic misconceptions, therefore, contribute to both the literature on algorithmic accountability, as well as law and policymaking.

2.2. Algorithmic Misconceptions

When it comes to new technologies, history tells us that their introduction most often goes hand-in-hand with a broad range of projected hopes and fears, which gives rise to various myths and misconceptions (Natale & Ballatore, 2020). These misconceptions and myths should be seen as dynamic constructs that give meaning and represent an important part of the collective mentality of (a group of) people (Mosco, 2004). Based on a thorough literature review in the area of (machine learning) algorithms, we identified five important misconceptions. This list might not be exhaustive, but it does certainly comprise the major misconceptions highlighted in recent academic work (e.g., de Saint Laurent, 2018; Emmert-Streib et al., 2020; First, 2018; Roffel & Evans, 2018).

The first major misconception refers to the idea that algorithms are completely independent from human influence. Algorithms are designed by humans to automate certain tasks in a highly optimized way (i.e., being much more efficient than humans) (Gillespie, 2014; Lee, 2018). Importantly, the degree of automation in algorithms can vary: Certain algorithms allow some degree of human involvement, whereas others take fully automated decisions and keep humans completely out of the loop (Diakopoulos, 2019; Parasuraman et al., 2000). In reality, many algorithms do not operate fully independently, but are closely monitored by human beings, they often rely on human-generated input and data and are the result of models and metrics developed by humans (Fry, 2019). That is, they are often used to improve a system's performance, without necessarily reducing human involvement (Shneiderman et al., 2018). In addition, algorithms are constantly being tweaked, tuned, re-written, repaired, or deleted; as such, they are not fully independent technical objects (Kitchin, 2014; Seaver, 2018) and co-evolve in their interactions with humans.

The second misconception is the idea that algorithms are operating neutrally and objectively, and thus, are free of bias. Indeed, algorithms "as such" are unbiased because they are inert and meaningless systems; the bias occurs when algorithms are paired with (human-generated) databases or models that determine their functioning (Gillespie, 2014). Algorithms can not only display the biases of those who make and operate them, but potentially also the values and (commercial) preferences of the companies that provide them, or the technical infrastructures in which they operate (Ananny & Crawford, 2018; Gillespie, 2010). So, in reality, all operating algorithms can have some kind of bias (de Saint Laurent, 2018). For instance, subtle human biases (e.g., ideologies, prejudices, and inequalities) can slip into the data inputted, the training of the data, and the algorithmic operation (Amoore & Piotukh, 2015; Beckett, 2019; van Dijck et al., 2018). On a technical level, research indicated that biases related to data representativeness and sampling can also occur (Eubanks, 2017; Fry, 2019; Hargittai, 2020). Therefore, algorithmic biases should be seen as reflections of more fundamental societal (and technical) biases (Bucher, 2018).

A third misconception entails that algorithms can replace the high-level critical reasoning and human thought. To illustrate this misconception, take the example of neural networks algorithms. These algorithms can learn to make quicker and more accurate decisions based on experience: The more examples they are exposed to, the more accurate they become (Chesney & Citron, 2019; Dack, 2019). That is why people came to believe that algorithms mimic the decision-making processes in our human brains. However, as argued by Emmert-Streib et al. (2020), assuming that these algorithmic models perform just like human brains is not plausible nor realistic. In fact, a major downside of algorithms is rooted in their inability to make critical decisions, cope with unanticipated scenarios, make subjective value-based judgments, and display creativity (Diakopoulos, 2019; Shneiderman et al., 2018). Therefore, scholars argue that algorithms cannot (yet) reason in the same way as humans (Roffel & Evans, 2018).

A fourth misconception is that algorithms can solve every problem in society. In the past decade, many people came to believe that every societal problem or difficulty has a solution based on technology, which has been referred to as "technological solutionism" (Morozov, 2014). This trend toward finding quick technological fixes is combined with an at times somewhat naive trust in the infallibility of technology. The reality is far more nuanced: Algorithms are usually used for solving very specific (rule-based) tasks or problems (Fry, 2019; Roffel & Evans, 2018) as they excel at executing routine, tedious, and error-prone tasks highly efficiently, tirelessly, and consistently (Diakopoulos, 2019; Shneiderman et al., 2018). Therefore, scholars have cautioned against the idea of considering algorithms as silver bullets that will solve everything (Morozov, 2014; Roffel & Evans, 2018). Rather, they should be seen as tools that have become very efficient in solving narrow problems.

The fifth misconception is that algorithms will replace human workers in the media sector. A good example would be automation in the newsroom: Some people foresee the elimination of jobs, with human journalists being replaced by algorithms. In reality, algorithms are unlikely to replace journalists, but instead, are often being used to design efficient and effective systems that support workflows and reporting (Beckett, 2019; Diakopoulos, 2019). As with other technological revolutions, it is possible that certain tasks or even professions in the newsroom may become obsolete, but at the same time, the introduction of algorithmic processes also introduces entirely new roles and tasks (Ferrer-Conill & Tandoc, 2018). Also from the perspective of managerial staff, editors, and journalists, there is no real immediate concern of replaceability in the newsroom; for them, algorithms represent a supplementary (useful) toolkit (Schapals & Porlezza, 2020). As this illustration shows,

algorithms should be seen as tools that can support rather than replace human decision-making (Fry, 2019).

2.3. Research Questions

Many of the misconceptions discussed above may have (deeply) infiltrated our society. On a societal level, this could lead to some serious concerns, such as a polluted (and myth-based) public debate, but also misjudging the role that humans have, e.g., in the process of spreading misinformation or contributing to algorithmic biases. In addition, algorithmic misconceptions can also seriously undermine the full potential of algorithms in our society (de Saint Laurent, 2018), since governments and other institutions might have a misguided lack of trust (as a result of misconceptions) in the use of algorithmic solutions for societal problems. On the level of the individual user, one of the biggest concerns is that these misconceptions might not be universally distributed in the population, and thus, that they might be overrepresented in certain (more vulnerable) parts of the population, resulting in new forms of digital exclusion. That is, unequal skills and knowledge (including misconceptions) can result in new forms of digital divide, e.g., the "algorithmic divide" (Carmi & Yates, 2020). In addition, when (certain groups of) people have numerous misconceptions, they might develop a distorted and ill-informed mindset about how algorithms work, which could undermine their ability to make correct and rational judgments about the information that algorithms present to them online and misjudge their own role in the process.

Therefore, this study aims to investigate the prevalence and the main sources of these algorithmic misconceptions in the population. Broadly speaking, when it comes to ICT knowledge and digital literacy skills, many studies have already acknowledged the importance of individual differences by looking at demographic characteristics (e.g., Hargittai, 2010; Schreurs et al., 2017; van Dijk & van Deursen, 2014). In the context of algorithms, a recent study found a relationship between algorithmic knowledge and socioeconomic background, indicating a worrisome (digital) knowledge inequality (Cotter & Reisdorf, 2020). Therefore, in this study, we explore the prevalence and differences in algorithmic misconceptions among certain demographic groups (age, gender, and education), as well as investigate whether these demographic variables can predict algorithmic misperceptions. In addition, we also aim to look into the main information sources of algorithmic misperceptions. In particular, knowing the (perceived) sources of misconceptions is essential to be able to refute them effectively (Menz et al., 2021). That is, to overcome the persistence of algorithmic misconception, we must have an idea of the main sources associated with these misconceptions. Therefore, we explore the main information sources that people attribute to their misconceptions and test whether there is a relationship between particular sources and the prevalence of algorithmic misconceptions. Based on this, we suggest the following research questions: RQ1) How prevalent are algorithmic misconceptions in the population, and how are they distributed among socio-demographic characteristics (age, gender, and education)?; RQ2) What are the main information sources that people attribute to their algorithmic (mis)conceptions?; and RQ3) Are these demographic characteristics (gender, age, and education) and information sources significant predictors of algorithmic misconceptions (is there a significant association)?

3. Methods

3.1. Sample

We used data from a larger panel wave study which was distributed among a representative sample of the Dutch population. The larger panel study focused on the societal impact of communication technologies and algorithms. Representativeness was achieved based on age, gender, education, and region. The fieldwork was carried out by a research company. The total sample size was N = 2,106. To achieve this net sample size, a gross sample of 6,000 people was initially contacted, which means that the overall response rate was 35%. The data collection took place from July 19 to August 9, 2019 (21 days). The respondents had a mean age of 54.18 (SD = 15.59 years), and 48% of them were women. All respondents successfully completing the survey received an incentive (bonus points) from the research company. A demographic overview of the sample is presented in Table 1.

3.2. Measures

We measured algorithmic misconceptions by presenting respondents with five true/false statements. To make these statements less abstract, they were preceded by a short explanation about algorithms: "The following questions are about your awareness of the use of algorithms in the media (e.g., algorithms that recommend relevant content to you)." This short introduction was followed by five statements that we discussed above, in the literature review. More precisely, we asked that respondents "indicate whether you believe the following statements about algorithms in the media are true or false," with the following items: (1) Algorithms are completely independent, without human influence; (2) algorithms always operate neutrally and objectively, and thus, are free of bias; (3) algorithms can solve every problem in society; (4) algorithms have the same level of critical reasoning and intelligence as humans; (5) algorithms will replace humans workers in the media sector. All items were misconceptions, i.e., responses needed to be "false" to be correct. Thus, a respondent answering "true" on an item (which is an incorrect answer), is considered as someone holding that specific misconception. We adopted this format, in which all correct responses are "false," from the study of Taylor and Kowalski (2004). In addition,

Table 1. Socio-demographic characteristics of the representative Dutch sample.

	Percentage (%)	Frequency (*N*)
Age categories (M_{age} = 54.18, SD_{age} = 15.59)		
18–34 years	14.62	308
35–54 years	33.67	709
55+ years	51.71	1,089
Gender		
Female	47.96	1,010
Male	52.04	1,096
Education		
Low	30.82	649
Moderate	50.47	1,063
High	18.71	394
Region		
North	11.16	235
East	21.37	450
South	24.17	509
West	28.40	598
Three large cities (Amsterdam, Rotterdam, and The Hague)	14.91	314

we also created an aggregated measure for algorithmic misconceptions (for the purpose of multivariate analyses). To do this, all incorrect answers were coded with 1, and summed to compute an index with scores ranging from 0 to 5. A higher score on this index means that person holds more algorithmic misconceptions (*M* = 2.31, *SD* = 1.75).

To investigate the information sources of algorithmic misconceptions, we used a similar approach as Menz et al. (2021). We asked respondents to indicate what sources contributed to their acquisition of algorithmic information. They were given multiple answer options: (1) own experiences, (2) media (offline and online), (3) school, (4) friends and/or family, (5) no information source, (6) other. We manually checked all responses of the "other," and many of the responses could easily be categorized in the five other options, which we did. Therefore, this category was not included in the analyses.

In terms of socio-demographic information, we measured respondents' age, gender, education level, and geographical region. Age was measured on a continuous level (for the multivariate analyses) and was also recoded into a categorical variable consisting of three age groups. Gender was measured based on two response choices: female and male. Education was measured based on a detailed list of seven categories (tailored to the Dutch education system). This categorization can be re-coded into three education levels: low (no education or primary education), moderate (secondary education), and high (post-secondary and higher education). The variable region was based on Nielsen's regional division of the Netherlands, which is the gold standard in market research. We provide the (sample) descriptives of this variable, but we do not include them in the statistical analyses.

4. Results

4.1. Prevalence and Distribution of Algorithmic Misconceptions (RQ1)

Table 2 presents the general prevalence of all five algorithmic misconceptions among all respondents (first row), as well as a more narrowed overview of the prevalence in specific demographic groups. The numbers in the table refer to the percentage of people that gave an incorrect answer on a misconception item (see measures), which means the proportion of respondents holding that misconception. Looking at the first row, i.e., the general prevalence numbers, we see that the first, second, fourth, and fifth misperception are supported by more than half of the respondents. The proportion related to the third misperception is slightly lower (43.64%). In particular, misconception five, i.e., that algorithms will replace human workers, is the most widespread among the respondents (63.96%). In terms of age groups, Table 1 illustrates that age is significantly associated with the prevalence of all algorithmic misconceptions (see χ^2 tests), except for misconception five. The Z-tests (indicated by means of superscripts) provide a more detailed overview, specifying which proportions differ from each other. Based on these tests, we see that older age groups have more algorithmic misconceptions than younger respondents. Gender was also found to be consistently associated with the prevalence of all misconceptions. More precisely, a higher proportion of women held algorithmic misconceptions than men. Finally, education level was also significantly associated with the extent to which people hold all five algorithmic misconceptions. That is, lower-educated respondents were more likely to have

Table 2. Percentage of people giving an incorrect answer on the misconception items.

	MC1	MC2	MC3	MC4	MC5
Total sample (%)	53.66	52.85	43.64	54.23	63.96
Age categories (%)					
18–34 years	44.81[a]	42.21[a]	39.61[a]	46.43[a]	60.06[a]
35–54 years	49.51[a]	43.86[a]	37.52[a]	48.52[a]	63.05[a]
55+ years	58.86[b]	61.71[b]	48.76[b]	60.15[b]	65.66[a]
χ^2-test	26.48***	71.64***	24.45***	32.23***	3.64 (*ns*)
Gender (%)					
Male	50.27[a]	45.89[a]	35.31[a]	49.18[a]	60.04[a]
Female	57.33[b]	60.40[b]	52.67[b]	59.70[b]	68.22[b]
χ^2-test	10.16**	44.36***	64.43***	23.45***	15.26***
Education (%)					
Low	63.02[a]	67.64[a]	58.86[a]	65.02[a]	70.72[a]
Moderate	51.36[b]	51.83[b]	42.62[b]	53.25[b]	63.78[b]
High	44.42[c]	31.22[c]	21.32[c]	39.09[c]	53.30[c]
χ^2-test	38.66***	131.42***	141.39***	67.28***	32.32***

Notes: For each variable, proportions in the same column with different superscripts ([a], [b], [c]) differ significantly at least at $p < 0.05$ (z-test). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. MC1: Algorithms are completely independent, without human influence. MC2: Algorithms always operate neutrally and objectively, and thus, are free of bias. MC3: Algorithms can solve every problem in society. MC4: Algorithms have the same level of critical reasoning and intelligence as humans. MC5: Algorithms will replace humans workers in the media sector.

algorithmic misconceptions compared to moderately and higher-educated respondents.

4.2. Information Sources of Algorithmic Misconceptions (RQ2)

Table 3 gives a summary of the main information sources of respondents holding algorithmic misconceptions. In this table, we used the algorithmic misconception measure, the index ranging from 0 to 5, indicating the number of misconceptions held (with 5 meaning that people hold all five misconceptions; and 0, none). Based on this table, we conclude that respondents with more misconceptions tend to rely less on their own experience, media, and school as sources of information. Conversely, they more commonly rely on friends and/or family as information sources, and are particularly likely to have no information source at all (up to 74.48%). These results suggest that one's own experiences, media, and school are associated with having fewer algorithmic misconceptions, whereas having friends/family as sources or having no information source at all are associated with having more algorithmic misperceptions.

4.3. Predictors of Algorithmic Misconceptions (RQ3)

Multiple regression analysis was performed to explore which variables predict algorithmic misconceptions among respondents. The regression model is presented in Table 4. Age was found to be a significant predictor of algorithmic misconceptions: The older people get, the more algorithmic misperceptions they have ($\beta = 0.09$, $p < 0.001$). Gender was revealed as a positive predictor of misconceptions, with women have significantly more algorithmic misperceptions than men ($\beta = 0.11$, $p < 0.001$). For education, the regression analysis revealed that respondents with moderate ($\beta = -0.08$, $p < 0.01$) and high education ($\beta = -0.15$, $p < 0.001$) levels have significantly fewer algorithmic misconceptions than respondents with a low education level. Altogether,

Table 3. Information sources in function of people's algorithmic misconception index score.

	Algorithmic misconception index score					
	0	1	2	3	4	5
Own experiences (%)	55.30[a]	56.80[a]	45.61[b]	43.77[b]	28.09[c]	10.48[d]
Media (%)	51.89[a]	55.20[a]	55.37[a]	45.45[b]	32.34[c]	12.38[d]
School (%)	9.09[a]	9.87[a]	8.78[a]	6.06[a,b]	2.55[b]	1.52[b]
Friends and/or family (%)	10.29[a]	16.77[b]	18.30[b]	22.44[b]	21.21[b]	23.73[b]
No information source (%)	20.45[a]	16.53[a]	20.73[a]	27.61[b]	43.83[c]	74.48[d]

Notes: Proportions in the same row with different superscripts ([a], [b], [c], [d]) differ significantly at least at $p < 0.05$ (z-test).

Table 4. The predictors of algorithmic misconceptions.

		B	SE	β	t-value	Sig.
Constant		1.77	0.21		8.42	***
Block 1: Demographics						
Age		0.01	0.01	0.09	4.60	***
Gender		0.37	0.07	0.11	5.48	***
Moderate education		−0.26	0.08	−0.08	−3.34	**
High education		−0.66	0.10	−0.15	−6.41	***
R^2 (%)	0.110					***
Block 2: Information sources						
Own experiences		−0.44	0.09	−0.12	−5.06	***
Media		−0.32	0.09	−0.09	−3.59	***
School		−0.12	0.14	−0.02	−0.83	*ns*
Friends and/or family		0.18	0.09	0.04	1.97	*
No information source		0.91	0.11	0.25	8.20	***
Incremental R^2	0.135					***
Total R^2 (%)	0.245					***

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Reference category for gender is "male"; reference category for education is "low education."

these demographic variables explained 11% of the variance. In the second block, we included the information sources. On the one hand, we see that a person's own experience ($\beta = -0.12$, $p < 0.001$) and media ($\beta = -0.09$, $p < 0.001$) are both sources that are negatively associated with having algorithmic misconceptions. On the other, friends/family ($\beta = 0.04$, $p < 0.05$) and no information source ($\beta = 0.25$, $p < 0.001$) were positively linked to algorithmic misconceptions. Information from school was not significantly linked to algorithmic misconceptions ($\beta = -0.02$, *ns*).

5. Discussion

This study showed that misconceptions about algorithms in the media are highly prevalent among the general population in the Netherlands (see Table 2). This prevalence is significantly more pronounced among very specific socio-demographic groups. Results showed that age, education, and gender were significant predictors of algorithmic misperceptions. More precisely, we found that erroneous representations about media algorithms are more common among (1) older people (vs. younger people), (2) lower-educated people (vs. higher-educated), and (3) women (vs. men). In addition, this study also explored the information sources that might contribute to these algorithmic misperceptions. It was found that people who have no information sources about algorithms, and those who rely on their friends and family for such information, were more likely to have algorithmic misconceptions. On the other hand, media, school, and people's own experience were found to be sources associated with having fewer algorithmic misconceptions, suggesting that these three are important sources to convey correct and accurate information about algorithms to the general public. All in all, these results tend to suggest that there is a clear variation in algorithmic misconceptions in society (with a higher prevalence among certain vulnerable parts of the population), which might contribute to new digital divides or inequalities (Cotter & Reisdorf, 2020).

These findings have important contributions to the literature of algorithmic awareness, and the digital divide. As mentioned earlier, the literature on algorithmic awareness is characterized by: 1) a limited body of research; 2) findings that are not entirely conclusive; 3) studies that—almost—exclusively focused on algorithms in specific mediated contexts (e.g., social media algorithms, news algorithms, search algorithms). The contribution of this study is that it focused on people's (mis)conceptions about algorithms on a more general level (without context-specificity) and that it presents insights that might indicate a significant lack of algorithmic awareness among the general population (particularly among certain vulnerable demographic groups). This point then brings us to the second important contribution, i.e., to the digital divide literature. The current findings raise the issue of whether algorithms are expanding digital divides, rather than closing them. It is important that people all have equal skills and knowledge to benefit from algorithmic systems (or at least, have equal opportunities to develop such skills and knowledge); if not, this may create what has been called "algorithmic divides" (Carmi & Yates, 2020). This divide manifests itself in parts of the population having a clear idea on how to benefit most of algorithmic technologies (e.g., young, high-educated individuals), whereas other parts of the population, including more vulnerable groups, might be excluded from the advantages of these technologies (e.g., older, low-educated individuals) and

fail to understand the role of algorithms in the media or the role that humans can play in algorithmic processes. As this study made clear differences in misconception visible, we hope that our findings raise awareness of algorithmic misconception as a factor that can contribute to digital exclusion and divides.

Finally, the findings from our study also entail some important contribution to both the literature and the practice of algorithmic accountability. According to recent findings, the Netherlands supposedly belongs to the leading group of EU countries when it comes to digital literacy (highest proportion of residents skilled in using tech; Eurostat, 2020; "The Netherlands ranks," 2020). It is therefore disturbing to see that even in a country with a relatively high level of digital literacy a rather significant part of the population holds algorithmic misconceptions. Recently, the European Commission set out a path to boosting the investment in, and widespread implementation of AI and algorithms (European Commission, 2020b). Europe is on its way to becoming an algorithmic society—a society shaped by the interplay of humans and the coding and processing of information through algorithms, and that increasingly depends on data-driven processes and decision-making systems. The main questions then are: (1) Will citizens be able to ask the right critical questions about the role and functioning of digital technology?, (2) does the population possess the necessary level of literacy to benefit from these systems?, and (3) are users sufficiently prepared to recognized and protect themselves from possible negative consequences of these technologies? A society in which a significant share of users hold (serious) misconceptions about the potential and workings of algorithms is hardly able to engage critically with algorithmic solutions. Such a society might not be prepared to decide where and how (not) to use algorithms and may be unable to understand their own role in algorithmic processes or to compel those who wield algorithmic power to respect their fundamental rights and public values. The lack of critical digital citizenship, again, can become a potential source of societal problems, such as the spread of misinformation and deep fakes, data-driven manipulation and re-enforcing stereotypes, and inequalities and discrimination (Eubanks, 2017; Mohamed et al., 2020). Our study raises a number of critical follow-up questions, for example regarding the literature on algorithmic transparency and accountability. Transparency is often discussed as a tool to overcome the information asymmetries between users, governments, and corporations, but is it also the role of transparency to correct misconceptions? Or would this require different interventions? To what extent can transparency enhancing measures even contribute to the creation of misconceptions? Can consent to data processing be considered "informed" in the sense of the GDPR if it based on misconceptions? What additional regulatory, policy, and organizational safeguards are needed to empower users to be able to hold algorithmic power to account? But also, where are the limits to "accountability by user empowerment" if users are not able to make fully informed decisions because these are based on misconceptions?

Our study also underlines the urgency of digital literacy education programs and more attention to the role of algorithms in the media. In such literacy initiatives, it is important to balance the need to equip citizens with protective strategies to face the harmful consequences of online algorithms, but at the same time, also to focus on empowering them with a nuanced appreciation of what algorithms are—and what they are not—and the ways in which algorithms may benefit individuals and society (Hobbs, 2020). These initiatives can be implemented within education programs at school or offered via mediated channels—two sources that we found to be particularly influential in debunking algorithmic misconceptions. It is also important to pay particular attention to vulnerable groups that are much harder to reach via schools and media, such as older age groups and lower-educated people. For these groups, a more tailored approach might be needed. Related to this, the results of this study seem to align with the argument that we should be wary of labelling young people as "digital natives" (Kirschner & De Bruyckere, 2017; Kirschner & van Merriënboer, 2013). Although the younger respondents in this study had lower levels of algorithmic misconceptions compared to older age groups, we can still conclude that these misconceptions are present among a considerable share of young adults. Based on these insights, labelling them as digital natives might obscure their need for support in developing the necessary skills to correctly understand algorithmic processes in the media. Therefore, future discussions about educational policy and practice should not be embedded in a mindset that considers young people—by default—as well-versed in algorithmic technology (i.e., digital natives), but rather from the perspective that further education and training is needed to teach them about the uses and consequences of algorithms.

Finally, we also want to address some limitations that could inspire future research. First, this research is based on a very exploratory and descriptive analysis of algorithmic misconceptions. We, therefore, encourage scholars to examine this topic in more depth, with the current study serving as a starting point. There are still many questions that remain unanswered, such as: What are the root causes of these misconceptions? Is the prevalence of algorithmic misconceptions in other countries comparable? Are these misconceptions caused by high levels of trust in the capabilities of algorithms (technological solutionism), or by a lack of critical thinking? What are the consequences of these misconceptions for the willingness to use or trust digital technology? A second limitation relates to the misconceptions selected in this study. They were chosen because of their importance and prominence in the literature. But, our list of misconceptions is by no means exhaustive, meaning that there can be still other algorithmic misconceptions in our

society. Therefore, we encourage scholars to capture the full breadth of misperceptions. This will help to get a better image of algorithmic awareness of citizens. A third limitation relates to the unpredictability of future algorithmic developments. That is, the algorithmic misconceptions discussed in this study might not necessarily be misconceptions in the future. For instance: Algorithms are not—significantly—replacing media workers at this point, but it would not be unreasonable to expect that in some far-distant future, humans might be replaced by efficient algorithms in certain media sectors. So, whether these misconceptions will eventually stand the test of time, remains to be seen and so keeping track of how these misconceptions develop through time would provide an interesting avenue for future research. Finally, this study did not measure variables related to familiarity with technology and digital media. It would therefore be interesting to investigate how these misconceptions relate to technological savviness and digital literacy and to explore how important these factors are in forming algorithmic misconceptions.

Acknowledgments

This work was funded by the Research Priority Area ICDS from the University of Amsterdam.

Conflict of Interests

The authors declare no conflict of interests.

References

Amoore, L., & Piotukh, V. (2015). Life beyond big data: Governing with little analytics. *Economy and Society*, *44*(3), 341–366. https://doi.org/10.1080/03085147.2015.1043793

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Beckett, C. (2019, November 18). New powers, new responsibilities: A global survey of journalism and artificial intelligence. *Polis*. https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities

Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, *20*(1), 1–13. https://doi.org/10.1080/1369118X.2016.1216147

Bucher, T. (2018). *If…then: Algorithmic power and politics*. Oxford University Press.

Carmi, E., & Yates, S. J. (2020). What do digital inclusion and data literacy mean today? *Internet Policy Review*, *9*(2). https://doi.org/10.14763/2020.2.1474

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, *107*(6), 1753–1820.

Cotter, K., & Reisdorf, B. C. (2020). Algorithmic knowledge gaps: A new dimension of (digital) inequality. *International Journal of Communication*, *14*, 745–765.

Council of Europe. (2020). *Recommendation CM/Rec(2020)1 of the committee of ministers to member states on the human rights impacts of algorithmic systems*. https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154

Dack, S. (2019, March 20). Deep fakes, fake news, and what comes next. *The Henry M. Jackson School of International Studies*. https://jsis.washington.edu/news/deep-fakes-fake-news-and-what-comes-next

de Saint Laurent, C. (2018). In defence of machine learning: Debunking the myths of artificial intelligence. *Europe's Journal of Psychology*, *14*(4), 734–747. https://doi.org/10.5964/ejop.v14i4.1823

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.

Elish, M. C., & boyd, d. (2018). Situating methods in the magic of big data and AI. *Communication Monographs*, *85*(1), 57–80. https://doi.org/10.1080/03637751.2017.1375130

Emmert-Streib, F., Yli-Harja, O., & Dehmer, M. (2020). Artificial intelligence: A clarification of misconceptions, myths and desired status. *Frontiers in Artificial Intelligence*, *3*, 1–7.

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015). "I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in news feeds. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems—CHI '15* (pp. 153–162). ACM. https://doi.org/10.1145/2702123.2702556

Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police, and punish the poor* (1st ed.). St. Martin's Press.

European Commission. (2020a). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on The European democracy action plan* (COM(2020) 790). European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:790:FIN

European Commission. (2020b). *Shaping Europe's digital future*. Publications Office of the European Union. https://ec.europa.eu/info/sites/default/files/communication-shaping-europes-digital-future-feb2020_en_4.pdf

Eurostat. (2020). *Individuals' level of digital skills* [Data set]. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_sk_dskl_i&lang=en

Ferrer-Conill, R., & Tandoc, E. C. (2018). The audience-oriented editor: Making sense of the audience in the newsroom. *Digital Journalism*, *6*(4), 436–453. https://doi.org/10.1080/21670811.2018.1440972

First, D. (2018). Will big data algorithms dismantle the foundations of liberalism? *AI & SOCIETY*, *33*(4),

545–556. https://doi.org/10.1007/s00146-017-0733-4

Frank, M., Roehrig, P., & Pring, B. (2017). *What to do when machines do everything: How to get ahead in a world of AI, algorithms, bots, and big data*. Wiley.

Fry, H. (2019). *Hello world: How to be human in the age of the machine*. Transworld Publishers.

Gillespie, T. (2010). The politics of "platforms." *New Media & Society*, *12*(3), 347–364. https://doi.org/10.1177/1461444809342738

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies* (pp. 167–194). MIT Press. https://doi.org/10.7551/mitpress/9780262525374.003.0009

Hargittai, E. (2010). Digital na(t)ives? Variation in internet skills and uses among members of the "net generation." *Sociological Inquiry*, *80*(1), 92–113. https://doi.org/10.1111/j.1475-682X.2009.00317.x

Hargittai, E. (2020). Potential biases in big data: Omitted voices on social media. *Social Science Computer Review*, *38*(1), 10–24. https://doi.org/10.1177/0894439318788322

Hargittai, E., Gruber, J., Djukaric, T., Fuchs, J., & Brombach, L. (2020). Black box measures? How to study people's algorithm skills. *Information, Communication & Society*, *23*(5), 764–775. https://doi.org/10.1080/1369118X.2020.1713846

Hintze, A., Dencik, L., & Wahl-Jorgensen, K. (2019). *Digital citizenship in a datafied society*. Polity.

Hobbs, R. (2020). Propaganda in an age of algorithmic personalization: Expanding literacy research and practice. *Reading Research Quarterly*, *55*(3), 521–533. https://doi.org/10.1002/rrq.301

Kirschner, P. A., & De Bruyckere, P. (2017). The myths of the digital native and the multitasker. *Teaching and Teacher Education*, *67*, 135–142. https://doi.org/10.1016/j.tate.2017.06.001

Kirschner, P. A., & van Merriënboer, J. J. G. (2013). Do learners really know best? Urban legends in education. *Educational Psychologist*, *48*(3), 169–183. https://doi.org/10.1080/00461520.2013.804395

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures & their consequences*. SAGE.

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), 14–29. https://doi.org/10.1080/1369118X.2016.1154087

Koene, A., Clifton, C., Hatada, Y., Webb, H., Patel, M., Machaod, C., LaViolette, J., Richardson, R., & Reisman, D. (2019). *A governance framework for algorithmic accountability and transparency*. European Parliament. https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1). https://doi.org/10.1177/2053951718756684

Meijer, A., & Grimmelikhuijsen, S. (2021). Responsible and accountable algorithmization: How to generate citizen trust in governmental usage of algorithms. In M. Schuilenburg & R. Peters (Eds.), *The algorithmic society: Technology, power, and knowledge* (pp. 53–66). Routledge.

Menz, C., Spinath, B., & Seifried, E. (2021). Where do pre-service teachers' educational psychological misconceptions come from? The roles of anecdotal versus scientific evidence. *Zeitschrift Für Pädagogische Psychologie*, *35*(2/3), 1–14. https://doi.org/10.1024/1010-0652/a000299

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, *33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Morozov, E. (2014). *To save everything, click here: Technology, solutionism, and the urge to fix problems that don't exist*. Penguin Books.

Mosco, V. (2004). *The digital sublime: Myth, power, and cyberspace*. MIT Press.

Natale, S., & Ballatore, A. (2020). Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence: The International Journal of Research into New Media Technologies*, *26*(1), 3–18. https://doi.org/10.1177/1354856517715164

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Powers, E. (2017). My news feed is filtered? *Digital Journalism*, *5*(10), 1315–1335. https://doi.org/10.1080/21670811.2017.1286943

Rader, E., & Gray, R. (2015). Understanding user beliefs about algorithmic curation in the Facebook news feed. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems—CHI '15* (pp. 173–182). ACM. https://doi.org/10.1145/2702123.2702174

Ricci, F. (2015). *Recommender systems handbook*. Springer.

Roffel, S., & Evans, I. (2018, July 16). The biggest misconceptions about AI: The experts' view. *Elsevier*. https://www.elsevier.com/connect/the-biggest-misconceptions-about-ai-the-experts-view

Schapals, A. K., & Porlezza, C. (2020). Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions. *Media and Communication*, *8*(3), 16–26. https://doi.org/10.17645/mac.v8i3.3054

Schreurs, K., Quan-Haase, A., & Martin, K. (2017). Problematizing the digital literacy paradox in the context of older adults' ICT use: Aging, media discourse, and self-determination. *Canadian Journal of Communication*, *42*(2). https://doi.org/10.22230/cjc.2017v42n2a3130

Seaver, N. (2018). What should an anthropology of algorithms do? *Cultural Anthropology*, *33*(3), 375–385. https://doi.org/10.14506/ca33.3.04

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., & Elmqvist, N. (2018). *Designing the user interface: Strategies for effective human-computer interaction*. Pearson.

Taylor, A. K., & Kowalski, P. (2004). Naïve psychological science: The prevalence, strength, and sources of misconceptions. *The Psychological Record*, *54*(1), 15–25. https://doi.org/10.1007/BF03395459

The Netherlands ranks among the EU top in digital skills. (2020, February 2). *CBS*. https://www.cbs.nl/en-gb/news/2020/07/the-netherlands-ranks-among-the-eu-top-in-digital-skills

van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society*. Oxford University Press.

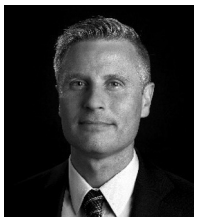
van Dijk, J. A. G. M., & van Deursen, A. J. A. M. (2014). *Digital skills: Unlocking the information society*. Palgrave Macmillan. https://doi.org/10.1057/978113743703

Zarouali, B., Boerman, S. C., & de Vreese, C. H. (2021). Is this recommended by an algorithm? The development and validation of the algorithmic media content awareness scale (AMCA-scale). *Telematics and Informatics*, *62*, Article 101607. https://doi.org/10.1016/j.tele.2021.101607

About the Authors

Brahim Zarouali is an assistant professor in persuasion and new media technologies at the Amsterdam School of Communication Research (ASCoR) of the University of Amsterdam. His interests can be situated at the intersection of persuasive communication and new media technologies (e.g., recommendation algorithms, chatbots, virtual assistants, etc.), with a specific focus on how these technologies shape individuals' views, attitudes, and behaviours. He usually focuses on the broader implications of these technologies on our society as a whole.

Natali Helberger is a distinguished university professor of law and digital technology, with a special focus on AI, and affiliated with the Institute for Information Law (IVIR) of the University of Amsterdam. Her research focus is on how digitization, algorithms, and AI are transforming the media and the implications for public values, diversity in the media landscape, and the media's democratic role.



Claes H. de Vreese is a distinguished faculty professor of artificial intelligence, data, and democracy and professor of political communication at the Amsterdam School of Communication Research (ASCoR) and University of Amsterdam (UvA). His research interests include the role of data and artificial intelligence in democratic processes, microtargeting, comparative journalism research, the effects of news, public opinion, and European integration.

COGITATIO

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 4, Pages 145–157
https://doi.org/10.17645/mac.v9i4.4162

Article

Algorithmic Self-Tracking for Health: User Perspectives on Risk Awareness and Coping Strategies

Noemi Festic [1,*], Michael Latzer [1] and Svetlana Smirnova [2]

[1] Department of Communication and Media Research, University of Zurich, Switzerland;
E-Mails: n.festic@ikmz.uzh.ch (N.F.), m.latzer@ikmz.uzh.ch (M.L.)
[2] Department of Media and Communications, London School of Economics and Political Science, UK;
E-Mail: s.smirnova@lse.ac.uk

* Corresponding author

Submitted: 7 February 2021 | Accepted: 21 September 2021 | Published: 18 November 2021

Abstract
Self-tracking with wearable devices and mobile applications is a popular practice that relies on automated data collection and algorithm-driven analytics. Initially designed as a tool for personal use, a variety of public and corporate actors such as commercial organizations and insurance companies now make use of self-tracking data. Associated social risks such as privacy violations or measurement inaccuracies have been theoretically derived, although empirical evidence remains sparse. This article conceptualizes self-tracking as algorithmic-selection applications and empirically examines users' risk awareness related to self-tracking applications as well as coping strategies as an option to deal with these risks. It draws on representative survey data collected in Switzerland. The results reveal that Swiss self-trackers' awareness of risks related to the applications they use is generally low and only a small number of those who self-track apply coping strategies. We further find only a weak association between risk awareness and the application of coping strategies. This points to a cost-benefit calculation when deciding how to respond to perceived risks, a behavior explained as a privacy calculus in extant literature. The widespread willingness to pass on personal data to insurance companies despite associated risks provides further evidence for this interpretation. The conclusions—made even more pertinent by the potential of wearables' track-and-trace systems and state-level health provision—raise questions about technical safeguarding, data and health literacies, and governance mechanisms that might be necessary considering the further popularization of self-tracking for health.

Keywords
algorithmic selection; coping strategies; mHealth; risk awareness; self-tracking apps; self-quantification; societal risks; user perception; wearables

Issue
This article is part of the issue "Algorithmic Systems in the Digital Society" edited by Sanne Kruikemeier (University of Amsterdam, The Netherlands), Sophie Boerman (University of Amsterdam, The Netherlands) and Nadine Bol (Tilburg University, The Netherlands).

1. Introduction

Algorithms are shaping many domains of our datafied lives, from the curation of news content to recommendations for what to buy. Self-tracking for health is no exception: this digital variant of self-surveillance is performed with the help of wearable devices (e.g., sports bracelets, smart jewelry) and mobile applications. It typically involves continuous data collection, storage, and analysis, which results in algorithmically-derived health recommendations, quasi-human motivational communication, and competitive benchmarking against peers. While self-trackers measure various aspects of their lives, the central focus of this article is on health, fitness, and

wellness tracking, which revolves around measuring and analyzing aspects of physical and mental well-being (e.g., sleep, diet, stress) and athletic performance.

In the last decade, self-tracking has grown exponentially in popularity and reach. In 2020, close to half a billion wearables were in use worldwide. The market of related mobile applications is highly concentrated: From more than 300,000 healthcare apps available, 36 account for more than half of all downloads (estimates by IMS Institute for Healthcare Informatics, 2015). Similarly, the market for wearables is split between five dominant players—Apple, Xiaomi, Fitbit, Samsung, and Huawei—accounting for nearly two-thirds of devices sold (Statista, 2020).

Self-tracking applications have in common that they rely on *algorithmic selection*, defined as a special kind of selection that builds on the automated assignment of relevance to certain pieces of information (Latzer et al., 2016). Risks that can be associated with the employment of algorithmic selection in widespread online services are receiving much public and academic attention. Personalized algorithmic selection shapes the practice of self-trackers in multiple and unknown ways. The self-tracking industry has developed a persuasive narrative that values self-optimization, personalization, prediction, and self-management of health. Not least owing to the opacity of these applications and the sensitive, health-related data they use, self-tracking applications have come under public scrutiny. A glance at the historical evolution of the adoption of self-tracking applications reveals that the need for a debate on their risks and benefits has amplified: While such applications were initially designed for personal use only and data was maybe shared with peers on social networks for comparison and motivation, the stakes for users have dramatically increased. A rapidly growing number of public and corporate actors are promoting the use of these services, using the data and linking financial benefits to achieving certain objectives, thereby exacerbating the potential for a variety of social risks: Self-tracking applications have not only been shown to be of dubious scientific quality (Mercurio et al., 2020), but the industry is also poorly regulated, especially when it comes to handling personal data. The European General Data Protection Regulation (GDPR) has, for instance, been assessed as ineffective in adequately accounting for the fast-paced evolution of self-tracking practices (Marelli et al., 2020). Consequently, different governance options such as self-help protection behaviors by users are likely to play an important role in coping with the risks associated with algorithmic-selection applications for health self-tracking (Ireland, 2020). Coping strategies allow users to exert agency against the "panoptic practices" that companies apply (De Certeau, 1984): By monitoring, measuring, and controlling internet user data, they transform their users into measurable types and classify them based on their habitus that mirrors different aspects of their social disposition. Thereby, these internet platforms and services co-construct users' realities by "mirroring their social dispositions in the form of scorings, recommendations, search results or advertisements" (Latzer & Festic, 2019, p. 10). In the context of self-tracking applications, this specifically involves health-related recommendations or scorings, which have an influence on the users' perceptions of themselves and the world. This article defines coping strategies as internet users' counterparts to the companies' data collection and analysis strategies that induce certain risks for users. This understanding is related to Kitchin and Fraser's (2020) notion of "slow computing," which captures a way for users to regain autonomy over their digital lives in the face of ever-accelerating and increasingly encompassing data grabbing infrastructures on the internet. In the context of self-tracking applications, one exemplary risk, induced by their algorithmic nature, is the inaccurate measurements and resulting fitness recommendations that are scientifically unfounded and inapt for the respective user (Depper & Howe, 2017). Double-checking measurements with the aid of different tools is one possible coping strategy for users to regain autonomy (Kitchin & Fraser, 2020) and mitigate risks.

Extant research has not sufficiently studied self-tracking for health in the wider context of the social power of algorithms—although personalized algorithmic selection lies at the core of these applications and provides a helpful framework to investigate associated risks. The call for more representative empirical research from a user perspective (see Albrecht, 2016) has so far not been sufficiently answered. Against the conceptual backdrop of algorithmic selection, this article first contributes to filling these gaps by empirically investigating how aware self-trackers are of the risks associated with health applications and how they cope with them. Second, this article contributes to the understanding of the coping behavior observed. While we know little about risk awareness and coping strategies by individual users in the realm of self-tracking for health, scholarship on online privacy lends a helpful concept to consider: the privacy calculus, which describes cost-benefit calculations that internet users perform when negotiating their online behavior in response to perceived risks to their privacy (see Baruh et al., 2017). As we described above, social risks associated with self-tracking applications for health have been linked to the growing interest of corporate actors in this data. Using the example of sharing personal self-tracking data with insurance companies as a case study, this article empirically explores self-trackers' behaviors in response to risks and in light of benefits attached to sharing personal data. In combination with the first aim introduced above, this article contributes to our (empirical) understanding of the relationship between risk awareness and coping strategies, which could help to shed light on how self-trackers evaluate risks and deal with them.

To fulfil these tasks, this article draws on representative survey data from Switzerland, a highly digitized

country where 95% of the population use the internet and self-tracking applications for health are gaining popularity: while 29% of internet users reported using them in 2017, this share has risen to 41% in 2021 (Latzer et al., 2021).

This article begins by conceptualizing self-tracking applications for health as algorithmic-selection applications. We then present a review of the existing literature on associated risks and coping strategies and introduce the concept of the privacy calculus. After the methodological approach is explained, the results section outlines our empirical insights. Lastly, the findings are interpreted and we conclude by identifying further research directions.

2. Theoretical Background and Review of Relevant Literature

2.1. Self-Tracking as an Algorithmic-Selection Application

While research on self-tracking applications and their implications is emerging, engagement with literature on algorithms often remains superficial. Bol et al. (2019, p. 2) are some of the few who explicitly address the personalized nature of self-tracking applications by referring to "customization," which captures users' "ability to *self*-tailor…mobile health app content and features." While this user-driven self-tailoring as an affordance of self-tracking applications is included in our understanding of algorithmic selection as introduced below, it goes beyond user-initiated personalization and also includes the automated selection of contents that is outside of what users are aware of and can influence.

In general, algorithmic selection describes the process that transforms *input* with the help of automated computational procedures (*throughput*) into *output* (Cormen et al., 2009; Latzer et al., 2016). Figure 1 illustrates how this model aids to understand the functionality of self-tracking applications for health.

The starting point for this algorithmic-selection process embedded in widely used self-tracking applications for health is a user request (e.g., for a training plan) paired with available user characteristics such as personal demographic factors (e.g., gender, age), user behaviors (e.g., levels of physical activity, diet), and personal goals. These user requests and characteristics combined with a basic data set are used as input by these applications to derive output that ranges from graphs of daily step counts and motivational reminders to be physically active, to an alarm being triggered automatically during a specific stage of sleep to ease waking or a prompt to meditate in response to rising stress levels. The inner functioning of algorithmic-selection applications (throughput) remains largely obscure to users, can form the basis for different biases, and relies on computational operations (Latzer et al., 2016). This process of algorithmic selection functions as follows in the context of a specific type of self-tracking for health: Based on data about fitness levels, past running experience, and age (input), a health application and its designated algorithms (throughput) can identify the ideal training strategy and make personalized recommendations to prepare someone for a marathon (output).

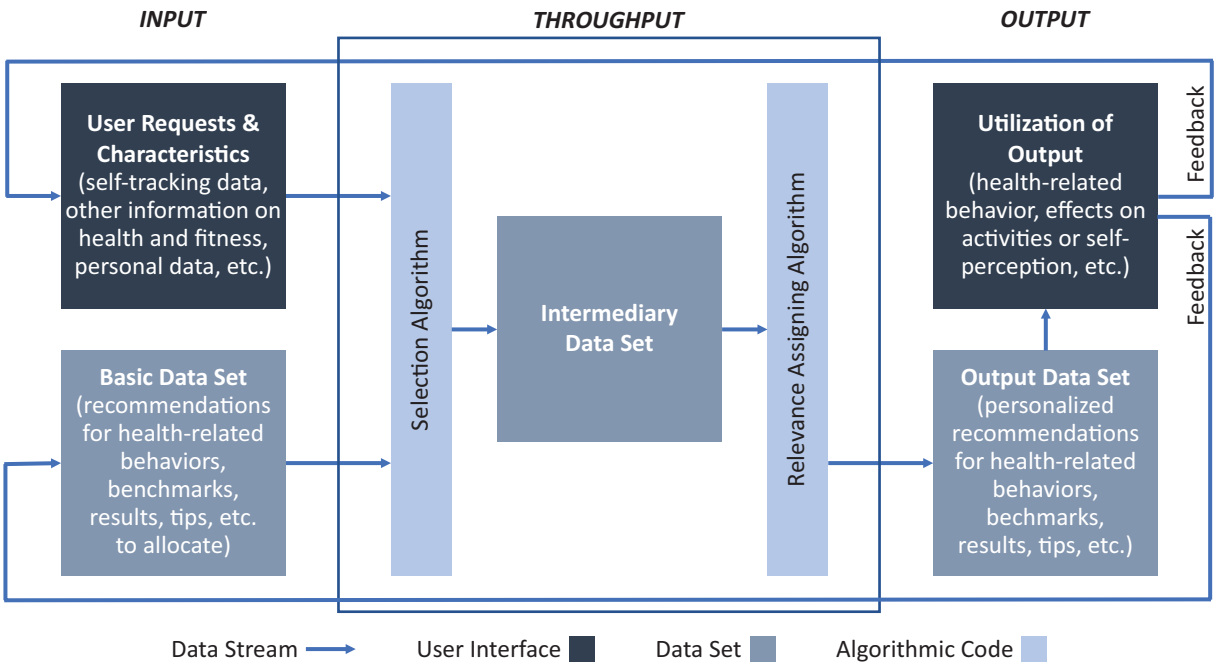


Figure 1. Input–throughput–output model of algorithmic selection applied to self-tracking applications for health and fitness. Source: Adapted from Latzer et al. (2016).

This conceptual understanding of self-tracking applications for health will guide and structure the following considerations on related risks and coping strategies.

2.2. Algorithmic Self-Tracking: Risks and Coping Strategies From a User Perspective

The central arguments of critical scholarship regarding users' risk awareness and coping strategies can be summarized as follows.

While there has been much discussion identifying the *risks* of the spread of algorithmic-selection applications in all domains of life, empirical evidence is only just emerging. Most of the critique directed at algorithmic-selection applications for self-tracking is derived from theoretical reasoning and does not rely on empirical data from a user perspective (for a critique of visualization and analytics, see Fawcett, 2015; and Hepworth, 2017; for a critique of Western-centered, ableist assumptions embedded in tracking systems, see Elias & Gill, 2018; Elman, 2018; Mills & Hilberg, 2020). Risks such as the spread of misinformation (Albrecht, 2016) or use-errors and resulting wrong treatments (Israelski & Muto, 2012) have also only been theoretically derived so far. In their SWOT (strengths, weaknesses, opportunities, and threats) analysis, Li and Hopfgartner (2016) recognize over-tracking and erosion of privacy as weaknesses and negative societal consequences in terms of privacy as a threat of self-tracking applications.

Lack of transparency, particularly in relation to medical evidence, is of special concern given the health-focused nature of the practice. There is robust empirical evidence revealing that expert involvement and adherence to medical evidence is low for various health applications (Chen et al., 2015; Subhi et al., 2015) and longitudinal comparisons reveal that smartphone health apps are not improving in terms of safety or quality (Mercurio et al., 2020). Empirical evaluations of self-tracking applications for weight loss (Mercer et al., 2016) concluded that goals were not adequately backed up by science, sponsorships were not disclosed, sources of information were not cited, and major behavior change techniques were missing.

Qualitative, user-centered research has revealed a variety of self-trackers' concerns, especially considering the output of self-tracking devices: accuracy of data and analysis, inability to edit erroneous entries, weak analytics, and unusable feedback. To exemplify, the accuracy of measurements, the universality of benchmarks (e.g., 10,000 steps or eight hours' sleep at night) and embedded heteronormative assumptions have been sources of concern (Barassi, 2017; Depper & Howe, 2017; Matthews et al., 2017).

Furthermore, privacy remains a significant issue that has been explored in relation to the practice. The risks related to privacy include data trading and access by third parties, lack of legal protection merited by the sensitive nature of data, extensive collection of data irrelevant to the functioning of the application, and users' inability to foresee the extent of data collected on them (Cyr et al., 2014; Daly, 2015; Katuska, 2019). In regard to privacy-related risks, earlier studies showed that self-trackers underestimated the amount of data they shared with companies and lacked knowledge of the conditions of data storage, sharing, and retention, as well as privacy policies, and what they could do to minimize unwanted privacy invasions (Goodyear et al., 2019; Lupton & Michael, 2017; Spiller et al., 2017; Vitak et al., 2018). Recent studies have also suggested that while self-trackers might know about their data being used and believe that harm may come from that (e.g., ovulation data used by an employer for human resources planning), they also think that such scenarios are unlikely to affect them personally (Alqhatani & Lipford, 2019; Gabriele & Chiasson, 2020), which is why they might not engage in mitigation strategies.

As one of the few studies with large-scale survey data in the field, Grzymek and Puntschuh (2019) found across all EU member states that people have little awareness of the potential of algorithms to assist in diagnosing diseases and there was significant concern about medical decisions made by algorithms.

In the realm of coping strategies, existing scholarship suggests that self-trackers use a range of techniques to deal with concerns related to their self-tracking. For example, ethnographic studies have explored how intermediation and reflection are employed by users to cope with problems of inaccuracy, data incompletion, and device breakage (Pink & Fors, 2017a, 2017b; Pink et al., 2017). Alternatively, multiple qualitative studies have illustrated how self-trackers engage in reframing their data, paying selective attention to some data points, or resisting the use of devices as designed (Gorm & Shklovski, 2019; Mopas & Huybregts, 2020; Sjöklint et al., 2015). Other than general research on privacy protection behavior, there is, to the best of our knowledge, no quantitative empirical evidence on how users cope with potential risks in the context of self-tracking applications.

Overall, there is a lack of representative, nation-level data that addresses how aware self-trackers are of various risks and how they cope with them. The discussion of related risks has so far lacked conceptual clarity and not sufficiently taken into account the algorithmic nature of self-tracking applications. When assessing the current state of research with the input-throughput-output model of algorithmic selection in mind, it becomes apparent that most research on risks and coping strategies is limited to the output dimension. We derive the following two research questions from the extant literature for this article:

> RQ1: How aware are Swiss self-trackers of the risks associated with the applications they use and how do they cope with them?

RQ2: How is risk awareness related to the employment of coping strategies among Swiss internet users?

Since the process of personalized algorithmic selection, which underlies the commonly used self-tracking applications, relies heavily on personal data, this topic is intertwined with critical scholarship on online privacy, which has been concerned with questions about how worried internet users are about their data online and how they attempt to protect it. From an (empirical) communication science perspective, privacy-related risks are among those studied most extensively in terms of internet users' awareness and their behavioral and cognitive reactions to it. While early research in the field revealed a seemingly paradoxical relationship between privacy concerns and behavior (e.g., Barnes, 2006; Norberg et al., 2007), more recent studies have replaced this image of ignorant internet users who do not protect their personal data online despite being concerned about their privacy with one where they constantly perform cost-benefit calculations: People engage in online behaviors if the benefit of disclosing personal data or not engaging in protective behaviors, respectively, outweighs the cost (Baruh et al., 2017). Bol et al. (2018) provided experimental empirical evidence for such a "cost-benefit trade-off" in the context of health websites, indicating that both privacy risk perception and perceived benefits were associated with the participants' willingness to self-disclose personal data. When it comes to protection behavior, extant research has shown that—based, for instance, on protection motivation theory—low levels in protective behaviors may be explained by a low perceived self-efficacy despite of high perceived severity of related threats (Boerman et al., 2018). For a convenience sample, Kordzadeh et al. (2016) found empirical proof of a privacy calculus effect on self-disclosure in virtual health communities. Dienlin and Metzger (2016) expanded the privacy calculus framework to include not only self-disclosure, but also self-withdrawal (e.g., deleting posts)—accounting for internet users' co-existing desires for disclosing and withholding information predicted by communication privacy management theory (see Petronio, 2012)—and found empirical evidence for this extended model for a representative sample of adult Facebook users in the US.

Applying this calculus logic to the research interest at hand provides indications for engaging in self-tracking and not applying coping strategies despite being aware of potential risks because the benefits outweigh the cost. A specific, real-world example for these cost-benefit calculations is provided by the rising interest of insurance companies in self-tracking data, offering financial benefits in exchange for personal tracking data. Sharing highly sensitive data on one's health with a third party through an opaque algorithmic-selection application despite a multitude of risks that can arise from this behavior in the short and long run can arguably only be explained if the perceived benefits of this behavior (i.e., a financial compensation) exceed the perceived cost (i.e., any harms from the risks). We use insurance settings as a case study to explore if user behavior is consistent with a calculus logic in the context of self-tracking applications by answering the following question:

RQ3: To what extent are Swiss self-trackers willing to share their data with insurance companies for financial benefit?

An extensive body of research has repeatedly shown that traditional societal fault lines are replicated in the digital space: Male, younger, more affluent members of a society tend to reap more benefits from their internet use and are able to deal with associated risks better (see van Dijk, 2020). Therefore, this article analyzes risk awareness and coping strategies in the realm of self-tracking for health against this backdrop of sociodemographic differences, too.

3. Method

3.1. Data Collection

The empirical section of this article relies on a representative survey of Swiss internet users conducted between October 2018 and February 2019. The survey covered the significance of algorithmic selection for everyday life (Latzer et al., 2020) and included questions on the frequency and purpose of tracking device use, attitudes, risk awareness, and coping strategies, as well as on the willingness to share personal data with insurance companies for financial benefit.

The survey was conducted as part of a larger project in which we also collected internet use tracking data: All participants, who were actively recruited from an existing mobile tracking panel by the LINK Institute, received installation instructions for a passive metering software for their desktop or laptop device (provided by Wakoopa) at the beginning of the field phase. We collected tracking data on private mobile and desktop or laptop devices. The following variables were collected: URL of visited webpages or name of visited app, duration and time of the visit, device, and operating system. On completion of the tracking, the participants received an invitation to complete the online survey questionnaire. While the research questions of this article will be empirically answered with the survey data, the sample description below includes relevant results from the tracking data on the use of self-tracking applications to provide context for the interpretation of the survey results.

3.2. Sample

The original survey sample consisted of $N_{participants}$ = 1,715. As part of the aforementioned questionnaire, the

participants were asked to evaluate the relevance they assign to various online and offline services and activities (e.g., self-tracking applications, offline contacts, search engines) for obtaining information on their personal health. They rated how relevant they believed each of these sources to be for their health information on a scale from 1 = *not at all relevant* to 5 = *very relevant*. For this study, we used a subsample of those participants who assigned some relevance (>1) to an application or device that automatically monitors their fitness or health (N = 716).

The tracking sample consisted of $N_{\text{tracked events}}$ = 13,486,101. We compiled a list of 675 websites and applications which allow their users to automatically track their fitness and health or connect to a wearable device (e.g., a watch) by systematically searching the Apple App Store, Google PlayStore, and Microsoft Store, and by conducting an extensive Google search. By searching the tracking data for occurrences of these app and website names and extracting these cases from the data set, we filtered all uses of self-tracking applications for health from the tracking data set to get descriptive results on the use of these applications in the sample.

Before addressing the guiding research questions, descriptive statistics on self-trackers in Switzerland are presented. Based on the survey data, one in 10 users of tracking applications (11%) reported using such services several times a day and a quarter (25%) reported using them daily. The majority used them either at least weekly (32%) or less than monthly (29%). There were no major differences in the frequency of use of these applications with regard to gender, age, or education. The most common purposes that the respondents reported using their devices for (multiple responses were possible) were fitness and sports (79%), sleep (28%), nutrition (16%), and documenting symptoms associated with a disease (11%).

Of all tracked events, .5% (N = 65,753) were uses of self-tracking applications. We identified 24 unique services used. Table 1 reveals the 10 most used self-tracking applications in descending order (as a share of all tracked use events of self-tracking applications for health). As becomes apparent from the most widespread services, Swiss internet users who engage in self-tracking through mobile applications almost exclusively track their physical activity (e.g., steps, training) and potentially related vital data (e.g., heart rate).

Table 1. Most used self-tracking applications in Switzerland (based on tracking data).

Name	% of self-tracking events
Fitbit	93.14% (N = 61,243)
Google Fit	3.14% (N = 2,062)
TomTom Sports	<.01% (N = 562)
Mi Fit	<.01% (N = 550)
Beurer HealthManager	<.01% (N = 357)
VeryFitPro	<.01% (N = 283)
Huawei Health	<.01% (N = 197)
Sports Tracker	<.01% (N = 136)
Visana-App	<.01% (N = 81)
FunDo Pro	<.01% (N = 57)

These descriptive characteristics of the self-tracking population are important to be kept in mind when interpreting the subsequent empirical answers to this article's guiding research questions.

3.3. Survey Measures

Based on existing literature introduced in Section 2.2, risk awareness was measured for four key risks: The respondents answered on a five-point Likert scale (1 = *do not agree at all*, 5 = *totally agree*) how strongly they agreed that they used their tracking device too much (overuse), were uncertain about the accuracy of their device's measurements (measurement inaccuracy), did not know how their device calculated the results it provides (lack of transparency), and were concerned about what happens with their data (loss of control over data).

To measure coping strategies, the respondents answered how often (1 = *never*, 2 = *rarely*, 3 = *sometimes*, 4 = *frequently*) they checked the accuracy of the measurements by comparing them to other results (checking measurements), how often they did not blindly trust their tracking device's results but actively thought about their meaning (reflecting on results) and how often they consciously refrained from using their tracking device (conscious non-use). Some of these risk awareness and coping strategy items can be clearly situated at one level in the input-throughput-output model of algorithmic selection (e.g., lack of transparency at the throughput level; checking measurements at the output level), others transcend this categorization and concern multiple levels. The goal of this empirical approach was to cover all levels in the measurement of both risk awareness and coping strategies.

The respondents indicated their willingness to share personal data with their insurance company by stating their agreement on a five-point Likert scale (1 = *strongly disagree*, 5 = *strongly agree*) to the following statement: "I would be willing to give my insurance access to my data if I received financial advantages for doing so." While potential risks (i.e., the cost) of using self-tracking applications were not explicitly part of the question, they were made salient to the respondents through multiple questions on risk awareness placed prior in the questionnaire.

The respondents were further asked to report their gender (female, male) as well as their age in years, which was recoded into four groups (16–29, 30–49, 50–69, 70–85) for certain analyses below. They also reported their completed levels of educational attainment, which were recoded into three levels: individuals whose highest completed education level was compulsory schooling were assigned the value *low* and those with tertiary qualifications were assigned the value *high*.

3.4. Data Analysis

Data analysis for RQ1 and RQ3 relied on descriptive statistics. To test the relationship between risk awareness and coping strategies (RQ2), we estimated a path model with the lavaan package in *R* (Rosseel, 2012). For the path model, we used all items separately with the raw scales introduced in Section 3.3. This allowed a detailed analysis of the relationship between different risks and coping strategies. A positive relationship between a risk awareness and a coping strategy item in the model can therefore be interpreted as follows: "stronger agreement with a risk is associated with applying coping strategies more frequently." We freely estimated the covariances between the items for risk awareness and coping strategies, respectively (the script for the analysis and further results are available in the Supplementary Material).

4. Results

The following sections detail our empirical results for the three research questions based on the survey data.

To answer RQ1, we address how widespread the awareness of risks associated with self-tracking applications and the employment of coping strategies is. Figure 2 shows the distribution of responses to the survey questions about risk awareness (*N* = 716).

Overall, awareness of the surveyed risks was low: About four out of ten (39%) to seven out of ten (69%) self-tracking users were not concerned about the risks associated with their self-tracking practice. For overuse and lack of transparency, "do not agree at all" was the modal category: About half of the internet users did not agree at all that they use their tracking device too much (48%) and disagreed or fully disagreed that they do not know how their application calculates health results (54%). Loss of control over data and measurement inaccuracy were different in that the responses were roughly equally distributed: 27% and 30%, respectively, agreed (4) or fully agreed (5) with the statements. Users of self-tracking applications felt more at risk of losing control over their data or being presented with inaccurate measurements than they feared overusing their device or not knowing how their results are calculated.

The application of coping strategies, which can counteract these risks, was distributed as shown in Figure 3 (*N* = 716).

Figure 3 shows that the practice of cross-checking tracking measurements was uncommon: almost half of users (46%) never do this and only a quarter (24%) engage in the practice at least sometimes. One third (33% and 34%, respectively) of self-trackers never consciously decide to not use their tracking device or engage in this practice at least sometimes. Reflecting on one's

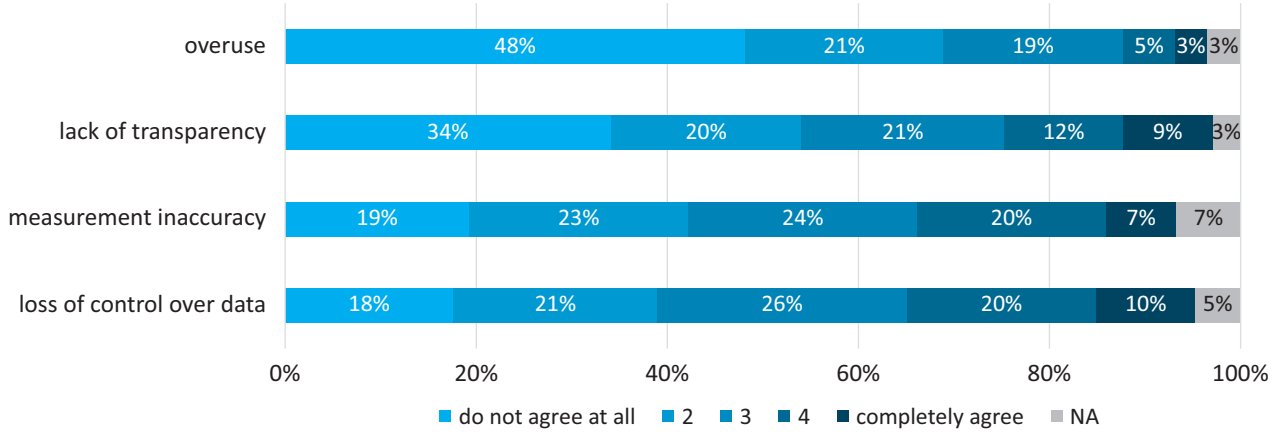


Figure 2. Distribution of indicators of risk awareness.

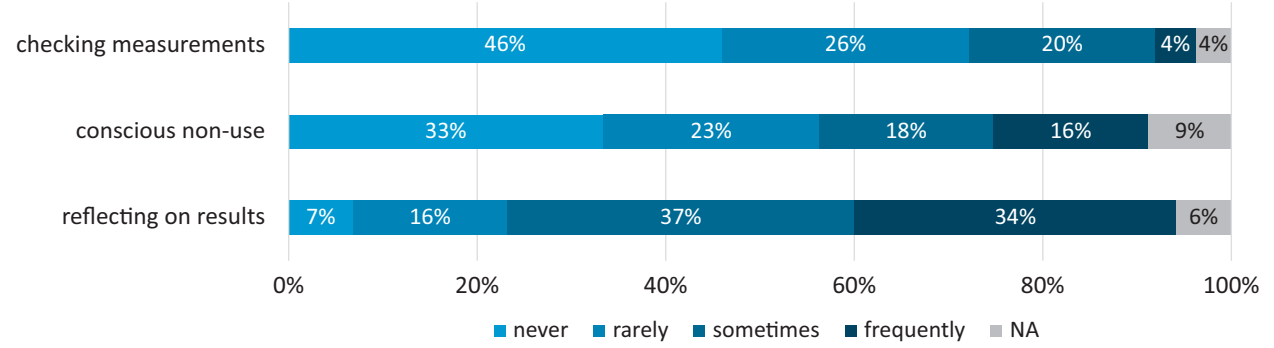


Figure 3. Distribution of indicators of coping strategies.

results was the most widespread coping strategy: only 7% never do this, while 71% of users engage in this practice at least sometimes.

To answer RQ2, we assessed the relationship between risk awareness and coping strategies. The awareness of specific risks and the frequency with which self-tracking users employed coping strategies was only weakly correlated both for the single items and for the two respective mean score indices (for further results see the Supplementary Material).

Figure 4 depicts a path model for the relationship between risk awareness and coping strategies. While gender and education were not significantly related to the two variables of interest, age was added as a control variable.

The model fit the data well: $\chi^2(3, N = 716) = 3.433$ ($p = .330$), $\chi^2/df = 1.144$, CFI = .999, TLI = .991, RMSEA = .014, SRMR = .012. Overall, the awareness of risks related to self-tracking devices explained only very small proportions of the variance in coping strategies. While there were some indications for a positive association between risk awareness and coping strategies—i.e., awareness of the risk to overuse self-tracking was positively associated with double-checking measurements and awareness of the risk of losing control over one's data was positively associated with consciously not using self-trackers—these effects were weak. Age was only significantly (and negatively) associated with the awareness of the risk of measurement inaccuracy.

While the application of coping strategies as a protection behavior does not appear to be meaningfully explained by risk awareness, this article also investigates whether Swiss self-trackers are willing to self-disclose their self-tracking data to insurance companies despite having been made aware of associated risks. RQ3 can be empirically answered as follows: 43% of tracking-device users in Switzerland agreed (4) or completely agreed (5) that they would generally be willing to share their data with their insurance company if they received financial advantages for doing so. This willingness was relatively uniformly distributed across all societal groups (see Figure 5). There was a weak tendency for older people and females to be less willing to share their data. Female self-trackers aged 70 and over reported the lowest willingness to share their data with an insurance company. There were no differences regarding education.

The following section discusses our empirical findings and details how they contribute to answering our research questions.

5. Discussion

Overall, our results reveal that awareness of risks associated with algorithmic self-tracking applications is relatively low and coping strategies are not regularly used. In the realm of risks, the results highlight that users perceive some risks—inaccuracy of measurements and losing control over their data—as more pertinent than others. However, even for those risks, less than a third of Swiss self-trackers reported awareness (RQ1). It is not necessarily the case that those who are more aware of risks engage in coping strategies more often (RQ2). This seemingly paradoxical result could be explained by a "calculus'' logic: Although Swiss self-trackers are somewhat aware of the risks they face, they still engage in the practice and do not apply many coping strategies because they rate the benefits higher than potential risks. Their willingness to share their self-tracked data with insurance companies (when there are direct financial benefits attached) further reiterates the plausibility of this explanation (RQ3). This result extends the extant literature on the privacy calculus (see e.g., Masur, 2019), from which this calculus logic was derived, to other types of risks associated with a specific type of everyday internet use that is dominated by algorithmic selection: self-tracking for health. In accordance with Dienlin

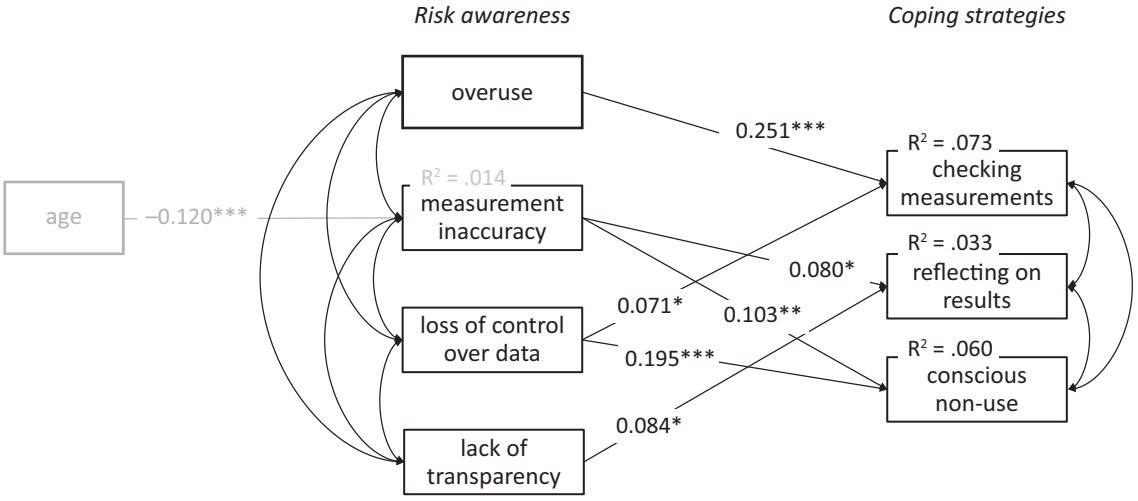


Figure 4. Path model: Risk awareness and coping strategies. Notes: Standardized estimates are shown; only significant paths are shown; ***$p < .001$, **$p < .05$, *$p < .1$

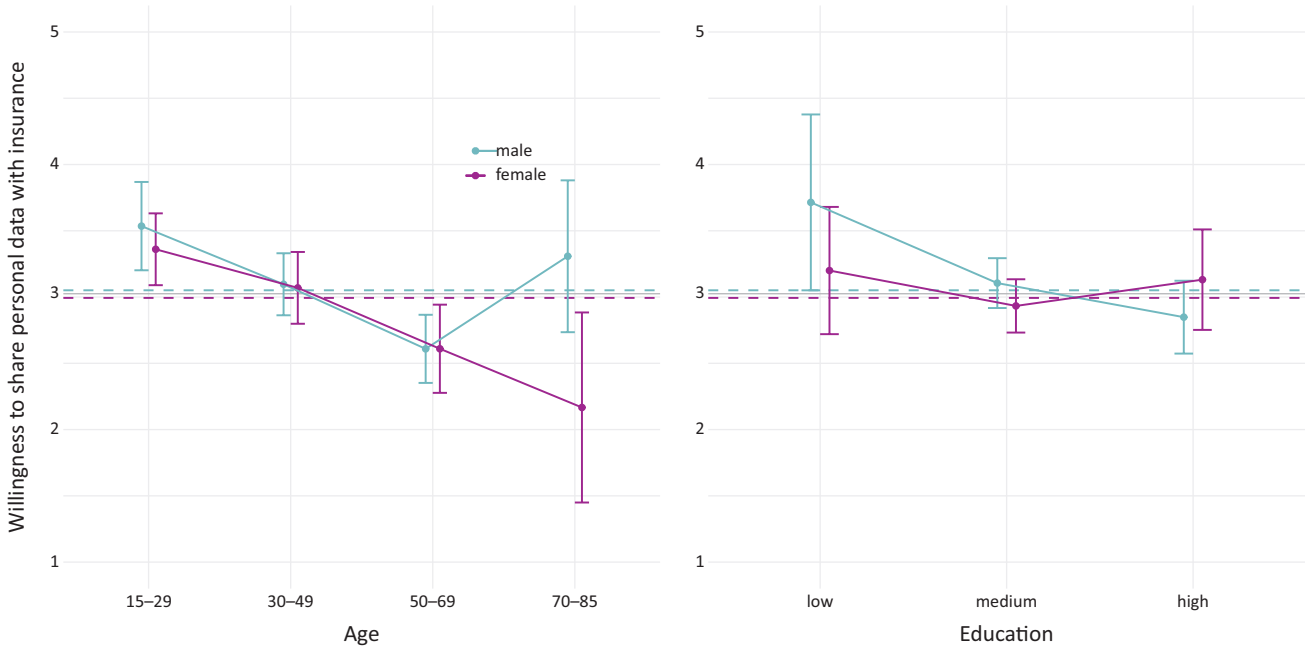


Figure 5. Willingness to share personal data with insurance company: Mean scores by gender, age, and education. Notes: Vertical bars represent 95% confidence intervals; horizontal lines represent overall (solid) and group means (dashed); Y-axis indicates means on a continuous scale (1 = *do not agree at all* to 5 = *completely agree*); *N* = 692.

and Metzger's (2016) empirical results, this effect was also likely present for coping strategies that reflect self-withdrawal behavior (i.e., conscious non-use).

From a public-policy perspective, these are important results to keep in mind when assessing the need for regulatory interventions to mitigate the possibility of certain risks occurring: While users may be familiar with some aspects of algorithmic selection and associated risks, this understanding does not deter them from engaging in the practice of self-tracking in their everyday lives. Alternative interpretations of this weak relationship could include skepticism about the efficacy of coping strategies (Boerman et al., 2018) or mediating effects of personality traits, internet skills, or more general concerns about being online. Our path model for the relationship between risk awareness and coping strategies (see Figure 4) also showed that coping strategies that are arguably effective in light of certain risks (e.g., conscious non-use as a coping strategy in response to awareness about the risk of overuse) were empirically not those most strongly associated with the respective risks. This provides further indications for the aforementioned interpretations and substantiates the need for further research on this relationship.

There are limitations to acknowledge when considering the results and implications of this study. Both survey and tracking data can be subject to biases such as effects of social desirability in surveys or the self-selection of people with certain personal characteristics into tracking samples. Another limitation concerns the list of risks included in this article. We examined a limited number of risks that we perceived as key, but future research should also consider emerging risks that have been associated with self-tracking, such as distorted self-perceptions (Strübing, 2021).

We found that existing research conceives self-tracking applications as a homogenous group. However, such applications and devices vary in the services they offer, the volume, type, and sensitivity of data they collect, the algorithms they employ, and the outputs they provide. Accordingly, the social risks we addressed in this article carry a different weight depending on the context of the self-tracking practice: While the potential risks of incorrect recommendations or data leaks for a chronically ill person relying on a self-tracking device for reminders of their medicine intake may be detrimental for their life chances, the effects of the same events in the context of a healthy person using a step counter are much less significant. This could be an additional, different explanation for the weak association between risk awareness and the application of coping strategies we found in our representative data set, which was almost exclusively composed of individuals who track arguably non-sensitive data (e.g., step counts) and where the potential for harm is therefore comparatively low. With this in mind, our data offer some specific indications that those who are chronically ill or require medical assistance are a group that future research should specifically focus on: Those in the sample who reported engaging in self-tracking to monitor symptoms in connection with a disease were more concerned about losing control over their data (38%, vs. 30% in the entire sample) and less willing to share their data with an insurance company for financial benefit (36%, vs. 43% in the

entire sample)—arguably because the potential harms are much more detrimental for them, even if their occurrence is unlikely. Future research should account for this diversity in self-tracking applications when investigating their uses, implications, and the need for governance interventions. In any circumstance, throwing all self-tracking applications into one basket and proposing generalized, one-size-fits-all explanations or solutions is unpromising for a realistic assessment of their harms and benefits. The identified tensions raise further research, normative, and regulation questions. For instance, it remains an open question if users would be more concerned about the implications of their self-tracking practice if their life chances were more transparently linked to its outcomes (e.g., by tracked data having an impact on premiums).

Examining users' understanding of algorithmic selection embedded in self-tracking applications and associated risks is becoming more pressing as the practice permeates deeper into formal medical settings and drives up the costs of opting out (Lupton, 2015). Today, dominant corporate quantification players are expanding their reach into organizational settings: For example, Fitbit, has developed a dedicated product that is marketed to employers, and a health insurance provider has integrated the use of Apple watches into their wellness plans (United Healthcare, 2021). Organizations (e.g., Target, Barclays, BP, Emory University) and nation-states alike (e.g., Singapore, the UK National Health Service) have initiated the integration of self-quantification into their health delivery operations. Results from more fine-grained studies will be particularly relevant in light of the fast-paced evolution of the adoption of self-tracking applications: from being mere tools for measuring health-related indicators for personal use only, they have more recently attracted the interest of powerful, profit-maximizing institutions that are looking to capitalize on individuals' self-tracking practices and are increasingly pervading private domains such as sleep, mental health, and family planning.

In terms of governance conclusions, we can derive from our results that self-help by individual internet users in the form of coping strategies alone is not a promising path forward when it comes to mitigating the risks associated with algorithmic self-tracking applications that apply panoptic practices. Is there a need for self-, co-, or state regulation and if so, how might the transnational nature of dataflows hinder such efforts? Should the functioning of algorithmic selection (throughput) be made more transparent? While there are attempts such as the mHealth App Trustworthiness checklist (van Haasteren et al., 2019) to systematically assess and improve the quality of self-tracking applications, these studies should take into account that algorithms are at the core of these applications and consider scholarship in the field of critical algorithm studies to advance these endeavors.

6. Conclusion

This article makes two central contributions: On the conceptual level, we have elaborated on the functionality of self-tracking as algorithmic-selection applications and discussed related risks and coping strategies. On the empirical level, we have provided hitherto missing representative evidence of the relationship between risk awareness and coping strategies. Based on tracking data, we also found evidence of a highly concentrated usage of self-tracking applications in Switzerland.

The findings highlight that users recognize some risks associated with algorithmic selection for shaping their practice; however, this awareness is sparse and mostly limited to the applications' input and output levels. The findings also suggest that users employ a limited range of coping strategies to mitigate these risks. Based on these conclusions, we argue that limited awareness of algorithmic functioning and the associated risks does not deter users from adopting self-tracking practices in their everyday lives. In that vein, this article also provides empirical indication for a cost-benefit calculus derived from the weak relationship between risk awareness and coping strategies as well as from the high willingness to share personal data with insurance companies. The blind spots in risk awareness and the toothless nature of coping strategies, however, call for further consideration as the practice continues to permeate medical, corporate, educational, legal, and nation-state settings. Our results substantiate the need for a more differentiated analysis of self-tracking applications, taking into account different types of applications, user groups, and data with different degrees of sensitivity.

Acknowledgments

This project received funding from the Swiss National Science Foundation.

Conflict of Interests

The authors declare no conflict of interests.

Supplementary Material

Supplementary material for this article is available at https://osf.io/ekjx9/?view_only=a513ba966d204966ac388079dfe84d62

References

Albrecht, U.-V. (Ed.). (2016). *Chances and risks of mobile health apps*. Hannover Medical School. https://doi.org/10.24355/dbbs.084-201210110913-73

Alqhatani, A., & Lipford, H. (2019). "There is nothing that I need to keep secret": Sharing practices and concerns of wearable fitness data. In H. R. Lipford (Ed.), *Proceedings of the fifteenth sympo-*

sium on usable privacy and security (pp. 421–434). USENIX. https://www.usenix.org/sites/default/files/soups2019_full_proceedings_interior.pdf

Barassi, V. (2017). BabyVeillance? Expecting parents, online surveillance and the cultural specificity of pregnancy apps. *Social Media + Society*, *3*(2), 1–10. https://doi.org/10.1177/2056305117707188

Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. *First Monday*, *11*(9). http://journals.uic.edu/ojs/index.php/fm/article/view/1394/1312

Baruh, L., Secinti, E., & Cemalcilar, Z. (2017). Online privacy concerns and privacy management: A meta-analytical review. *Journal of Communication*, *67*(1), 26–53. https://doi.org/10.1111/jcom.12276

Boerman, S. C., Kruikemeier, S., & Zuiderveen Borgesius, F. J. (2018). Exploring motivations for online privacy protection behavior: Insights from panel data. *Communication Research*, *48*(7), 953–977. https://doi.org/10.1177/0093650218800915

Bol, N., Dienlin, T., Kruikemeier, S., Sax, M., Boerman, S. C., Strycharz, J., Helberger, N., & de Vreese, C. H. (2018). Understanding the effects of personalization as a privacy calculus: Analyzing self-disclosure across health, news, and commerce contexts. *Journal of Computer-Mediated Communication*, *23*(6), 370–388. https://doi.org/10.1093/jcmc/zmy020

Bol, N., Høie, N. M., Nguyen, M. H., & Smit, E. S. (2019). Customization in mobile health apps: Explaining effects on physical activity intentions by the need for autonomy. *Digital Health*, *5*, 1–12. https://doi.org/10.1177/2055207619888074

Chen, J., Cade, J. E., & Allman-Farinelli, M. (2015). The most popular smartphone apps for weight loss: A quality assessment. *JMIR mHealth and uHealth*, *3*(4), Article e104. https://doi.org/10.2196/mhealth.4334

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms* (3rd ed.). MIT Press.

Cyr, B., Horn, W., Miao, D., & Specter, M. (2014). *Security analysis of wearable fitness devices (Fitbit)*. MIT. https://pdfs.semanticscholar.org/f4ab/ebef4e39791f358618294cd8d040d7024399.pdf

Daly, A. (2015). The law and ethics of "self-quantified" health information: An Australian perspective. *International Data Privacy Law*, *5*(2), 144–155. https://doi.org/10.1093/idpl/ipv001

De Certeau, M. (1984). *The practice of everyday life*. University of California Press.

Depper, A., & Howe, P. D. (2017). Are we fit yet? English adolescent girls' experiences of health and fitness apps. *Health Sociology Review*, *26*(1), 98–112. https://doi.org/10.1080/14461242.2016.1196599

Dienlin, T., & Metzger, M. J. (2016). An extended privacy calculus model for SNSs: Analyzing self-disclosure and self-withdrawal in a representative U.S. sample. *Journal of Computer-Mediated Communication*, *21*(5), 368–383. https://doi.org/10.1111/jcc4.12163

Elias, A. S., & Gill, R. (2018). Beauty surveillance: The digital self-monitoring cultures of neoliberalism. *European Journal of Cultural Studies*, *21*(1), 59–77. https://doi.org/10.1177/1367549417705604

Elman, J. P. (2018). "Find your fit": Wearable technology and the cultural politics of disability. *New Media & Society*, *20*(10), 3760–3777. https://doi.org/10.1177/1461444818760312

Fawcett, T. (2015). Mining the quantified self: Personal knowledge discovery as a challenge for data science. *Big Data*, *3*(4). https://doi.org/10.1089/big.2015.0049

Gabriele, S., & Chiasson, S. (2020, April 25–30). *Understanding fitness tracker users' security and privacy knowledge, attitudes and behaviors* [Paper presentation]. CHI Conference on Human Factors in Computing Systems, Honolulu, HI, US. https://dl.acm.org/doi/proceedings/10.1145/3313831

Goodyear, V. A., Kerner, C., & Quennerstedt, M. (2019). Young people's uses of wearable healthy lifestyle technologies: Surveillance, self-surveillance and resistance. *Sport, Education and Society*, *24*(3), 212–225. https://doi.org/10.1080/13573322.2017.1375907

Gorm, N., & Shklovski, I. (2019). Episodic use: Practices of care in self-tracking. *New Media & Society*, *21*(11/12), 2505–2521. https://doi.org/10.1177/1461444819851239

Grzymek, V., & Puntschuh, M. (2019). *What Europe knows and thinks about algorithms* (Discussion Paper Ethics of Algorithms #10). Bertelsmann Stiftung. https://doi.org/10.11586/2019008

Hepworth, K. (2017). Big data visualization: Promises & pitfalls. *Communication Design Quarterly Review*, *4*(4), 7–19. https://doi.org/10.1145/3071088.3071090

IMS Institute for Healthcare Informatics. (2015). *Patient adoption of mhealth: Use, evidence, and remaining barriers to the mainstream acceptance*. https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf

Ireland, L. (2020). Predicting online target hardening behaviors: An extension of routine activity theory for privacy-enhancing technologies and techniques. *Deviant Behavior*. Advance online publication. https://doi.org/10.1080/01639625.2020.1760418

Israelski, E., & Muto, W. (2012). Human factors risk management for medical products. In P. Carayon (Ed.), *Handbook of human factors and ergonomics in health care and patient safety* (2nd ed., pp. 475–506). CRC Press.

Katuska, J. (2019). Wearing down HIPAA: How wearable technologies erode privacy protections. *Journal of Corporation Law*, *44*(2), 385–401.

Kitchin, R., & Fraser, A. (2020). *Slow computing: Why we need balanced digital lives*. Bristol University Press.

Kordzadeh, N., Warren, J., & Seifi, A. (2016). Antecedents

of privacy calculus components in virtual health communities. *International Journal of Information Management*, *36*(5), 724–734. https://doi.org/10.1016/j.ijinfomgt.2016.04.015

Latzer, M., Büchi, M., Kappeler, K., & Festic, N. (2021). *Internetverbreitung und digitale Bruchlinien in der Schweiz 2021* [Internet diffusion and digital fault lines in Switzerland 2021]. University of Zurich. http://mediachange.ch/research/wip-ch-2021

Latzer, M., & Festic, N. (2019). A guideline for understanding and measuring algorithmic governance in everyday life. *Internet Policy Review*, *8*(2), 1–19. https://doi.org/10.14763/2019.2.1415

Latzer, M., Festic, N., & Kappeler, K. (2020). *Use and assigned relevance of algorithmic-selection applications in Switzerland*. University of Zurich. https://mediachange.ch/research/algosig

Latzer, M., Hollnbuchner, K., Just, N., & Saurwein, F. (2016). The economics of algorithmic selection on the internet. In J. Bauer & M. Latzer (Eds.), *Handbook on the economics of the internet* (pp. 395–425). Edward Elgar. https://doi.org/10.4337/9780857939852

Li, N., & Hopfgartner, F. (2016). To log or not to log? SWOT analysis of self-tracking. In S. Selke (Ed.), *Lifelogging* (pp. 305–325). Springer. https://doi.org/10.1007/978-3-658-13137-1_17

Lupton, D. (2015). Quantified sex: A critical analysis of sexual and reproductive self-tracking using apps. *Culture, Health and Sexuality*, *17*(4), 440–453. https://doi.org/10.1080/13691058.2014.920528

Lupton, D., & Michael, M. (2017). "Depends on who's got the data": Public understandings of personal digital dataveillance. *Surveillance & Society*, *15*(2), 254–268. https://doi.org/10.24908/ss.v15i2.6332

Marelli, L., Lievevrouw, E., & Hoyweghen, I. V. (2020). Fit for purpose? The GDPR and the governance of European digital health. *Policy Studies*, *41*(5), 447–467. https://doi.org/10.1080/01442872.2020.1724929

Masur, P. K. (2019). *Situational privacy and self-disclosure*. Springer. https://doi.org/10.1007/978-3-319-78884-5

Matthews, M., Murnane, E., & Snyder, J. (2017). Quantifying the changeable self: The role of self-tracking in coming to terms with and managing bipolar disorder. *Human-Computer Interaction*, *32*(5/6), 413–446. https://doi.org/10.1080/07370024.2017.1294983

Mercer, K., Li, M., Giangregorio, L., Burns, C., & Grindrod, K. (2016). Behavior change techniques present in wearable activity trackers: A critical analysis. *JMIR mHealth and uHealth*, *4*(2). https://doi.org/10.2196/mhealth.4461

Mercurio, M., Larsen, M., Wisniewski, H., Henson, P., Lagan, S., & Torous, J. (2020). Longitudinal trends in the quality, effectiveness and attributes of highly rated smartphone health apps. *Evidence Based Mental Health*, *23*(3), 107–111. https://doi.org/10.1136/ebmental-2019-300137

Mills, C., & Hilberg, E. (2020). The construction of mental health as a technological problem in India. *Critical Public Health*, *30*(1), 41–52. https://doi.org/10.1080/09581596.2018.1508823

Mopas, M. S., & Huybregts, E. (2020). Training by feel: Wearable fitness-trackers, endurance athletes, and the sensing of data. *The Senses and Society*, *15*(1), 25–40. https://doi.org/10.1080/17458927.2020.1722421

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, *41*(1), 100–126. https://doi.org/10.1111/j.1745—6606.2006.00070.x

Petronio, S. (2012). *Boundaries of privacy: Dialectics of disclosure*. SUNY Press.

Pink, S., & Fors, V. (2017a). Being in a mediated world: Self-tracking and the mind-body-environment. *Cultural Geographies*, *24*(3), 375–388. https://doi.org/10.1177/1474474016684127

Pink, S., & Fors, V. (2017b). Self-tracking and mobile media: New digital materialities. *Mobile Media & Communication*, *5*(3), 219–238. https://doi.org/10.1177/2050157917695578

Pink, S., Sumartojo, S., Lupton, D., & Heyes La Bond, C. (2017). Mundane data: The routines, contingencies and accomplishments of digital living. *Big Data & Society*, *4*(1), 1–12. https://doi.org/10.1177/2053951717700924

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(1), 1–36. https://doi.org/10.18637/jss.v048.i02

Sjöklint, M., Constantiou, I., & Trier, M. (2015). The complexities of self-tracking: An inquiry into user reactions and goal attainment. In J. Becker, J. vom Brocke, & M. de Marco (Eds.), *ECIS 2015 completed research papers* (Paper No. 170). European Conference on Information Systems. https://doi.org/10.18151/7217479

Spiller, K., Ball, K., Bandara, A., Meadows, M., Mccormick, C., Nuseibeh, B., & Price, B. A. (2017). Data privacy: Users' thoughts on quantified self personal data. In B. Ajana (Ed.), *Self-tracking: Empirical and philosophical investigations* (pp. 111–124). Palgrave Macmillan.

Statista. (2020). *Wearable technology: Statistics & facts*. https://www.statista.com/topics/1556/wearable-technology

Strübing, J. (2021). Selbstvermessung als Subjektivierungsweise [Self-quantification as a way of subjectivation]. In K. Brümmer, A. Janetzko, & T. Alkemeyer (Eds.), *Ansätze einer Kultursoziologie des Sports* [Approaches to a cultural sociology of sports] (pp. 231–248). Nomos.

Subhi, Y., Bube, S. H., Bojsen, S. R., Thomsen, A. S. S., & Konge, L. (2015). Expert involvement and adherence to medical evidence in medical mobile phone apps: A systematic review. *JMIR MHealth and UHealth*,

3(3), Article e79. https://doi.org/10.2196/mhealth.4169

United Healthcare. (2021). *Wellness & rewards programs*. https://www.uhc.com/employer/communication-resources/wellness-and-rewards-programs

van Dijk, J. (2020). *The digital divide*. Polity.

van Haasteren, A., Gille, F., Fadda, M., & Vayena, E. (2019). Development of the mhealth app trustworthiness checklist. *Digital Health*, *5*, 1–12. https://doi.org/10.1177/2055207619886463

Vitak, J., Liao, Y., Kumar, P., Zimmer, M., & Kritikos, K. (2018). Privacy attitudes and data valuation among fitness tracker users. In G. Chowdhury, J. McLeod, V. Gillet, & P. Willett (Eds.), *Transforming digital worlds* (pp. 229–239). Springer. https://doi.org/10.1007/978-3-319-78105-1_27

About the Authors

Noemi Festic is a research and teaching associate in the Media Change & Innovation Division, Department of Communication and Media Research (IKMZ), University of Zurich, Switzerland. Her research interests include internet use, and the use of algorithmic-selection applications in particular, and its implications on everyday life and personal well-being. Her current research focuses on how computational methods can contribute to a better empirical understanding of the role of algorithmic selection for everyday life.

Michael Latzer is professor of communications at the Department of Communication and Media Research (IKMZ), University of Zurich, Switzerland, where he chairs the Media Change & Innovation Division. His research focuses on the co-evolution of technical, economic, political, and social innovations in the convergent communications sector, in particular on information society issues, internet research, and the significance of algorithmic selection. For details, see mediachange.ch.

Svetlana Smirnova has recently completed her PhD at the Department of Media and Communications of the London School of Economics and Political Science. Her current research interests include self-tracking and self-quantification, digital selfhood, research design and methodologies. Most recently, Svetlana has served as a post-doctoral researcher on a research and development initiative focused on the use of age-assurance and parental control tools.

COGITATIO

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 4, Pages 158–169
https://doi.org/10.17645/mac.v9i4.4085

Article

Political Microtargeting and Online Privacy: A Theoretical Approach to Understanding Users' Privacy Behaviors

Johanna Schäwel *, Regine Frener and Sabine Trepte

Department of Communication Science: Media Psychology, University of Hohenheim, Germany;
E-Mails: johanna.schaewel@uni-hohenheim.de (J.S.), regine.frener@uni-hohenheim.de (R.F.),
sabine.trepte@uni-hohenheim.de (S.T.)

* Corresponding author

Submitted: 29 January 2021 | Accepted: 17 August 2021 | Published: 18 November 2021

Abstract
Social media allow political parties to conduct political behavioral targeting in order to address and persuade specific groups of users and potential voters. This has been criticized: Most social media users do not know about these microtargeting strategies, and the majority of people who are aware of targeted political advertising say that it is not acceptable. This intrusion on personal privacy is viewed as problematic by users and activists alike. The overarching goal of this article is to elaborate on social media users' privacy perceptions and potential regulating behaviors in the face of political microtargeting. This work is theoretical in nature. We first review theoretical and empirical research in the field of political microtargeting and online privacy. We then analyze how privacy is experienced by social media users during political microtargeting. Building on our theoretical analysis, we finally suggest clear-cut propositions for how political microtargeting can be researched while considering users' privacy needs on the one hand and relevant political outcomes on the other.

Keywords
online privacy; political microtargeting; social media affordances; social media privacy model

Issue
This article is part of the issue "Algorithmic Systems in the Digital Society" edited by Sanne Kruikemeier (University of Amsterdam, The Netherlands), Sophie Boerman (University of Amsterdam, The Netherlands) and Nadine Bol (Tilburg University, The Netherlands).

1. Introduction

Political microtargeting can be regarded as a pivotal tool amongst the different campaign instruments that exist. Oftentimes, microtargeting takes place on social media (Papakyriakopoulos et al., 2017). Consequently, the vast majority of users report having encountered political ads on social media (Media Authority of North Rhine-Westphalia [MANRW], 2019). However, despite its importance, political microtargeting was not extensively discussed until 2015, when a former digital analyst at Cambridge Analytica leaked the company's illegitimate practices of extracting, using, and combining user data for targeting purposes. Then, microtargeting became a standard campaign practice—especially in the US—despite being debated controversially (e.g., Donald Trump's team invested 44 million dollars, and Hillary Clinton's team 28 million dollars in digital advertising during the 2016 presidential campaign; Frier, 2018). Digital political advertising and microtargeting can be observed in Europe as well: During the German federal election campaign in 2017 and the European election campaign in 2019, German political parties invested in digital advertising on Facebook and Google. In 2017, the Left (*Die Linken*) invested 450,000 euros, the Greens (*Die Grünen*) two million euros, and the Free Democratic Party (FDP) 500,000 euros (Scherfig, 2017). In 2019, German parties invested up to 558,001 euros in digital advertising, e.g., the Christian Democratic Union (CDU) spent 296,001 euros on Facebook ads and 261,200 euros

on digital advertising on Google (Hegelich & Medina Serrano, 2019). An analysis of microtargeting strategies showed that the Greens reached more women than men and that the Social Democratic Party (SPD) and FDP tended to reach people aged 25 to 44, indicating that the parties employed demographic targeting strategies (Hegelich & Medina Serrano, 2019). Hence, although political parties in the US invest more resources in terms of time and money in political microtargeting than political parties in Germany, and despite the fact that (a) the European Union's restrictive General Data Protection Regulation (GDPR) regulates data collection, storage, and usage in Germany (Zuiderveen Borgesius et al., 2018); and (b) an even more far-reaching code of conduct was recently co-developed by the digital industry and the European Commission (2019), political microtargeting is a hot topic in Germany that has evolved in parallel with technological developments in this field.

In fact, a large majority of the public is concerned: 62% of US citizens say that political targeting is not acceptable (Smith, 2018), while 89% of Germans demand more transparent labeling and regulation of political ads (MANRW, 2019), demonstrating the relevance of analyzing and clarifying targeting processes and their implications for users. In addition, voters view the violation of their privacy as problematic and are afraid of losing control over their personal data (MANRW, 2019), which can be a serious problem from a psychological perspective (see Section 5).

The goal of our work is to theoretically examine the relations between microtargeting and online privacy in the context of elections and social media affordances. We aim to contribute to current research on political microtargeting by providing an in-depth understanding of what social media users need and expect in terms of their individual online privacy. We elaborate on how online privacy behavior might evolve over time—from the initial assessment of one's goals when using social media, to exposure to political targeting, subsequent privacy considerations and behaviors—by analyzing previous research in the field of online privacy and political behavioral targeting and drawing upon the social media privacy model (Trepte, 2020; cf. Figure 1).

2. Political Targeting: The Relevance of a Psychological Perspective

Political microtargeting is a specific kind of campaign tool. As part of microtargeting, behavioral (e.g., website visits), sociodemographic (e.g., gender, age), network (e.g., communication partners), and meta data (e.g., time and place of a message) are collected, analyzed, and processed (Dobber et al., 2019; Papakyriakopoulos et al., 2017). This data is used to identify groups of similar users (Queck, 2018). These groups are then exposed to messages tailored to their assumed needs and preferences (Beer et al., 2019; Bol et al., 2020). This kind of collected *behavioral* data is often enriched and aggregated with *psychometric* data, making it possible to match adverts to users' personality, which can further increase persuasiveness and influence actual behavior (e.g., 50% more purchases of a product after matching the appearance of a product ad to users' personality; see Matz et al., 2017). Psychometric measures are either extracted from paralinguistic traits or provided by the users themselves. For example, users may actively fill out personality tests on Facebook (e.g., myPersonality App), with which companies connect different kinds and sources of data (see also Kosinski et al., 2013). Hence, behavioral political targeting is oftentimes combined with psychological targeting. In this article, we refer to both kinds of targeting as well as their combination as (political) microtargeting.

Microtargeting is deployed at the back-end by political parties who strive to inform, steer, and persuade potential voters. However, this is not noticeable to users at the front-end. Oftentimes, targeted information is perceived as conventional social media information or even as independent news. Only one third of social media users are aware that political targeting takes place (Dobber et al., 2018). However, even if users are aware of the practice of political targeting, they are not able to completely shield their posts and profiles from psychological or behavioral profiling. Thus, they may have to deal with the potentially uncomfortable sentiment of being objectified and assigned to a certain cluster.

By identifying target groups, it becomes possible to address users' concrete political attitudes, needs, and fears (Queck, 2018). For political parties, the advantages of microtargeting lie in the higher probability of addressing voters' specific expectations, in resource efficiency, and in staying competitive with other parties (König, 2020; Zuiderveen Borgesius et al., 2018). This approach was particularly evident during Barack Obama's 2008 election campaign, when the campaign team analyzed data sets from around 150 million people and divided them into interest groups that could be specifically targeted through various channels—such as email, social media advertisements, and home visits—although this example received little public attention at the time (Aaker & Chang, 2009). Then, after Donald Trump became US president in 2017, it was revealed that the British political consulting company Cambridge Analytica used Facebook users' data to create psychometric personality profiles for over 50 million individuals that were used for microtargeting purposes during Trump's campaign (Beuth & Horchert, 2018). In spring 2018, whistleblower and former Cambridge Analytica employee Christopher Wylie leaked background information on how Cambridge Analytica had set up an extensive system of websites and blogs to target voters with precisely tailored information (Baetz & Zilm, 2018).

Academics, lawyers, activists, and journalists (Bennett & Lyon, 2019; Potthast, 2019; Rebiger, 2018; Reihs, 2019) have criticized political microtargeting as intrusive and manipulative, because targeted users are often unaware of their exposure to this campaign

strategy. Despite restrictions regarding the processing of personal data in the EU (e.g., due to the GDPR), Twitter's official prohibition on political microtargeting (Fanta, 2018), and contextual limitations such as budget or party structures (Kruschinski & Haller, 2017), users provide a great deal of data on social media that can be used to target them *despite* these legal and contextual restrictions (Papakyriakopoulos et al., 2017). Subsequently, social network sites such as Facebook still present adverts and content to their users based on their *likes*, interests, and provided information, which is sometimes related to political topics (Facebook Help Center, 2021). At the same time, politicians increasingly use social media to directly address potential voters (Hegelich & Shahrezaye, 2015). Dobber et al. (2019, p. 7) summarize: "In sum, Europe's privacy laws do not categorically prohibit microtargeting. Still, Europe's privacy laws make microtargeting more difficult than in, for instance, the US."

While users' concerns are evident, the effects of political targeting are ambiguous. Research on the direct effect of political targeting on "outcome" variables such as voting behavior is scarce, and studies reveal a heterogeneous picture.

We suggest three main reasons for why it is difficult to find a linear relationship between exposure to political microtargeting and political participation outcomes. First, it is questionable whether the actions that facilitate microtargeting (e.g., tracking, tracing, or buying user data) provide information that is not already available through traditional sources like voter rolls and past voting behavior (Hersh, 2015). Second, potential effects of microtargeting on political outcomes can only be measured clearly if microtargeting presents unique information the user is not also exposed to through other channels. For example, if a social media user is targeted via both canvassing and microtargeting, the differential effect of microtargeting can only be measured if different information is conveyed through these two kinds of campaigning. Third, targeting is oftentimes applied coarsely. For example, German parties usually target based on broad categories such as gender and region (Hegelich & Medina Serrano, 2019). Such categories might not have strong effects on political participation.

The belief that targeting might have no direct effect on voters' decisions may be a source of relief—but should it be? We doubt this. Instead, we pose the overarching question of what exactly the "outcome" of targeting practices is. Therefore, it is important to significantly broaden our understanding of this "outcome." We seek to consider not only the narrow behavioral outcome, but also the question of whether and how political microtargeting affects social media users' and therefore voters' subjective self-perception of informational self-determination as well as perceived privacy and privacy concerns. Users' privacy perceptions may in turn also mediate their voting behavior. In other words, a lacking linear relationship between exposure to political microtargeting and political behavior could stem from a missing link to privacy mechanisms. Since identifying the psychological processes underlying how microtargeting is perceived with regard to privacy would require comprehensive theoretical and empirical investigations that go beyond the scope of a single journal article, we decided to begin working on this task theoretically. While previous theoretical work has analyzed political microtargeting and its potential consequences from a normative and communication science perspective (Haller & Kruschinski, 2020; König, 2020), a psychological perspective is still missing.

3. Users' Assessment of Privacy and Political Microtargeting

The concept of online privacy has been researched and defined in many distinct disciplines, such as communication science, psychology and sociology, applying descriptive, empirical, and normative perspectives (Masur, 2019; Schäwel, 2019; Sevignani, 2016; Trepte & Reinecke, 2011). Originally, privacy was normatively defined as the human "right to be let alone" (Warren & Brandeis, 1890, p. 193). The level of access an individual feels comfortable with and individual communication goals are crucial for privacy decisions (Dienlin, 2014; Trepte, 2020): In an "initial assessment" (cf. Figure 1, first row), users evaluate their individual level of access (e.g., high access through the disclosure of personal information like gender or political attitudes) and consider it in light of their individual communication goals (e.g., letting others know their personal information). The level of access to the self represents a pivotal component of personal privacy. Westin (1967, p. 7) defined privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others." Social media privacy is always defined in reference to certain others (e.g., institutions, people, or entities; Trepte, 2020). When writing a post to share on Twitter or Facebook, different people might assess the situation in different ways. One person might evaluate their privacy with regard to the service provider, while another person might consider not the provider but rather their followers. Hence, the service provider is an object of privacy assessments for the first person, but not the second. This level of access determines the decision an individual will make in a particular privacy-relevant situation. It is inter-individually different, has some intra-individual stability, and is context-dependent.

4. Social Media Boundary Conditions: Content and Affordances

Privacy regulation behavior is influenced by the social media context and its "boundary conditions" (cf. Figure 1, second row). Social networking sites such as Twitter or Facebook constitute important sources of political information. Social media as a context must therefore be

understood comprehensively in order to understand privacy, and in turn, the relevance of the mechanisms of control, trust, and communication (Nissenbaum, 2010). Hence, we will consider social media affordances that shape users' individual perceptions of privacy, control, trust, and communication options. The term affordance (Gibson, 2014) captures the idea that the environmental properties of an entity are perceived and experienced differently by different people. Here, the environmental properties would be social media affordances such as persistence, and the entity would be the social media site itself. While using social media, certain properties are actively used and emphasized, while others are overlooked (Trepte, 2015). Due to the importance of social media affordances for users' perceptions of online privacy, we will consider them in our theoretical investigation of the privacy-relevant context of political microtargeting.

We concentrate on the four affordances addressed in the social media privacy model (Trepte, 2020): anonymity, editability, association, and persistence (Evans et al., 2017; Treem & Leonardi, 2012). Other affordances discussed in the literature are visibility, navigability, interactivity (Evans et al., 2017), and paralinguistic affordances (Hayes et al., 2016).

In the social media context, "anonymity" means that other users, institutions, and companies do not know the source of a message (Evans et al., 2017), which can increase senders' perception of privacy. However, using social media anonymously is rare, because anonymity reduces contacts and social support, which are the main benefits of using social media (Rainie et al., 2013). Additionally, users leave data traces while searching the internet even when they appear anonymous to their online contacts. Companies specialized in collecting and aggregating data traces can create profiles that make it possible to identify the person and connect this data to their online personas. Therefore, anonymity with respect to companies and institutions is not guaranteed on social media (Trepte, 2020). Parties can use these data traces to optimize algorithms for microtargeting, for instance, by linking geospatial data with user interests (Dobber et al., 2019) or expressions of political views and opinions (e.g., through *likes*), which in turn allows them to better match the advertisements presented to users (Papakyriakopoulos et al., 2018). Furthermore, by joining a social media site and by accepting its terms and conditions, users (have to) consent to the further processing and use of their data for commercial or other purposes (Papakyriakopoulos et al., 2018) and thus implicitly—and often unawarely—also consent to political microtargeting. If users encounter tailored political advertisements and recognize that these are based on their private information, they might feel that their anonymity, and thus an essential part of their privacy, has been threatened. The question is, how would this impression affect the perception of potential privacy regulation mechanisms (see also Section 5)? Since political parties can acquire metadata that allows for targeted and personalized political advertisements from data broker companies; control cannot be used as a possible privacy mechanism without abandoning the use of the internet or a certain app (Dobber et al., 2018; Papakyriakopoulos et al., 2018). Legal norms, in turn, do not yet have a firm grip on targeting practices that are seen critically and even as illegitimate by activists and scientists. Therefore, legal norms offer only limited protection, meaning that the only available privacy mechanism is trust that one's data will be used responsibly. More interested and active voters might also consider communication with political parties as a privacy mechanism; however, there are no firm research results on whether social media users take advantage of this deliberative option. The mechanisms of control, trust, norms, and communication will be explained in more detail in Section 5.

"Editability" gives users the opportunity to adjust with whom they communicate in which manner (Treem & Leonardi, 2012) by modifying their posts or applying specific privacy settings (e.g., blocking people or designating the audience for specific posts). Users can also manage their self-presentation via social media functionalities that afford editability (e.g., editing a photo) and in this way regulate their privacy. For instance, a user might blur a photo or crop a picture in which she participates in a (political) demonstration to only show certain details. Messages or one's profile name can also be edited by means of functionalities that afford editability. Editing one's profile name can also be related to the anonymity affordance, such as when changing one's real name into a fake name.

A central affordance of social media is the "association" between different interaction partners (Ellison & boyd, 2013; Treem & Leonardi, 2012), which seldom allows a person to maintain control over the subjective regulation of privacy and therefore might reduce perceived privacy. The prevalence of the association affordance can influence users' number of contacts, quantity of interactions, network structures, the visibility of visited events and locations, group memberships, or pictures on Facebook. Fox and McEwan (2017) showed that associations between social media users negatively affect their sense of control. To counteract this, users can rely on alternative privacy mechanisms such as trust (see Section 5). Studies demonstrate that users with stronger associations trust their social media friends and acquaintances more (Hofstra et al., 2016). However, trust in Twitter and Facebook is comparatively low when it comes to political issues (Paus & Börsch-Supan, 2019). If a discrepancy between high trust in people and low trust in platforms or the source of a political advertisement is detected, the privacy mechanism of social and legal norms gains relevance. Users recall that social media platforms must adhere to social and legal norms ensuring that their personal data is used only in an acceptable, non-invasive way that follows data protection laws.

"Persistence" refers to the permanency and replicability of online statements and content (boyd, 2014). Data remains available over unknown periods of time and can be accessed by different and unexpected users, e.g., (future) employers (Evans et al., 2017; Treem & Leonardi, 2012) or political parties—although the GDPR requires "storage limitation" by stipulating that "personal data may not be retained for unreasonably long periods" (GDPR, 2018). Users do in fact see the lack of control over their personal information that results from the persistence of online information as a problem for their privacy (Teutsch et al., 2018).

According to the social media privacy model, the outlined affordances (i.e., social media boundary conditions) interact with users' ideal level of access to provide to their personal information and their communication goals (i.e., initial assessment), which in turn shape users' expectations about how they can react to potential privacy harms by using prevalent privacy mechanisms to regulate their privacy (cf. Figure 1, first to third row).

5. Available Privacy Mechanisms and the Experience of Privacy in the Context of Targeting

Control has long been and still is an essential part of the definition and understanding of privacy (Altman, 1974; Burgoon, 1982; Petronio, 2002). The basic assumption is that the amount of perceived privacy and corresponding informational self-determination depends on the control people perceive to have over their private information. Thus, more control equals more privacy. However, this linear relationship has not been supported by research so far (see Trepte, 2020). A decreasing amount of control does not necessarily mean having no or limited privacy. If a social media user trusts a provider like Twitter or a political party to handle their personal data responsibly, they rely on trust as a privacy mechanism, resulting in a perception of individual privacy. Changes in the person's privacy perceptions of Twitter's or the political party's trustworthiness will influence their privacy regulatory behavior. Hence, users are not restricted to one privacy mechanism, but can consider different mechanisms depending on their current availability and perceived impact. In the following paragraphs, we will describe each of the "available privacy mechanisms" (cf. Figure 1, third row) in detail.

Informational "control" means the ability to hold information back (Crowley, 2017) and the user's ability to freely choose whether to disclose certain information (e.g., political attitudes) or not (Tavani, 2007). In contrast to other privacy mechanisms such as communication, the individual him- or herself steers control behavior (i.e., "egocentric regulation," cf. Figure 1). Users can exercise control by anonymizing or editing their posts or profiles (e.g., by providing fake information). We assume that users who *feel* to be in control also *experience* more privacy than those who have no access to this mechanism. Users exercising control should therefore feel less susceptible to being targeted with personalized advertisements. However, control is hard to achieve and is only one aspect influencing the perception of privacy (Trepte, 2020). If users feel a lack of control, trust in the communication partner (e.g., a political party) becomes relevant. Trust in an online shop, for instance, is associated with a lower perception of risk regarding the disclosure of personal information (Gurung & Raja, 2016). Accordingly, if users have a strong feeling of trust towards a political party, they might feel a lower need for control in order to protect their privacy against this party's microtargeting practices. If, on the other hand, the user receives political advertising from a political party they highly mistrust, this could reduce their experience of privacy and increase the relevance of control or alternative privacy mechanisms.

"Interpersonal communication" is understood here as interactions between users, or between users on the one hand and institutions or companies on the other. For example, users might discuss among one another whether or not certain (political) opinions should be shared on Facebook. Furthermore, if the current privacy situation is not satisfactory, e.g., because users' social contacts might leak private information or no laws to protect privacy exist, users can engage in interpersonal communication with peers, companies or political parties to change the situation. In the case of political targeting, when privacy-invasive practices are recognized or gain public attention, such communication might take the form of problem-oriented interactions with peers or parties. We assume that users who anticipate that they can get in touch with the political party experience more privacy than users without access to such interpersonal communication. On the other hand, users sometimes feel powerless when communicating with companies about data deletion or terms of consent (Teutsch et al., 2018). Thus, interpersonal communication is not always possible or expedient for privacy regulation.

Instead, "trust" as the result of previous successful communication or adherence to norms can serve as a privacy regulatory mechanism. Trust is defined as the expression of balanced communication and the anticipation that normatively correct behavior will be implemented (Green, 2007). Trust and communication influence each other in the sense that a minimal level of trust is needed for communication, and trust can increase as a result of a successful communication (Saeri et al., 2014). Henderson et al. (2016) found that engaging in communication based on collectively established communication norms can predict trust in virtual teams. Common norms in online communities have a direct influence on users' trust in community members (Blanchard et al., 2011). Furthermore, trust can even reduce privacy concerns (Taddei & Contena, 2013), suggesting that people might be less concerned about a political party or social media site they trust. However, Lankton et al. (2012) demonstrated that trust cannot be a full substitute for control. A study on political microtargeting conducted

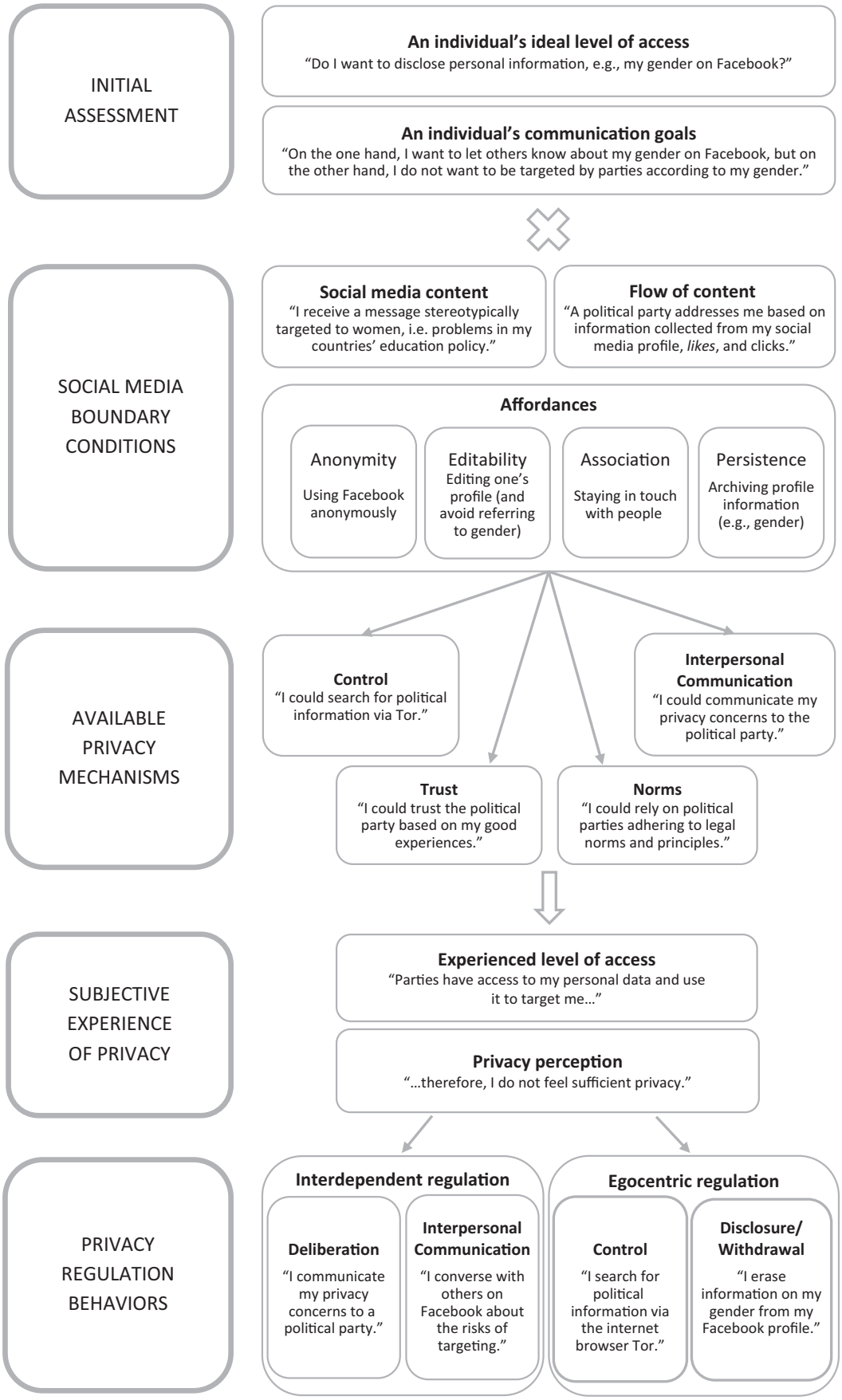


Figure 1. The social media privacy process as experienced by users confronted with political behavioral targeting: From the initial assessment of individual needs to the behaviors and choices ultimately executed to regulate one's privacy. Source: Trepte (2020).

in the Netherlands found that a political party post that was clearly marked as an advertisement had no effect on trust in this party. Still, users were less willing to share this post that they knew was advertising by a political party (Kruikemeier et al., 2016). The study's authors conclude that users resist sharing political messages that they know to be personalized political ads (Kruikemeier et al., 2016).

Next, social and legal "norms" play an important role in privacy regulation. If social media users have the feeling that the existing norms in place protect them sufficiently, their experience of privacy should be correspondingly high. Social norms can evolve either as a result of observations, i.e., what do I *see* others doing (Lewis, 2011), or as a result of assumptions, i.e., what do I *believe* others are doing regarding their privacy (Spottswood & Hancock, 2017) or protecting others' privacy (e.g., third parties follow the law in order to protect users' privacy). While the influence of social norms on social media users' privacy has already been investigated (Utz & Krämer, 2009), there has been no research into how legal norms and regulations influence privacy behaviors in the context of political microtargeting. The perception and awareness of legal norms may be associated with and affected by current law, which demands "lawfulness, fairness, and transparency," "purpose limitation," "data minimization," "accuracy," "storage limitation," "integrity and confidentiality," and "accountability" when processing personal data (GDPR, 2018). Thus, users may rely on third parties adhering to these principles and consequently following not only the law but also legal norms of transparency and fairness. In such a case, they should experience more privacy. This is also related to the trust mechanism. Users who trust political parties probably trust them to follow these principles as well.

This process of users' initially assessing (communication) goals, perceiving social media affordances and available privacy mechanisms, and arriving at a subjective experience of privacy and subsequent privacy behavior (which is explained in more detail in Section 6) is visualized in the social media privacy model (Trepte, 2020), which we enriched with concrete examples regarding political targeting.

6. Users' Privacy Regulation in the Face of Targeting

According to the social media privacy model (Trepte, 2020), the individual perception of privacy can vary depending on available privacy mechanisms and in turn lead to different regulatory behaviors, namely interdependent (deliberation/interpersonal communication) or individual (control/disclosure or withdrawal) regulation. This means, for example, that if the privacy mechanism control is available and the experienced level of privacy is low, no deliberative communication is performed. Instead, control would be exercised as a regulating behavior (e.g., using the Tor internet browser to search for political information; cf. Figure 1, last row). However, when the control mechanism is not available, the availability of alternative mechanisms becomes relevant, which in turn affects users' perceived privacy and regulatory behaviors. Consequently, either interdependent (e.g., negotiating with third parties or communicating with users) or individual (e.g., limiting personal disclosures) regulatory strategies are enacted (cf. Figure 1, last row). Regulation of behavior should be increasingly implemented the lower the perception of privacy is (e.g., by rejecting specific cookies or services). However, a current survey conducted in Germany revealed that 20% of 1,065 participants did not use any settings to actively protect their privacy (i.e., privacy regulation) during the past year, although 82% expressed concerns about their privacy (Kozyreva et al., 2020).

Still, users might have a limited perception of political targeting because it is designed to not be perceived by users. Hence, privacy behaviors are not a direct reaction to exposure to political targeting, but presumably a reaction to some general (perhaps limited) knowledge of political targeting practices, associated attitudes, and expectations about which privacy mechanisms might successfully combat these kinds of practices. We will now refer to how future research and debates may address this particular circumstance.

7. Discussion

Privacy is a higher-order need and as such oftentimes remains in the background while predominantly serving the fulfilment of other needs, such as participation in democratic processes (Trepte & Masur, 2017). The need for privacy particularly comes into play and causes friction when it is unfulfilled. This is the case when it comes to political targeting. Based on the social media privacy model, we propose considering users' assessment of access and communication goals; social media boundary conditions, including prevalent affordances; available privacy mechanisms; subjective experiences of privacy; and potential interdependent or egocentric privacy regulation behaviors in the context of microtargeting processes. As such, individual privacy regulation becomes visible and is not modeled simply as disclosure or withdrawal, but also as a form of political action. Accordingly, there are interdependent regulation strategies like interpersonal communication and deliberation, in which the individual communicates conscious decisions about privacy and levels of access.

Our theoretical analysis showed that the effects of political microtargeting are determined by users' need for privacy and their assessment of the social media context in light of this need. As an analytical result of our theoretical discussion, we present three propositions for future research.

Our first proposition is to further consider the complexity of the social media context, users' perception of it, and its affordances. It is important to understand what kinds of targeting users are exposed to via which

channels (i.e., context). Then, not only exposure, but also users' perception of the context and information processing must be measured (i.e., perceptions). Finally, the perceived social media functionalities and boundaries must be evaluated (i.e., affordances). Exposure to political targeting does not necessarily mean that users are aware of it. Indeed, only one-third of users are even aware that targeting takes place (Dobber et al., 2018). Thus, only when we know what users experience can we understand how this affects privacy and informational self-determination, as well as ultimately the dependent outcome variable. This consideration has a crucial impact on methodology, which will be discussed in the third proposition (see also Bol et al., 2020).

Our second proposition is to bring privacy to the fore and to understand users' privacy perception and evaluation as underlying psychological processes that influence or even mediate the consequences of microtargeting (i.e., the outcome). Our theoretical analysis showed that the level of individual privacy is a core aspect of self-determination and a precondition for valuable online experiences, which in turn affect numerous decisions, actions, and behaviors. This is of interest with regard to the effectiveness and potential intrusiveness of political microtargeting strategies—from both political parties' as well as researchers' point of view.

Our third proposition is to conduct research that aligns with the ethical principles formulated for social science. Our theoretical analysis showed that users feel uncomfortable being observed, evaluated, and targeted. Another ethical concern is that most users are not aware *that* targeting takes place (Dobber et al., 2018). Even if they are aware of it, they cannot influence *what* data is being seen and used, and *when and how* this data is reflected back to them in the form of targeted advertising (Matz et al., 2017, 2020; Noecker et al., 2013).

This ethical criticism is closely connected to empirical possibilities and research practices in the field of political behavioral targeting (e.g., tracking or tracing user data). In future research on targeting and privacy, it will be important to rely on observational studies and computational approaches to gather useful data, for which ethical boundaries will pose one of the most serious challenges. One reason why observational measures are needed is because users' self-reports are often not reliable in the context of political microtargeting: Users have difficulties identifying situations in which they were targeted and how they felt. Therefore, more advanced observational and experimental research designs are needed (Bol et al., 2020). In tracking studies, for instance, participants would install a browser plug-in on their computer or smartphone to log their online behavior and allowing to draw conclusions based on their clicks (which might in turn have been guided by specific ads). However, this method does also not allow for investigating users' experience of privacy. Thus, even more comprehensive and intelligent measures are required, e.g., combining users' log data and self-reports to identify the moment and source of targeting and initiate a direct request for a user self-report. The crucial point with such tracking or tracing methods is that they are based on similar mechanisms as microtargeting (i.e., observing users and specifically targeting them based on these observations). Consequently, user-centered research on the effects of political microtargeting presumes certain ethical standards that should also hold in the field of targeting resarch itself.

8. Conclusion

The goal of this theoretical investigation of privacy and political microtargeting on social media was to derive propositions for analyzing political microtargeting in a way that considers users' privacy needs, relevant political outcomes, and ethical implications. We conclude by highlighting the importance of: (a) considering the complexity of the social media context and its affordances as well as users' perceptions of these, (b) positioning privacy as a relevant research topic by understanding how users' privacy experiences influence and mediate the outcome of microtargeting, and (c) conducting research in accordance with ethical guidelines in order to establish research practices that meet the standards we as scholars set for the social media industry.

Acknowledgments

We would like to thank Jennifer Müller for her support and valuable feedback.

Conflict of Interests

The authors declare no conflict of interests.

References

Aaker, J., & Chang, V. (2009). Obama and the power of social media and technology. *The European Business Review*, 16–32. https://jaaker.people.stanford.edu/sites/g/files/sbiybj2966/f/obamaandthepowerofsocialmediafinal2009.pdf

Altman, I. (1974). Privacy: A conceptual analysis. In S. T. Margulis (Ed.), *Man-environment interactions: Evaluations and applications* (pp. 3–28). Dowden, Hutchinson & Ross.

Baetz, B., & Zilm, K. (2018, April 10). *Daten ohne Schutz—Zuckerberg in Bedrängnis* [Data without protection—Zuckerberg in trouble]. Deutschlandfunk. https://www.deutschlandfunk.de/der-facebook-skandal-daten-ohne-schutz-zuckerberg-in.724.de.html?dram:article_id=415251

Beer, D., Redden, J., Williamson, B., & Yuill, S. (2019). *Landscape summary: Online targeting: What is online targeting, what impact does it have, and how can we maximise benefits and minimise harms?* Centre for Data Ethics and Innovation. http://orca.cf.ac.uk/126114

Bennett, C. J., & Lyon, D. (2019). Data-driven elections: Implications and challenges for democratic societies. *Internet Policy Review*, *8*(4). https://doi.org/10.14763/2019.4.1433

Beuth, P., & Horchert, J. (2018, March 20). Was treibt eigentlich Cambridge Analytica? [What is Cambridge Analytica doing?]. *Spiegel*. https://www.spiegel.de/netzwelt/netzpolitik/cambridge-analytica-das-steckt-hinter-der-datenanalyse-firma-a-1198962.html

Blanchard, A. L., Welbourne, J. L., & Boughton, M. D. (2011). A model of online trust. *Information, Communication & Society*, *14*(1), 76–106. https://doi.org/10.1080/13691181003739633

Bol, N., Strycharz, J., Helberger, N., van de Velde, B., & de Vreese, C. H. (2020). Vulnerability in a tracked society: Combining tracking and survey data to understand who gets targeted with what content. *New Media & Society*, *22*(11), 1996–2017. https://doi.org/10.1177/1461444820924631

boyd, d. (2014). *It's complicated: The social lives of networked teens*. Yale University Press.

Burgoon, J. K. (1982). Privacy and communication. *Communication Yearbook*, *6*(1), 206–249. https://doi.org/10.1080/23808985.1982.11678499

Crowley, J. L. (2017). A framework of relational information control: A review and extension of information control research in interpersonal contexts. *Communication Theory*, *27*(2), 202–222. https://doi.org/10.1111/comt.12115

Dienlin, T. (2014). The privacy process model. In S. Garnett, S. Halft, M. Herz, & J. M. Mönig (Eds.), *Medien und Privatheit* [Media and privacy] (pp. 105–122). Karl Stutz.

Dobber, T., Ó Fathaigh, R., & Zuiderveen Borgesius, F. J. (2019). The regulation of online political microtargeting in Europe. *Internet Policy Review*, *8*(4). https://doi.org/10.14763/2019.4.1440

Dobber, T., Trilling, D., Helberger, N., & de Vreese, C. H. (2018). Spiraling downward: The reciprocal relation between attitude toward political behavioral targeting and privacy concerns. *New Media & Society*, *21*(6), 1212–1231. https://doi.org/10.1177/1461444818813372

Ellison, N. B., & boyd, d. (2013). Sociality through social network sites. In W. H. Dutton (Ed.), *The Oxford handbook of internet studies* (pp. 151–172). Oxford University Press.

European Commission. (2019). *Guidelines on ethical standards for the participation of the members of the european commission in the election campaign*. https://ec.europa.eu/info/sites/info/files/guidelines_election_campaign_en.pdf

Evans, S. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2017). Explicating affordances: A conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication*, *22*(1), 35–52. https://doi.org/10.1111/jcc4.12180

Facebook Help Center. (2021). *How does Facebook decide which ads to show me?* Facebook. https://www.facebook.com/help/516147308587266/how-ads-work-on-facebook/?helpref=hc_fnav

Fanta, A. (2018). *EU-kommission: 80 Prozent der Europäer wollen wissen, wer für politische Werbung im Netz zahlt* [EU Commission: 80 percent of Europeans want to know who pays for political advertising on the web]. Netzpolitik. https://netzpolitik.org/2018/eu-kommission-80-prozent-der-europaeer-wollen-wissen-wer-fuer-politische-werbung-im-netz-zahlt

Fox, J., & McEwan, B. (2017). Distinguishing technologies for social interaction: The perceived social affordances of communication channels scale. *Communication Monographs*, *84*(3), 298–318. https://doi.org/10.1080/03637751.2017.1332418

Frier, S. (2018, April 3). Trump's campaign said it was better at Facebook. Facebook agrees. *Bloomberg*. https://www.bloomberg.com/news/articles/2018-04-03/trump-s-campaign-said-it-was-better-at-facebook-facebook-agrees

General Data Protection Regulation. (2018). *Art. 5: Principles relating to processing of personal data*. https://gdpr-info.eu/art-5-gdpr

Gibson, J. J. (2014). *The ecological approach to visual perception*. Routledge.

Green, M. C. (2007). Trust and social interaction on the internet. In A. N. Joinson, K. Y. A. McKenna, T. Postmes, & U. D. Reips (Eds.), *The Oxford handbook of Internet psychology* (pp. 43–52). Oxford University Press.

Gurung, A., & Raja, M. K. (2016). Online privacy and security concerns of consumers. *Information and Computer Security*, *24*(4), 348–371. https://doi.org/10.1108/ICS-05-2015-0020

Haller, A., & Kruschinski, S. (2020). Politisches Microtargeting: Eine normative Analyse von datenbasierten Strategien gezielter Wähler_innenansprache [Political microtargeting: A normative analysis of data-based strategies of targeted voters]. *ComSoc Communicatio Socialis*, *53*(4), 519–530. https://doi.org/10.5771/0010-3497-2020-4-519

Hayes, R. A., Carr, C. T., & Wohn, D. Y. (2016). One click, many meanings: Interpreting paralinguistic digital affordances in social media. *Journal of Broadcasting & Electronic Media*, *60*(1), 171–187. https://doi.org/10.1080/08838151.2015.1127248

Hegelich, S., & Medina Serrano, J. C. (2019). *Microtargeting in Deutschland bei der Europawahl 2019* [Microtargeting in Germany in the 2019 European elections]. Media Authority of North Rhine-Westphalia. https://www.blm.de/files/pdf2/studie_microtargeting_deutschlandeuropawahl2019_hegelich-1.pdf

Hegelich, S., & Shahrezaye, M. (2015). The communication behavior of German MPs on Twitter: Preaching to the converted and attacking opponents. *Euro-*

pean Policy Analysis, *1*(2), 155–174. https://doi.org/10.18278/epa.1.2.8

Henderson, L. S., Stackman, R. W., & Lindekilde, R. (2016). The centrality of communication norm alignment, role clarity, and trust in global project teams. *International Journal of Project Management*, *34*(8), 1717–1730. https://doi.org/10.1016/j.ijproman.2016.09.012

Hersh, E. D. (2015). *Hacking the electorate: How campaigns perceive voters*. Cambridge University Press. https://doi.org/10.1017/CBO9781316212783

Hofstra, B., Corten, R., & van Tubergen, F. (2016). Understanding the privacy behavior of adolescents on Facebook: The role of peers, popularity and trust. *Computers in Human Behavior*, *60*, 611–621. https://doi.org/10.1016/j.chb.2016.02.091

König, P. D. (2020). Why digital-era political marketing is not the death knell for democracy: On the importance of placing political microtargeting in the context of party competition. *Statistics, Politics and Policy*, *11*(1), 87–110. https://doi.org/10.1515/spp.2019--0006

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Scienses*, *110*(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Kozyreva, A., Herzog, S., Lorenz-Spreen, P., Hertwig, R., & Lewandowsky, S. (2020). *Artificial intelligence in the online environment: A representative survey on public opinion in Germany*. Max Planck Institute for Human Development. https://www.conpolicy.de/en/news-detail/artificial-intelligence-in-the-online-environment-a-representative-survey-on-public-opinion-in-germ

Kruikemeier, S., Sezgin, M., & Boerman, S. C. (2016). Political microtargeting: Relationship between personalized advertising on Facebook and voters' responses. *Cyberpsychology, Behavior and Social Networking*, *19*(6), 367–372. https://doi.org/10.1089/cyber.2015.0652

Kruschinski, S., & Haller, A. (2017). Restrictions on data-driven political micro-targeting in Germany. *Internet Policy Review*, *6*(4). https://doi.org/10.14763/2017.4.780

Lankton, N. K., McKnight, D. H., & Thatcher, J. B. (2012). The moderating effects of privacy restrictiveness and experience on trusting beliefs and habit: An empirical test of intention to continue using a social networking website. *IEEE Transactions on Engineering Management*, *59*(4), 654–665. https://doi.org/10.1109/TEM.2011.2179048

Lewis, K. (2011). The co-evolution of social network ties and online privacy behavior. In S. Trepte & L. Reinecke (Eds.), *Privacy online: Perspectives on privacy and self-disclosure in the social web* (pp. 91–110). Springer.

Masur, P. K. (2019). *Situational privacy and self-disclosure: Communication processes in online environments*. Springer.

Matz, S. C., Appel, R. E., & Kosinski, M. (2020). Privacy in the age of psychological targeting. *Current Opinion in Psychology*, *31*, 116–121. https://doi.org/10.1016/j.copsyc.2019.08.010

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(48), 12714–12719. https://doi.org/10.1073/pnas.1710966114

Media Authority of North Rhine-Westphalia. (2019). *Informationsverhalten bei Wahlen und politische Desinformation* [Information behavior in elections and political disinformation]. https://www.medienanstalt-nrw.de/fileadmin/user_upload/lfm-nrw/Service/Pressemitteilungen/Dokumente/2019/Praesentation_forsa_Desinformation_LFMNRW.pdf

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

Noecker, J., Ryan, M., & Juola, P. (2013). Psychological profiling through textual analysis. *Literary and Linguistic Computing*, *28*(3), 382–387. https://doi.org/10.1093/llc/fqs070

Papakyriakopoulos, O., Hegelich, S., Shahrezaye, M., & Medina Serrano, J. C. (2018). Social media and microtargeting: Political data processing and the consequences for Germany. *Big Data & Society*, *5*(2). https://doi.org/10.1177/2053951718811844

Papakyriakopoulos, O., Shahrezaye, M., Thieltges, A., Medina Serrano, J. C., & Hegelich, S. (2017). Social Media und Microtargeting in Deutschland [Social media and microtargeting in Germany]. *Informatik-Spektrum*, *40*(4), 327–335. https://doi.org/10.1007/s00287-017-1051-4

Paus, I., & Börsch-Supan, J. (2019). *Alles auf dem Schirm?* [Everything in mind?]. Vodafone. https://www.vodafone-stiftung.de/alles-auf-dem-schirm

Petronio, S. (2002). *Boundaries of privacy*. State University of New York Press.

Potthast, K. C. (2019). *Political Microtargeting—Zwischen Regulierungsbegehren und Ungewissheit* [Political microtargeting—Between regulatory desire and uncertainty]. Juwiss. https://www.juwiss.de/103-2019

Queck, S. (2018). *Microtargeting—Definition, Einsatz und Beispiele* [Microtargeting—Definition, use and examples]. Marconomy. https://www.marconomy.de/microtargeting-definition-einsatz-und-beispiele-a-739666

Rainie, L., Kiesler, S., Kang, R., & Madden, M. (2013). *Anonymity, privacy, and security online*. Pew Research Center. http://www.pewinternet.org/2013/09/05/anonymity-privacy-and-security-online

Rebiger, S. (2018). *Offener brief: Europäische Parteien sollen auf Microtargeting verzichten* [Open letter: European parties should renounce microtargeting]. Netzpolitik. https://netzpolitik.org/2018/offener-

brief-eu-parteien-sollen-auf-microtargeting-verzichten

Reihs, V. (2019). *Politisches Microtargeting in Deutschland: Ich sehe was, was du nicht siehst* [Political microtargeting in Germany: I see something you don't]. Politik-Digital. https://politik-digital.de/news/politisches-microtargeting-in-deutschland-ich-sehe-was-was-du-nicht-siehst-155876

Saeri, A. K., Ogilvie, C., La Macchia, S. T., Smith, J. R., & Louis, W. R. (2014). Predicting Facebook users' online privacy protection: Risk, trust, norm focus theory, and the theory of planned behavior. *The Journal of Social Psychology*, *154*(4), 352–369. https://doi.org/10.1080/00224545.2014.914881

Schäwel, J. (2019). *How to raise users' awareness of online privacy* [Doctoral dissertation, University Duisburg-Essen]. DuEPublico2. https://doi.org/10.17185/duepublico/70691

Scherfig, L. (2017, January 26). *Wie die Parteien 2017 in den digitalen Wahlkampf ziehen* [How the parties are moving into the digital election campaign in 2017]. Berliner Morgenpost. https://www.morgenpost.de/politik/article209403515/Wie-die-Parteien-2017-in-den-digitalen-Wahlkampf-ziehen.html

Sevignani, S. (2016). *Privacy and capitalism in the age of social media*. Routledge.

Smith, A. (2018). *Algorithms in action: The content people see on social media*. Pew Research Center. https://www.pewresearch.org/internet/2018/11/16/algorithms-in-action-the-content-people-see-on-social-media

Spottswood, E. L., & Hancock, J. T. (2017). Should I share that? Prompting social norms that influence privacy behaviors on a social networking site. *Journal of Computer-Mediated Communication*, *22*(2), 26. https://doi.org/10.1111/jcc4.12182

Taddei, S., & Contena, B. (2013). Privacy, trust and control: Which relationships with online self-disclosure? *Computers in Human Behavior*, *29*(3), 821–826. https://doi.org/10.1016/j.chb.2012.11.022

Tavani, H. T. (2007). Philosophical theories of privacy: Implications for an adequate online privacy policy. *Metaphilosophy*, *38*(1), 1–22. https://doi.org/10.1111/j.1467-9973.2006.00474.x

Teutsch, D., Masur, P. K., & Trepte, S. (2018). Privacy in mediated and nonmediated interpersonal communication: How subjective concepts and situational perceptions influence behaviors. *Social Media + Society*, *4*(2), 1–14. https://doi.org/10.1177/2056305118767134

Treem, J. W., & Leonardi, P. M. (2012). Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association. *Communication Yearbook*, *36*(1), 143–189. https://doi.org/10.1080/23808985.2013.11679130

Trepte, S. (2015). Social media, privacy, and self-disclosure: The turbulence caused by social media's affordances. *Social Media and Society*, *1*(1), 1–2. https://doi.org/10.1177/2056305115578681

Trepte, S. (2020). The social media privacy model: Privacy and communication in the light of social media affordances. *Communication Theory*, *19*(4), 1–22. https://doi.org/10.1093/ct/qtz035

Trepte, S., & Masur, P. K. (2017). Need for privacy. In V. Zeigler-Hill & T. K. Shakelford (Eds.), *Encyclopedia of personality and individual differences*. Springer. https://doi.org/10.1007/978-3-319-28099-8_540-1

Trepte, S., & Reinecke, L. (Eds.). (2011). *Privacy online: Perspectives on privacy and self-disclosure in the social web*. Springer.

Utz, S., & Krämer, N. C. (2009). The privacy paradox on social network sites revisited: The role of individual characteristics and group norms. *Journal of Psychosocial Research on Cyberspace*, *3*(2), Article 2. http://cyberpsychology.eu/view.php?cisloclanku=2009111001&article=2

Warren, S. D., & Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review*, *4*(5), 193–220.

Westin, A. F. (1967). *Privacy and freedom*. Atheneum.

Zuiderveen Borgesius, F. J., Möller, J., Kruikemeier, S., Ó Fathaigh, R., Irion, K., Dobber, T., Bodo, B., & de Vreese, C. (2018). Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review*, *14*(1), 82. https://doi.org/10.18352/ulr.420

About the Authors

Johanna Schäwel (PhD) is a postdoctoral researcher at the Institute of Communication Science: Media Psychology at the University of Hohenheim (Germany). Her research focuses on causes and consequences of social media use, especially users' (privacy) needs and competencies.

Regine Frener (MA, University of Mannheim) is a PhD candidate at the Institute of Communication Science: Media Psychology at the University of Hohenheim (Germany). She is interested in gender studies and how it relates to privacy and self-disclosure.

Sabine Trepte is a professor for media psychology at the University of Hohenheim (Germany). Her research focuses online self-disclosure and privacy from a psychological perspective.

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 4, Pages 170–181
https://doi.org/10.17645/mac.v9i4.4164

Article

Algorithmic or Human Source? Examining Relative Hostile Media Effect With a Transformer-Based Framework

Chenyan Jia [1,*] and Ruibo Liu [2]

[1] Moody College of Communication, The University of Texas at Austin, USA; E-Mail: chenyanjia@utexas.edu
[2] Department of Computer Science, Dartmouth College, USA; E-Mail: ruibo.liu.gr@dartmouth.edu

* Corresponding author

Submitted: 7 February 2021 | Accepted: 7 October 2021 | Published: 18 November 2021

Abstract
The relative hostile media effect suggests that partisans tend to perceive the bias of slanted news differently depending on whether the news is slanted in favor of or against their sides. To explore the effect of an algorithmic vs. human source on hostile media perceptions, this study conducts a 3 (author attribution: human, algorithm, or human-assisted algorithm) × 3 (news attitude: pro-issue, neutral, or anti-issue) mixed factorial design online experiment (*N* = 511). This study uses a transformer-based adversarial network to auto-generate comparable news headlines. The framework was trained with a dataset of 364,986 news stories from 22 mainstream media outlets. The results show that the relative hostile media effect occurs when people read news headlines attributed to all types of authors. News attributed to a sole human source is perceived as more credible than news attributed to two algorithm-related sources. For anti-Trump news headlines, there exists an interaction effect between author attribution and issue partisanship while controlling for people's prior belief in machine heuristics. The difference of hostile media perceptions between the two partisan groups was relatively larger in anti-Trump news headlines compared with pro-Trump news headlines.

Keywords
algorithms; automated journalism; computational method; hostile media effect; source credibility

Issue
This article is part of the issue "Algorithmic Systems in the Digital Society" edited by Sanne Kruikemeier (University of Amsterdam, The Netherlands), Sophie Boerman (University of Amsterdam, The Netherlands), and Nadine Bol (Tilburg University, The Netherlands).

1. Introduction

With advances in machine learning techniques and the growing availability of big data, algorithms have become widely adopted in news agencies around the world (Jia & Johnson, 2021). Automated journalism is defined as a form of news production that can automatically produce news stories with little human intervention beyond the initial programming phase (Carlson, 2015; Graefe, 2016; Tandoc et al., 2020). Although automated journalistic writing is mostly restricted to factual and data-driven topics such as sports, finance, crime, weather, and disaster reporting, it has also been applied to other domains such as political news (Jia & Johnson, 2021; Wu, 2020). With the growing presence of automated journalism, this new technological affordance has altered how audiences consume and engage with news (Liu & Wei, 2019).

Increasing scholarly attention has been given to the perceptions of automated news (e.g., Graefe et al., 2018; Jia & Gwizdka, 2020; Wu, 2020). One recent meta-analysis shows that when reading the actual content written by humans and algorithms, people perceive no difference in terms of news credibility; however, people perceive news purportedly attributed to algorithms as slightly less credible than news attributed to humans (Graefe & Bohlken, 2020). Algorithmic author

attribution may reduce message credibility through an indirect pathway of source anthropomorphism because people prefer human rather than machine sources possibly due to the principle of similarity attraction (e.g., Byrne, 1997; Simons et al., 1970). Anthropomorphism is defined as the attribution of human traits, motivations, emotions, or behaviors to non-human and non-living entities (Airenti, 2015). Another explanation for why algorithmic sources are perceived as less credible than human sources is that people tend to be less familiar and knowledgeable with automation technologies (e.g., Haim & Graefe, 2017).

Despite the initial evidence of differences in source credibility between human and algorithmic sources, few studies have further examined whether machine source attribution can affect partisans' perceptions of news with an ideological slant. It remains unclear whether the manipulation of the source can affect people's perceptions of media bias, especially for partisans who are likely to fall prey to hostile media phenomena. Hostile media effect (HME) refers to the tendency for people who are highly involved in an issue to rate ostensibly neutral and balanced stories as biased due to their own biases (Arpan & Raney, 2003; Feldman, 2011; Giner-Sorolla & Chaiken, 1994; Vallone et al., 1985). The relative HME theory further argues that partisans tend to perceive the extent of bias of slanted news coverage differently depending on whether the news is slanted in favor of or against their points of view (Goldman & Mutz, 2011; Gunther et al., 2001). Previous studies have examined the relative HME by studying how partisans perceive news from sources with different levels of credibility (e.g., Arpan & Raney, 2003; Coe et al., 2008; Gunther & Liebhart, 2006). Gunther and Liebhart (2006), for instance, examine the influence of sources with different credibility (journalist vs. college student) on people's perceptions of bias. Their results suggest that more credible sources (journalist, large reach) yield more HME than lower credible sources (college student, small reach; Gunther & Liebhart, 2006). Few studies, however, have investigated how algorithmic sources will affect the relative HME compared with human sources. Will the attribution of an algorithmic source produce less HME than a human source because of its relatively lower credibility?

Given the increasing usage of AI in the online political context, it is especially important to examine whether algorithmic sources (as opposed to human sources) will increase or reduce relative HME. If an algorithmic cue can reduce the relative HME or perceived bias, it might increase partisans' exposure to cross-cutting information and help combat the extreme polarization (Jia & Johnson, 2021). In order to examine the effect of algorithmic sources on HME, this study conducts a 3 (author attribution: human, algorithm, or human-assisted algorithm) × 3 (news attitude: pro-issue, neutral, or anti-issue) mixed factorial design online experiment (N = 511). Using the computational method, this study adopts a transformer-based adversarial network to generate comparable stimuli. The framework was trained with a dataset of 364,986 news stories from 22 mainstream media outlets (Liu, Jia, & Vosoughi, 2021). The present research answers the following overarching questions: (a) Will stories purportedly written by humans produce different relative HME compared with those by algorithms? (b) How will source credibility (human vs. algorithmic sources) affect relative HME?

2. Literature Review

2.1. Relative Hostile Media Effect

HME refers to the tendency for partisans (i.e., individuals with a strong preexisting political stance) to perceive neutral media coverage as biased against their sides (e.g., Giner-Sorolla & Chaiken, 1994). The original HME theory assumes that news stories are balanced (Feldman, 2011). The relative HME theory expands the assumption of the original HME by making it applicable to news that is slanted rather than balanced (Gunther et al., 2001). More specifically, the relative HMEs theory suggests that partisans tend to perceive the same media content differently and perceive less bias in the news coverage leaning toward their views than their opponents (Feldman, 2011; Gunther & Chia, 2001). The HME theory has been tested in numerous contexts through both experimental and survey methods (Feldman, 2017; Perloff, 2015). A recent meta-analysis of 34 HME studies has shown that a considerable number of empirical studies have provided widespread evidence of HME (Hansen & Kim, 2011).

Researchers have attempted to provide multiple explanations for why HME manifests itself. One explanation is the idea of the selective process (or message-processing mechanisms; Feldman, 2011, 2017; Giner-Sorolla & Chaiken, 1994; Gunther & Liebhart, 2006). Partisans often selectively recall, categorize, or use different standard mechanisms to process unfavorable or attitude-challenging content (Hansen & Kim, 2011). The second core factor that may lead to HME is the level of involvement. Early HME studies often suggest that hostile media phenomenon is limited to partisans who have strong issue involvement (e.g., Giner-Sorolla & Chaiken, 1994; Vallone et al., 1985). Recent studies, however, view issue involvement as a moderator of HME (Perloff, 2015).

Another key factor that might explain the HME is source credibility (Hansen & Kim, 2011). Arpan and Raney (2003) suggest that the credibility ratings of news sources may affect hostile media perceptions (HMP). People's expectations about the media outlet affect the perceived hostility of the media (Giner-Sorolla & Chaiken, 1994). Past work indicates that people's prior beliefs about the source credibility (or a related concept "trust") give rise to biased processing of the content (Baum & Gussin, 2007). Partisans often perceive news sources providing confirmatory information as more credible than those that do not (Baum & Gussin, 2007). Some studies

have examined how source credibility affects HME by examining home-town newspapers vs. rival-town newspapers (Arpan & Raney, 2003), college students vs. journalists (Gunther & Liebhart, 2006), different cable television news programs (CNN, FOX, The Daily Show; Coe et al., 2008). Other studies have also examined relative HME by varying general characteristics of news coverage (circulation size, type of ownership; Gunther, 1992).

2.2. Automated Journalism

Very few studies have examined whether human sources and algorithmic sources will yield different relative HME. Many scholars have conducted empirical research on perceptions of automated news by examining both the actual content of automated news (e.g., Clerwall, 2014; Graefe et al., 2018; Haim & Graefe, 2017; Jia & Gwizdka, 2020; Wu, 2020) and the effects of machine vs. human source attribution (e.g., Jung et al., 2017; Tandoc et al., 2020; Waddell, 2019). In terms of perceptions of credibility, studies examining actual content produced by algorithms as opposed to humans yield different results from studies focusing exclusively on the effect of source attribution (e.g., Graefe & Bohlken, 2020; Jia, 2020; Waddell, 2018). One recent meta-analysis including 12 studies on automated news shows no difference in readers' perceptions of credibility when reading the actual content written by humans and algorithms (Graefe & Bohlken, 2020). In terms of the effect of source, however, findings of the meta-analysis revealed that people perceive news purportedly attributed to algorithms as slightly less credible than news attributed to humans (Graefe & Bohlken, 2020).

To understand the effect of machine sources as opposed to human sources, many recent studies examine the source attribution while controlling for the content (e.g., Jung et al., 2017; Tandoc et al., 2020; Waddell, 2019). Results are mixed on whether news purportedly written by a machine source is more or less credible. Some studies found that news attributed to a machine author is more credible than news attributed to a human author, especially for news that requires more information processing (Liu & Wei, 2019; Waddell, 2019). Others found no main difference in the perceived source credibility between news attributed to algorithmic and human authors (e.g., Tandoc et al., 2020). Overall, most previous studies suggest that news attributed to a human author is perceived as more credible than news attributed to an algorithmic author (Graefe & Bohlken, 2020; Jia & Johnson, 2021).

Very few studies, however, have further investigated whether the difference in source credibility (human vs. algorithm) will have an impact on relative HME. Therefore, this study aims to fill in the gap by examining whether HME also occurs in news stories attributed to algorithms. This work proposes the authorship (humans, algorithms, or human-assisted algorithms) of news stories as a novel source cue and examines how different sources affect issue partisans' HMP. Issue partisans are people who hold strong and even extreme attitudes toward an issue, especially a political issue (Feldman, 2011). Previous work suggests that partisans in favor of one issue often perceive anti-issue news as relatively more biased than partisans on the opposing side, regardless of the news source (e.g., Arpan & Raney, 2003). Pro-issue news refers to the news in favor of one issue whereas anti-issue news refers to the news standing against one issue. Adding to previous literature, this study predicts the following hypotheses:

> H1: For news headlines purportedly written by (a) humans, (b) algorithms, and (c) human-assisted algorithms, partisans on the supporting side of an issue will perceive the anti-issue news as relatively more biased than partisans on the opposing side.

> H2: For news headlines purportedly written by (a) humans, (b) algorithms, and (c) human-assisted algorithms, partisans on the opposing side of an issue will perceive the pro-issue news as relatively more biased than partisans on the supporting side.

2.3. Source Credibility

Source credibility was initially used to measure how the characteristics of speakers influence the receiver's acceptance of a message (Hovland et al., 1953). Factors such as the speaker's expertise, truthfulness, and motivation to tell the truth are major characteristics to determine source credibility (e.g., Hovland et al., 1953; Perloff, 2015). The concept of credibility is related to different theoretical concepts including trust and fairness (Engelke et al., 2019). People are inclined to judge news stories from traditional mainstream media as more credible than those from social media because they often employ journalistic values such as trustworthiness, fairness, professionalism, and balance in assessing credibility (Johnson & Kaye, 2013; Yamamoto et al., 2016). People also tend to perceive attitude-consistent information as more credible and trustworthy than that challenges their beliefs (Metzger & Flanagin, 2015).

A large body of literature suggests the influence of source credibility on perceptions of media bias (e.g., Arpan & Raney, 2003; Gunther & Liebhart, 2006). For instance, Gunther and Liebhart (2006) found that more traditionally credible sources (i.e., journalists over students, large over small reach) facilitate more relative HME. The majority of past studies on automated journalism suggest that news purportedly written by a human author is perceived as more credible than that by an algorithmic author (e.g., Graefe & Bohlken, 2020; Waddell, 2019). Adding to previous literature, this study also predicts that people will perceive human sources as more credible, and thus yield greater HME in news attributed to human sources rather than news attributed to algorithmic sources:

H3: People will perceive human sources as more credible than algorithms sources.

H4: Relative HME will be greater for news headlines purportedly written by humans than news headlines purportedly written by algorithms or human-assisted algorithms.

Readers who process information heuristically often rely on the reputation of heuristic cues and consider familiar sources more credible (Metzger & Flanagin, 2015; Metzger et al., 2010). Several automated journalism studies have shown that audiences have relatively low familiarity and lack of knowledge with automation technologies (e.g., Haim & Graefe, 2017) and thus may perceive algorithm attributed news differently from human attributed news in terms of credibility (e.g., Clerwall, 2014; Jung et al., 2017; Waddell, 2019). Readers may rely more on heuristic cues such as source credibility to make judgments rather than their own issue attitudes because of their low familiarity with the underlying mechanism of algorithms (Haim & Graefe, 2017). One previous study found that source credibility partially mediates the influence of issue partisanship on peoples' selective exposure to gun stories (Jia & Johnson, 2021). Another study also found that trust in source mediates people's perceptions of algorithmic products (Shin, 2020). Adding to past work, this present study predicts a mediating effect of source credibility on people's HMP:

H5: Source credibility will mediate the influence of issue partisanship on people's HMP.

3. Method

3.1. Experimental Design

The present study adopted a 3 (author attribution: human, algorithm, or human-assisted algorithm) × 3 (news attitude: pro-issue, neutral, or anti-issue) mixed factorial design online experiment. An online experiment (N = 511) embedded in Qualtrics was conducted in January 2021. Author attribution was a between-subjects variable whereas news attitude was a within-subjects variable. Participants were randomly assigned to read news headlines purportedly written by a human author (n = 168), algorithm (n = 168), or human-assisted algorithm (n = 175). Each participant was asked to read 15 news headlines about Donald Trump. The order of headlines was randomized.

3.2. Procedures

Before the experiment, participants were asked to report their political attitude, party affiliation, and attitude towards Donald Trump. Participants also need to answer several questions about source familiarity and source credibility for both human authors and algorithmic authors. Then, participants were randomly assigned to read news headlines purportedly written by human, algorithm, or human-assisted algorithm. After participants read each headline, they were asked to rate their perceived bias and credibility of news stories. Given the important role of author attribution in this experimental design, participants were asked if they could recall the author listed on the byline (adapted from Jia & Johnson, 2021; Waddell, 2019). Two attention checks were embedded in the experiment to exclude careless responses. Participants were asked to select point three on the first attention check question. In the second attention check question, participants were asked to add three to the first number they selected and to use the result as the answer to the second question.

3.3. Participants

For both the pre-test and the main experiment, participants were recruited from CloudResearch (formerly known as TurkPrime) which is an advanced online crowdsourcing platform for behavioral science data collection (Litman et al., 2017). Participants were all from the United States and above 18 years old. Each participant was required to have a HIT approval rate greater than 95%. Participants in both pre-test and main experiment were paid 75 cents for their participation. After ruling out repeated IP addresses, incomplete answers, and subjects who failed both attention checks, 511 participants remained in the main experiment. The average age of participants was 41.10 years old (SD = 12.77, $Median$ = 39). More than half of the participants (54%) were male, 45.8% were female, and 0.2% of participants chose other. Participants have received 16.27 years of education on average (SD = 2.30, $Median$ = 16). The majority of participants are White (73.6%), followed by 11.7% Asian/Pacific Islander, 7.8% Black/African American, 3.3% Hispanic/Latino/Latina, 2.7% Other/Multi-Racial, and 0.8% participants preferred not to respond to the race question. About a half of the participants (48.1%) self-identified as Democrats, 24.1% were Independent, and 27.8% self-identified as Republicans.

3.4. Stimuli and Computational Method

In order to guarantee the stimuli were comparable, this study used a computer-assisted method, the current state-of-the-art controllable headline generation model (Liu, Jia, & Vosoughi, 2021), to generate news headlines with different political ideologies (i.e., liberal, neutral, and conservative). Specifically, the model consists of two main modules: the polarity detection module and the polarity flipper module. The polarity detector leverages the self-attention mechanism of the Transformer framework to score the polarity of different spans of the text and outputs the biased part of the text. The polarity flipper only flips the detected biased content through

an adversarial network, with the preservation of semantic consistency (Liu, Jia, & Vosoughi, 2021). The model was trained on a dataset of 364,986 news stories from 22 mainstream media outlets (CNN, NYTimes, Fox, WSJ, etc.). News articles were collected from Allsides and Media Cloud. Each news story from Allsides was labeled with a political polarity label by an editing expert. News articles from Media Cloud were assigned with ideological polarity labels according to the polarity of media outlets using the rationale developed by Pew Research Center (Pew Research, 2014). All news articles were collected from June 2012 to May 2019.

We first selected eight news headlines about the 45th President of the United States, Donald Trump, from our dataset (four from liberal media outlets and four from conservative media outlets). News headlines were selected as stimuli because our model performed better in flipping and neutralizing headlines rather than body text. This can be explained by the fact that issue attitudes are oftentimes more obvious in the headlines of strategic news stories compared with the body text (Liu, Wang, et al., 2021). Using the transformer-based framework, this study automatically generates corresponding headlines either with the opposite polarity or being neutral. Some of the machine-generated headlines may be slightly ungrammatical due to the mechanism of machine learning algorithms (e.g., *Trump Said That Shouldn't Matter Those State of the Union Ratings*). However, we intentionally did not include an additional human editing process because the post-editing procedure of machine-generated texts may introduce more potential bias (Biswas & Rajan, 2020). All stimuli were generated by one unified transformer-based framework and thus have an overall consistent performance on the grammatical level. In order to avoid the potential influence of ungrammatical languages, this study used multiple sets of headlines to test the robustness of the experiment design.

A pre-test (N = 90) was conducted to examine the issue attitudes of these 24 headlines. Participants in the pre-test were asked to read 24 headlines and answer whether the news headline was strictly neutral or biased in favor of one side or the other followed by the 11-point scale with −5 indicating strongly biased against Trump, and +5 indicating strongly biased in favor of Trump, 0 indicating strictly neutral. Several one-sample T-tests (test value = 0) were conducted to test whether the stimuli were biased in the direction they were designed to be. Based on the results of the pre-test, 3 sets of headlines were excluded in the main experiment because they were not statistically significant in the direction they were designed to be. In total, 15 headlines were chosen as final stimuli of the main experiment (5 anti-Trump, 5 neutral, 5 pro-Trump), as shown in Table 1.

This study used Photoshop to make every stimulus looks like a screenshot from the same fictional news site. The bylines of articles were under each headline, following with the published time and three social media sharing buttons. In the byline, it either shows "by an automated journalism algorithm," "by staff reporter Jim Richard," or "by an automated journalism algorithm and staff reporter Jim Richard."

3.5. Measures

3.5.1. Issue Partisanship

Issue partisanship (M = 1.60, SD = 0.59) was measured by asking to what extent they support or oppose Trump. Responses were recorded on the 11-point scale ranging from −5 (*strongly support Trump*), 0 (*strictly neutral*), and +5 (*strongly oppose Trump*; adapted from

Table 1. News headlines stimuli.

Anti-Trump	Neutral	Pro-Trump
Donald Trump Lied About His State of The Union Ratings	Trump Said That Shouldn't Matter Those State of The Union Ratings	Trump Claims Highest His State of The Union Ratings
Trump Denies Asking Ex-FBI Director Comey to Drop Flynn Investigation	Trump Says: "I Never Asked Comey to Stop Investigating Flynn''	Trump Never Asked Ex-FBI Director Comey to Stop Investigating Flynn
Trump Administration Considers Tearing Families Apart in New Immigration Crackdown	Trump Immigration Plan Provides Path to Citizenship for Millions of Immigrants Illegally	Trump Offers Dreamers A Path to Citizenship, Tough on Other Immigrants
Trump Threatens to Abandon Puerto Rico Recovery Effort	Senate Narrowly Approves Budget, Paving Way for Tax Reform	Senate Approves Budget in Crucial Step for Trump's Tax Overhaul
Trump Administration Approves Plan to Separate Families at Border	Trump's New HHS Office Will Protect Health Care Workers Who Violate Abortion	Trump Administration Creates New Religious Protections for Health Care Workers

Notes: The bias ratings of anti-Trump stimuli are significantly lower than value 0 whereas ratings of pro-Trump stimuli are all significantly higher than value 0; the ratings of neutral headlines are not significantly different from value 0.

Feldman, 2011). Participants were then categorized into three groups using the cutoff value 0. Participants who selected 0 (strictly neutral) were classified as non-partisans and were excluded in further analysis. Participants scoring below 0 were classified as pro-Trump partisans (n = 148) whereas those scoring above 0 were classified as anti-trump partisans (n = 336).

3.5.2. Hostile Media Perception

The measure of perceived bias or slant in the news headlines was adapted from previous research on HMP (Giner-Sorolla & Chaiken, 1994; Gunther & Schmitt, 2004; Feldman, 2011). Participants were asked "Would you say that the above news headline was strictly neutral, or was it biased in favor of one side or the other?" followed by the 11-point scale with −5, *strongly biased against Trump*, and +5, *strongly biased in favor of Trump*. Two additional items asked participants to rate what percentage of the news headline was unfavorable and favorable, respectively, toward the focal news issue on 11-point scales ranging from 0 to 100%. Both items were converted to a −5 to +5 response scale. All three items were then averaged to form a scale, where positive scores represent the headline is perceived as biased favorable toward Trump and negative scores a bias unfavorable toward Trump. Three items were highly correlated and can be averaged to form a reliable index (M = 1.16, SD = 2.31, Cronbach's α = 0.78). This study also measured the HMP of authorship because the manipulation of the authorship is a key variable. Participants need to answer, "Would you say that the author of the above news headline was strictly neutral, or was it biased in favor of one side or the other?" on a −5 to +5 response scale (adapted from Gunther & Schmitt, 2004).

3.5.3. Source Credibility and Message Credibility

Credibility was measured as a multidimensional construct consisting of believability, fairness, accuracy, depth of information, and authenticity (adapted from Gaziano & McGrath, 1986; Metzger et al., 2003; Newhagen & Nass, 1989) on 7-point scales. Participants were asked to rate source credibility based on their expectations or previous experiences with human, human-assisted algorithm, and algorithmic authors. After reading each news headline, participants were asked to rate message credibility. This study reverse coded the authenticity item ("the story written by humans or algorithms is not authentic"). Five items were highly correlated and can be averaged into one measure (M = 3.99, SD = 1.32, Cronbach's α = 0.92).

3.5.4. Prior Belief in Machine Heuristics

Participants' prior beliefs in the machine heuristic were measured using four bipolar items "harmful/beneficial," "unethical/ethical," "unfavorable/favorable,'' "unnecessary/necessary." Participants were asked to rate the questions "What is your view on using machine learning software to replace or augment human journalists?" on 7-point scales. One item "unfavorable/favorable" was reverse coded. Four items were highly correlated and can be averaged into one measure (M = 3.15, SD = 1.42, Cronbach's α = 0.90).

3.6. Manipulation Check

To ensure the experimental manipulation was effective, participants answered a manipulation check to rate their perceived source anthropomorphism of the listed author(s) on four 7-point scales. Four semantic differential items "fake/natural," "unconscious/conscious," "artificial/life-like," and "mechanical/organic" were adapted from prior research to measure the perceived source anthropomorphism (Bartneck et al., 2007; Jia & Johnson, 2021). Four items were highly correlated and can be averaged to form a reliable index (Cronbach's α = 0.94). One-way Analysis of Variance (ANOVA) was conducted to compare the difference of the perceived source anthropomorphism among the three groups. The source anthropomorphism (M = 4.44, SD = 1.47) rated by participants (n = 168) who were assigned to the human group was significantly higher than the source anthropomorphism (M = 4.13, SD = 1.34) rated by participants (n = 175) who were assigned to the human-assisted algorithm group and that of (M = 3.02, SD = 1.44) the algorithmic author group (n = 168), $F(2, 508) = 46.79$, $p < 0.001$, which showed the manipulation was successful.

4. Results

A two-way repeated measures ANOVA was conducted to test H1 and H2. H1 predicted that for anti-Trump news headlines purportedly written by all types of authors, the pro-Trump group would perceive more bias than the anti-Trump group. Analysis showed that there was a significant difference between the anti-Trump and the pro-Trump group, $F(1,478) = 59.45$, $p < 0.001$, $\eta^2 = 0.11$. For anti-Trump news attributed to each author type, the pro-Trump group perceived significantly more bias than the anti-Trump group, $p < 0.001$. H1 was supported.

H2 predicted that for the pro-Trump news headlines written by all types of authors, anti-Trump partisans would perceive more bias than the pro-Trump group. Analysis showed that there was a significant difference between two groups, $F(1,478) = 17.95$, $p < 0.001$, $\eta^2 = .04$. The direction was as expected in each condition, as shown in Table 2. Therefore, H1 and H2 were supported.

H3 predicted that people would perceive human sources as more credible than algorithms sources. A one-way ANOVA was conducted to examine whether there existed significant differences in source credibility between algorithm-related authors and human authors. Results showed that human sources (M = 4.19, SD = 1.24) were perceived as significantly more credible

Table 2. Partisan group HMP means in three authorship conditions.

	Algorithm Author		Combined Author		Human Author	
	Anti-Trump group	Pro-Trump group	Anti-Trump group	Pro-Trump group	Anti-Trump group	Pro-Trump group
Anti-Trump News						
M	−1.95(0.13)	−3.02(0.19)	−2.13(0.18)	−3.25(0.21)	−1.98(0.13)	−2.91(0.18)
n	106	53	128	40	102	55
Neutral News						
M	0.51(0.11)	−0.37(0.15)	0.30(0.10)	−0.55(0.17)	0.38(0.11)	−0.10(0.15)
n	106	53	128	40	102	55
Pro-Trump News						
M	1.51(0.12)	0.73(0.17)	1.38(0.11)	0.76(0.20)	1.40(0.12)	1.22(0.17)
n	106	53	128	40	102	55

Note: Means for each group were presented as marginal means (with standard errors in parentheses).

than both the algorithm author (M = 3.75, SD = 1.35) and the combined author (M = 4.07, SD = 1.28), F (2,508) = 5.16, p < 0.01. Specifically, pure algorithmic sources received the lowest credibility score. Therefore, H3 was supported.

H4 predicted that the relative HME would be greater for news headlines purportedly written by humans compared with news headlines purportedly written by algorithms or human-assisted algorithms. Two-way repeated measures ANOVA showed that there was no significant main effect observed for source attribution in both anti-Trump, F (2,478) = 1.18, p = 0.31, η^2 = 0.01, and pro-Trump headlines, F (2,478) = 1.44, p = 0.24, η^2 = 0.01. The interaction between source attribution and issue attitudes was not significant for both anti-Trump, p = 0.85, and pro-Trump news headlines, p = 0.12. Therefore, H4 was not supported. In fact, in both pro- and anti-Trump news headlines, human authors produced smaller relative HME than algorithm-related authors, but the difference is not statistically different, as shown in Figure 1.

Several repeated measures ANCOVAs were conducted as additional analyses. Results show that while controlling for people's prior belief in machine heuristics, a marginally significant interaction effect between source attribution and issue attitudes was detected for anti-Trump news headlines, p = 0.052. There existed no significant interaction effect between source attribution and issue partisanship for pro-Trump news headlines, p = 0.11 and neutral headlines, p = 0.41 after controlling for people's prior belief in machine heuristics.

H5 predicted that source credibility would mediate the influence of issue partisanship on the perceived bias. To test this hypothesis, a mediation model was run by using the PROCESS macro (Hayes, 2013) based on nonparametric bootstrapping with 1,000 simulations and 95% confidence intervals (CIs). For anti-Trump news headlines, the effect of issue partisanship on HMP was partially mediated via source credibility. Ninety-five percent CIs for indirect, direct, and total effects did not include zero, which means all these effects were significant. As Figure 2 illustrates, for anti-Trump news, the indirect effect was a_1*b_1 = 0.016, CI = [0.001, 0.03]. The direct effect was c_1 = 0.13, CI = [0.095, 0.17]. The total effect was 0.15, CI = [0.11, 0.19]. Thus, H5 was supported for anti-Trump news headlines. For pro-Trump news headlines, a mediation effect of source credibility did not occur. The indirect effect was not significant, CI = [−0.03, 0.00]. Therefore, H5 was partially supported.

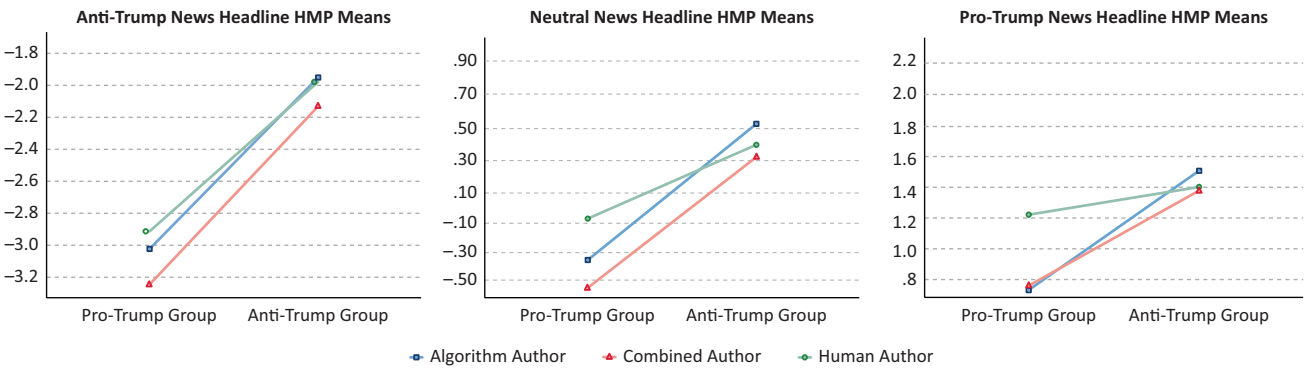


Figure 1. Perceived headline HMP means.

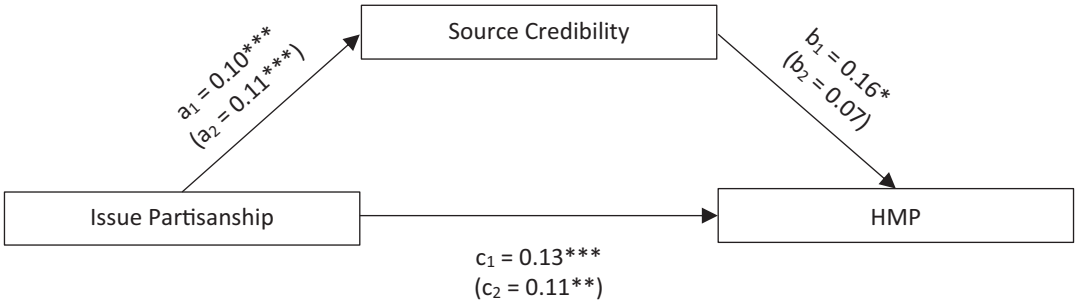


Figure 2. Mediation analysis of anti-Trump (pro-Trump) news headlines. Notes: Two mediation models were presented in the figure; source credibility partially mediated the influence of issue attitudes on HMP for anti-Trump news headlines; *p < 0.05, **p < 0.01 *** p < 0.001.

5. Discussion

5.1. Theoretical Contribution

In recent years, algorithms become ubiquitous in the contemporary media environment (Thurman et al., 2017). Our work adds to the previous literature in both fields of political communication and human-computer interaction. Previous works have studied how different types of traditional news sources such as different cable television channels (Coe et al., 2008), journalists vs. college students (Gunther & Liebhart, 2006) affected the relative HME. Our study introduced a novel type of source (algorithmic sources) and shed light on new trends in the era of an algorithms-driven society.

This study contributed to both automated journalism and HME literature. First, the results of this study showed that relative HME occurred in all types of author attributions (human, algorithm, and combined authors). This work revealed that partisans tend to perceive the bias of slanted news coverage differently depending on their own political ideology even when the authors are algorithms. This result was consistent with previous research that relative HME existed regardless of the news source (e.g., Arpan & Raney, 2003). This work also found that news attributed to a sole human source was perceived as more credible than two algorithm-related sources. News attributed to a sole algorithmic author was rated as the least credible among three types of declared authors. This result was consistent with most previous studies (see Graefe & Bohlken, 2020; Jia & Johnson, 2021) but contradicted Waddell's (2019) findings. Waddell (2019) suggests that messages attributed to human authors are perceived as the least credible than those attributed to machine and combined sources. One possible explanation is that Waddell (2019) not only attributed authorship (human, algorithm, combined) but also attributed media outlets as another source to indicate the ideological slant of the stimuli (i.e., Fox News or MSNBC). Therefore, the credibility of media outlets may affect participants' perceived credibility of messages as well. Also, the difference of perceived message credibility among news attributed to three types of authors is marginal significant in Waddell's (2019) study.

This study predicted that news headlines attributed to human authors would evoke a larger relative HME than news attributed to algorithmic sources because one previous study suggests that more traditionally credible sources yield larger HME than lower credible sources (Gunther & Liebhart, 2006). This hypothesis was not supported. In fact, for both pro- and anti-Trump news headlines, human authors produced smaller relative HME than algorithm-related authors, but the difference was not statistically different. An interesting pattern was that when controlling for people's prior belief in machine heuristics, an interaction effect was detected between source attribution and issue partisanship. It is worth noting that such interaction effect was only significant for anti-Trump news headlines. With that said, for anti-Trump news headlines, news attributed to a sole algorithmic author (low credibility) yields larger relative HME than news attributed to a sole human author (high credibility) when controlling for people's prior belief in machine heuristics. As a matter of fact, even Gunther and Liebhart (2006) acknowledged that it is possible that partisans simply considered the student a more trustworthy source than the journalist. Future studies can further explore whether the pattern of low credible source yield larger relative HME stays true or not in other realms.

Another interesting finding of this study is that the difference of perceived bias between two partisan groups was relatively larger in anti-Trump news headlines compared with pro-Trump news headlines. The pro-Trump group perceived anti-Trump news headlines as much more biased than the anti-Trump group did. This is not surprising because Trump supporters are less likely to interact with outgroups (Pettigrew, 2017), and thus may perceive attitude-challenging information as very biased.

This study further investigated how source credibility affects people's perceived bias and found that source credibility partially mediates the influence of issue partisanship on people's perceived bias for anti-Trump news. This finding was interesting as it suggested that when partisans read news headlines, source credibility plays an important role in hostile media perceptions. This study posits a possible theoretical model to predict the perceived bias through source credibility.

Another contribution of this study is the implementation of a computer-assisted method to generate comparable news headlines. Earlier studies often use news content from different media outlets (e.g., *The New York Times* and *The Wall Street Journal*) to manually manipulate the select news articles or headlines with different political ideologies (e.g., Van Duyn & Collier, 2019). Although the political ideology of news content can be manipulated by pre-testing the stimuli, it is inevitable to include potential bias caused by different writing quality or source credibility (Liu, Jia, & Vosoughi, 2021). Our study, however, proposes a novel approach that is capable of flipping the ideology without shifting semantic meaning and readability of news articles.

5.2. Limitations

Despite these contributions, this study still has certain limitations. First, this study chose Trump-related news as a partisan issue because Trump was one of the most salient topics in our dataset. Our transformer-based framework was trained using the media coverage data from June 2012 to May 2019. Therefore, the stimuli were somewhat outdated by the time when the experiment was conducted. Future studies can choose more conventional political topics such as gun control, abortion, immigration, and gay marriage (Knobloch-Westerwick et al., 2017; Wojcieszak, 2019). Furthermore, some of the stimuli were not as readable as real news headlines due to the auto-generation process. The topic of the stimuli was not always controlled in each set because we prioritized the fact whether the ideological direction of stimuli was as we expected from the pre-test.

Second, the majority of our sample self-reported as Democrats, which also leads to the imbalanced number of two-issue partisan groups (pro-Trump and anti-Trump). Even though CloudResearch overcomes many limitations of Amazon Mechanical Turk, it still cannot represent the overall population as participants self-select studies to participate (Litman et al., 2017). Future research can recruit more representative samples especially in terms of pollical ideology.

Third, both issue partisanship and source credibility were not manipulated experimentally, which limits the plausibility of mediation models. If the mediator is measured rather than manipulated, one cannot exclude the possibility that a confounding variable may influence the relationship (Spencer et al., 2005). As one recent review of mediation analysis suggests that elaborate statistical techniques for testing mediation cannot overcome the flaws of inadequate research design (Chan et al., 2020).

5.3. Conclusion

As technology diffuses, the importance of examining how algorithmic source attribution will reduce or increase relative HME is of importance because such study bears implications to media effects studies as well as the media industry. AI research in the political communication area is still at a nascent stage. Some scholars contend that people's perceptions of news bias may be attenuated when news is attributed to a machine author (e.g., Waddell, 2019; Wang, 2021) because AI is often perceived as fair, objective, unbiased, and with less political agenda (Gillespie, 2014). This study, along with many others, found that such positive perceptions of machine neutrality may not always be true. Results of this study showed that the relative HME still occurs when people read news headlines attributed to algorithmic authors. In fact, news headlines attributed to algorithmic authors exhibited larger relative HME than those attributed to human authors in terms of anti-Trump news while controlling for people's prior belief in machine heuristics. The current study sheds light on a better understanding of the role of machine cues in the political context.

Conflict of Interests

The authors declare no conflict of interests.

References

Airenti, G. (2015). The cognitive bases of anthropomorphism: From relatedness to empathy. *International Journal of Social Robotics*, *7*(1), 117–127. https://doi.org/10.1007/s12369-014-0263-x

Arpan, L. M., & Raney, A. A. (2003). An experimental investigation of news source and the hostile media effect. *Journalism & Mass Communication Quarterly*, *80*(2), 265–281. https://doi.org/10.1177/107769900308000203

Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. (2007). The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI and Society*, *21*, 217–230. https://doi.org/10.1007/s00146-006-0052-7

Baum, M. A., & Gussin, P. (2007). In the eye of the beholder: How information shortcuts shape individual perceptions of bias in the media. *Quarterly Journal of Political Science*, *3*, 1–31. http://doi.org/10.1561/100.00007010

Biswas, S., & Rajan, H. (2020). Do the machine learning models on a crowd sourced platform exhibit bias? An empirical study on model fairness. In P. Devanbu (Ed.), *Proceedings of the 28th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering* (pp. 642–653). ACM. https://doi.org/10.1145/3368089.3409704

Byrne, D. (1997). An overview (and underview) of research and theory within the attraction paradigm. *Journal of Social and Personal Relationships*, *14*, 417–431. https://doi.org/10.1177/0265407597143008

Carlson, M. (2015). The robotic reporter. *Digital Jour-*

nalism, *3*(3), 416–431. https://doi.org/10.1080/21670811.2014.976412

Chan, M., Hu, P., & K. F. Mak, M. (2020). Mediation analysis and warranted inferences in media and communication research: Examining research design in communication journals from 1996 to 2017. *Journalism & Mass Communication Quarterly*. Advance online publication. https://doi.org/10.1177/1077699020961519

Clerwall, C. (2014). Enter the robot journalist. *Journalism Practice*, *5*, 519–531. https://doi.org/10.1080/17512786.2014.883116

Coe, K., Tewksbury, D., Bond, B. J., Drogos, K. L., Porter, R. W., Yahn, A., & Zhang, Y. (2008). Hostile news: Partisan use and perceptions of cable news programming. *Journal of Communication*, *58*(2), 201–219. https://doi.org/10.1111/j.1460-2466.2008.00381.x

Engelke, K. M., Hase, V., & Wintterlin, F. (2019). On measuring trust and distrust in journalism: Reflection of the status quo and suggestions for the road ahead. *Journal of Trust Research*, *9*(1), 66–86. https://doi.org/10.1080/21515581.2019.1588741

Feldman, L. (2011). Partisan differences in opinionated news perceptions: A test of the hostile media effect. *Political Behavior*, *33*(3), 407–432.

Feldman, L. (2017). The hostile media effect. In K. Kenski & K. H. Jamieson (Eds.), *The Oxford handbook of political communication* (pp. 557–568). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199793471.013.011_update_001

Gaziano, C., & McGrath, K. (1986). Measuring the concept of credibility. *Journalism Quarterly*, *63*(3), 451–462. https://doi.org/10.1177/107769908606300301

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media technologies* (pp. 167–194). MIT Press. https://10.7551/mitpress/9780262525374.003.0009

Giner-Sorolla, R., & Chaiken, S. (1994). The causes of hostile media judgments. *Journal of Experimental Social Psychology*, *30*(2), 165–180. https://doi.org/10.1006/jesp.1994.1008

Goldman, S. K., & Mutz, D. C. (2011). The friendly media phenomenon: A cross-national analysis of cross-cutting exposure. *Political Communication*, *28*(1), 42–66. https://doi.org/10.1080/10584609.2010.544280

Graefe, A. (2016). *Guide to automated journalism*. Tow Center for Digital Journalism, Columbia University. https://doi.org/10.7916/D80G3XDJ

Graefe, A., & Bohlken, N. (2020). Automated journalism: A meta-analysis of readers' perceptions of human-written in comparison to automated news. *Media and Communication*, *8*(3), 50–59. https://doi.org/10.17645/mac.v8i3.3019

Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. (2018). Readers' perception of computer-generated news: Credibility, expertise, and readability. Journalism, *19*(5), 595–610. https://doi.org/10.1177/1464884916641269

Gunther, A. (1992). Biased press or biased public? Attitudes toward media coverage of social groups. *The Public Opinion Quarterly*, *56*(2), 147–167. www.jstor.org/stable/2749167

Gunther, A. C., & Chia, S. C. Y. (2001). Predicting pluralistic ignorance: The hostile media perception and its consequences. *Journalism & Mass Communication Quarterly*, *78*(4), 688–701. https://doi.org/10.1177/107769900107800405

Gunther, A. C., Christen, C. T., Liebhart, J. L., & Chih-Yun Chia, S. (2001). Congenial public, contrary press, and biased estimates of the climate of opinion. *Public Opinion Quarterly*, *65*(3), 295–320. https://doi.org/10.1086/322846

Gunther, A. C., & Liebhart, J. L. (2006). Broad reach or biased source? Decomposing the hostile media effect. *Journal of Communication*, *56*(3), 449–466. https://doi.org/10.1111/j.1460-2466.2006.00295.x

Gunther, A. C., & Schmitt, K. (2004). Mapping boundaries of the hostile media effect. *Journal of Communication*, *54*(1), 55–70. https://doi.org/10.1111/j.1460-2466.2004.tb02613.x

Haim, M., & Graefe, A. (2017). Automated news: Better than expected? *Digital Journalism*, *5*, 1044–1059. https://doi.org/10.1080/21670811.2017.1345643

Hansen, G. J., & Kim, H. (2011). Is the media biased against me? A meta-analysis of the hostile media effect research. *Communication Research Reports*, *28*(2), 169–179. https://doi.org/10.1080/08824096.2011.565280

Hayes, A. F. (2013). *Methodology in the social sciences. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford.

Hovland, C. I., Janis, I. L., & Kelley, H. (1953). *Communication and persuasion*. Yale University Press.

Jia, C. (2020). Chinese automated journalism: A comparison between expectations and perceived quality. *International Journal of Communication*, *14*(2020), 1–21.

Jia, C., & Gwizdka, J. (2020). An eye-tracking study of differences in reading between automated and human-written news. In F. D. Davis, R. Riedl, J. v. Brocke, P.-M. Léger, A. B. Randolph, & T. Fischer (Eds.), *Information systems and neuroscience* (pp. 100–110). NeuroIS Retreat. https://doi.org/10.1007/978-3-030-60073-0_12

Jia, C., & Johnson, T. (2021). Source credibility matters: Does automated journalism inspire selective exposure? *International Journal of Communication*, *15*(2021), 3760–3781.

Johnson, T. J., & Kaye, B. K. (2013). The dark side of the boon? Credibility, selective exposure, and the proliferation of online sources of political information. *Computers in Human Behavior*, *29*(4), 1862–1871. https://doi.org/10.1016/j.chb.2013.02.011

Jung, J., Song, H., Kim, Y., Im, H., & Oh, S. (2017). Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists. *Computers in Human Behavior*, *71*, 291–298. https://doi.org/10.1016/j.chb.2017.02.022

Knobloch-Westerwick, S., Mothes, C., & Polavin, N. (2017). Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Communication Research*, *47*(1). https://doi.org/10.1177/0093650217719596

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Liu, B., & Wei, L. (2019). Machine authorship in situ: Effect of news organization and news genre on news credibility. *Digital Journalism*, *7*(5), 635–657. https://doi.org/10.1080/21670811.2018.1510740

Liu, R., Jia, C., & Vosoughi, S. (2021). A transformer-based framework for flipping political polarity of news articles. *Proceedings of the ACM on Human–Computer Interaction*, *5*(CSCW1), 1–26. https://doi.org/10.1145/3449139

Liu, R., Wang, L., Jia, C., & Vosoughi, S. (2021). Political depolarization of news articles using attribute–aware word embeddings. *Proceedings of the 15th International AAAI Conference on Web and Social Media*, 15(1), 385–396. https://ojs.aaai.org/index.php/ICWSM/article/view/18069

Metzger, M. J., & Flanagin, A. J. (2015). Psychological approaches to credibility assessment online. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 445–466). John Wiley & Sons. https://doi.org/10.1002/9781118426456.ch20

Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & McCann, R. (2003). Credibility in the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. In P. Kalbfeisch (Ed.), *Communication yearbook* (Vol. 27, pp. 293–335). Erlbaum.

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, *60*, 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Newhagen, J., & Nass, C. (1989). Differential criteria for evaluating credibility of newspapers and TV news. *Journalism Quarterly*, *66*(2), 277–284. https://doi.org/10.1177/107769908906600202

Perloff, R. M. (2015). A three-decade retrospective on the hostile media effect. *Mass Communication and Society*, *18*(6), 701–729. https://doi.org/10.1080/15205436.2015.1051234

Pettigrew, T. F. (2017). Social psychological perspectives on Trump supporters. *Journal of Social and Political Psychology*, *5*(1), 107–116. https://doi.org/10.5964/jspp.v5i1.750

Pew Research. (2014). *Political polarization in the American public*. https://www.pewresearch.org/politics/2014/06/12/appendix-a-the-ideological-consistency-scale

Shin, D. (2020). User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. *Journal of Broadcasting & Electronic Media*, *64*(4), 541–565. https://doi.org/10.1080/08838151.2020.1843357

Simons, H. W., Berkowitz, N. N., & Moyer, J. R. (1970). Similarity, credibility, and attitude change: A review and a theory. *Psychological Bulletin*, *73*, 1–16. https://doi.org/10.1037/h0028429

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*, 845–851. https://doi.org/10.1037/0022---3514.89.6.845

Tandoc, E. C., Jr., Yao, L. J., & Wu, S. (2020). Man vs. machine? The impact of algorithm authorship on news credibility. *Digital Journalism*, *8*(4), 548–562. https://doi.org/10.1080/21670811.2020.1762102

Thurman, N., Dörr, K., & Kunert, J. (2017). When reporters get hands-on with robo-editing: Professionals consider automated journalism's capabilities and consequences. *Digital Journalism*, *5*(10), 1240–1259. https://doi.org/10.1080/21670811.2017.1289819

Vallone, R. P., Ross, L., & Lepper, M. R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology*, *49*(3), 577–585. https://doi.org/10.1037/0022-3514.49.3.577

Van Duyn, E., & Collier, J. (2019). Priming and fake news: The effects of elite discourse on evaluations of news media. *Mass Communication and Society*, *22*(1), 29–48. https://doi.org/10.1080/15205436.2018.1511807

Waddell, T. F. (2018). A robot wrote this? How perceived machine authorship affects news credibility. *Digital Journalism*, *6*(2), 236–255. https://doi.org/10.1080/21670811.2017.1384319

Waddell, T. F. (2019). Can an algorithm reduce the perceived bias of news? Testing the effect of machine attribution on news readers' evaluations of bias, anthropomorphism, and credibility. *Journalism & Mass Communication Quarterly*, *96*(1), 82–100. https://doi.org/10.1177/1077699018815891

Wang, S. (2021). Moderating uncivil user comments by humans or mchines? The effects of moderation agent on perceptions of bias and credibility in news content. *Digital Journalism*, *9*(1), 64–83. https://doi.org/10.1080/21670811.2020.1851279

Wojcieszak, M. (2019). What predicts selective exposure online: Testing political attitudes, credibility, and

social identity. *Communication Research*, *48*(5), 687–716. https://doi.org/10.1177/0093650219844868

Wu, Y. (2020). Is automated journalistic writing less biased? An experimental test of auto-written and human-written news stories. *Journalism Practice*, *14*(8), 1008–1028. https://doi.org/10.1080/17512786.2019.1682940

Yamamoto, M., Lee, T. T., & Ran, W. (2016). Media trust in a community context: A multilevel analysis of individual- and prefecture-level sources of media trust in Japan. *Communication Research*, *43*(1), 131–154. https://doi.org/10.1177/0093650214565894

About the Authors

Chenyan Jia is a PhD candidate and Harrington dissertation fellow in the Moody College of Communication at The University of Texas at Austin. Her research interests include algorithmic bias, human-computer interaction, and computational social science. She is especially interested in how new media technologies such as automated journalism and misinformation detection algorithms are impacting media organizations and the public.

Ruibo Liu is a PhD student in the Department of Computer Science at Dartmouth College. His research interests include natural language processing, human-centered AI, and social computing. He studies political polarity detection, generative models for mitigating bias, and better evaluation metrics that reflect human preference.

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 4, Pages 182–197
https://doi.org/10.17645/mac.v9i4.4167

Article

Epistemic Overconfidence in Algorithmic News Selection

Mariken van der Velden * and Felicia Loecherbach

Department of Communication Science, Vrije Universiteit Amsterdam, The Netherlands;
E-Mails: m.a.c.g.vander.velden@vu.nl (M.v.d.V.), f.loecherbach@vu.nl (F.L.)

* Corresponding author

Submitted: 9 February 2021 | Accepted: 30 August 2021 | Published: 18 November 2021

Abstract
The process of news consumption has undergone great changes over the past decade: Information is now available in an ever-increasing amount from a plethora of sources. Recent work suggests that most people would favor algorithmic solutions over human editors. This stands in contrast to public and scholarly debate about the pitfalls of algorithmic news selection—i.e., the so-called "filter bubbles." This study therefore investigates reasons and motivations which might lead people to prefer algorithmic gatekeepers over human ones. We expect that people have more algorithmic appreciation when consuming news to pass time, entertain oneself, or out of escapism than when using news to keep up-to-date with politics (H1). Secondly, we hypothesize the extent to which people are confident in their own cognitive abilities to moderate that relationship: When people are overconfident in their own capabilities to estimate the relevance of information, they are more likely to have higher levels of algorithmic appreciation, due to the third person effect (H2). For testing those two pre-registered hypotheses, we conducted an online survey with a sample of 268 US participants and replicated our study using a sample of 384 Dutch participants. The results show that the first hypothesis cannot be supported by our data. However, a positive interaction between overconfidence and algorithmic appreciation for the gratification of surveillance (i.e., gaining information about the world, society, and politics) was found in both samples. Thereby, our study contributes to our understanding of the underlying reasons people have for choosing different forms of gatekeeping when selecting news.

Keywords
algorithmic appreciation; algorithmic gatekeepers; algorithmic news selection; third person effect; uses and gratifications

Issue
This article is part of the issue "Algorithmic Systems in the Digital Society" edited by Sanne Kruikemeier (University of Amsterdam, The Netherlands), Sophie Boerman (University of Amsterdam, The Netherlands), and Nadine Bol (Tilburg University, The Netherlands).

1. Introduction

Nowadays, information is not only 24/7 available, but also in unprecedented amounts. There are more news outlets and stories than any human could use, from more and more sources. To separate the "signal" from "noise," news is selected by humans—i.e., journalists (e.g., see the seminal work by Shoemaker & Vos, 2009), friends, but also people who leave similar digital traces (for example, see Gil de Zúñiga et al., 2017)—or automated systems (i.e., algorithms, recommender systems; for an overview, see Nechushtai & Lewis, 2019). While scholars and pundits discuss the dangers and pitfalls of being drawn into the rabbit hole by algorithmic news selection (e.g., Helberger, 2020), the work by Thurman et al. (2019) and the framework of algorithmic appreciation (Logg et al., 2019) suggest that most people prefer news in general to be selected by algorithmic solutions rather than human editors.

We put forward that these seemingly opposing views emerge because the uses and gratifications of news (Diddi & LaRose, 2006; Katz et al., 1974; Lee, 2013; Ng & Zhao, 2020) are lumped together. The uses and gratifications approach proclaims that "instead of studying

what media do to people, we should be asking what people do with media, particularly the gratifications they aim to derive from the fare on offer" (Blumler, 2019, p. 1) and focuses on the reasons why people tend to news—for example to get political information (surveillance) or to be entertained. Using algorithmic news selection to keep up with the latest celebrity gossip or to follow your favorite sports league is less consequential in terms of problematic societal consequences (i.e., polarization, fragmentation) than using algorithmic news selection for getting information about political processes and opinions. However, especially the public discourse is still mainly focused on the negative consequences of news algorithms for fragmentation, polarization, and spread of false information (Harari, 2020; Rose, 2020; Schipper, 2020; Wong, 2020). By now, it has become clear that most of those claims are overstated (Dubois & Blank, 2018; Geiß et al., 2021)—still, the gut feeling that algorithms are "to blame for it" remains rather persistent in academic and public discourse.

We propose to instead look more closely at people's preferences for news selection mechanisms and argue that (a) gratifications for news usage coincide with preferences for news selectors, and (b) overconfidence in one's own cognitive ability moderates news selection preferences. We assume that when the main aim of news consumption is surveillance—i.e., gathering knowledge, keeping up with political news—the quality, accuracy, and diversity of information is of rather high importance for the user. Therefore, we expect the appreciation for expert (human) judgment to be higher when news is consumed for this gratification compared to the other gratifications of news usage (i.e., passing time, entertainment, and escapism). At the same time, algorithmic solutions of news selection mostly depend on user choices (i.e., past selections, friend circles). We thus propose that in cases of overconfidence in one's own ability to judge the relevance of information, algorithmic gatekeepers (i.e., algorithms that choose which news will be presented where and to whom) offer an easy way to get tailor-made, relevant content, even, or especially, when the aim of news consumption is surveillance. The algorithm amplifies one's own decisions—meaning that for overconfident people it is seen as making "good" choices. The negative public picture of news algorithms in this context can rather be seen as a consequence of a third person effect (Davison, 1983)—believing that others are more affected by negative consequences of algorithms than oneself.

We pre-registered our argument (see Supplementary Material), subsequently conducted an online survey with a sample of 250 US participants, and replicated our study using a sample of 400 Dutch participants (for research compendium, see Supplementary Material). Our expectation that people who consume news for surveillance have lower trust in algorithmic news selection is not supported by the data. However, when additionally including overconfidence in one's own abilities in the model, a significant positive interaction effect between the surveillance gratification (i.e., understanding the world, society, and politics) and overconfidence could be found. People with high levels of overconfidence appreciate algorithmic news selection more, the more they use news for staying up-to-date with political news, the opposite holds for low levels of overconfidence.

Our findings contribute to the understanding of the underlying reasons people have for preferring different forms of gatekeeping when selecting news. Despite their bad image, algorithms mostly help consumers to make decisions easier and faster. They thus play a larger role when looking for news with no particular purpose. Given the amount of information floating around these days, algorithmic solutions to news selection seem here to stay. Better understanding people's motives and characteristics for news consumption, and in particular news selectors, gives insights into what people are searching for when selecting gatekeepers. Our pre-registered and exploratory analyses indicate that there is much to be learned from looking at individual differences when it comes to news consumption, but especially the preference for news selectors. This in turn plays a large role in shaping the information environments of citizens and the relative importance of algorithmic intermediaries—a process in which our study can deliver an important first step.

2. Causes for Concern, Yet Appreciated: The Paradox of Algorithmic News Selection

While news consumption has always been influenced by processes of gatekeeping, the enormous amount of information requires to not only select *what* makes it into the news (since space limitations are far less of a concern), but it also brings up the question of *how* to make a selection. In this article, we focus on the latter. While it is still (mostly) journalists and editors that decide whether something happening in the world becomes a news article and gets written up—which certainly is a crucial part of the gatekeeping process—the decision of what actually appears on the front page of the online newspaper and gets suggested to users can be done by either editors or via algorithms. Welbers et al. (2018) showed that even when news selection is performed by human editorial teams, they already heavily depend on sources such as news agencies to handle the workload. Still, in this case, news selection resembles rather closely the traditional way seen in printed newspapers: Experts (trained journalists) judging the newsworthiness of articles for the population as a whole. An increasing amount of pre-selection and placement on the page as well as tailor-made personalization, however, is currently done through algorithmic selections—"gate-keeping no longer belongs to journalists or humans exclusively" (Nechushtai & Lewis, 2019, p. 7). Recommender systems play important roles in the process of algorithmic selection (Karimi et al., 2018; Ricci et al., 2011). Like human editors, they filter information and thereby reduce the information overload placed on news consumers today.

The pathways through which those algorithms select information have been the object of scholarly and public debates. Algorithmic news selectors allegedly cause "echo chambers" (Sunstein, 2009)—where only information that resembles one's own world view is displayed—and/or "filter bubbles" (Pariser, 2011)—where the decisions of the users (often shaped by processes of selective exposure) involuntarily lead to a narrowing of certain dimensions in the selected articles and especially perspectives expressed in the content. Many studies have challenged those assumptions and often found counteracting results (see e.g., Dubois & Blank, 2018; Geiß et al., 2021), still the narratives of echo chambers and filter bubbles remain persistent and continue to draw attention on a wide scale. At the same time, a recent study by Thurman et al. (2019) suggests that people view algorithms as more neutral compared to editors and recommendations from friends and therefore prefer them for their news selection. Their finding can be seen as in line with more general work on algorithm appreciation (Logg et al., 2019), showing that people often prefer algorithmic to human judgment. Sundar (2008) also refers to this phenomenon as the "machine heuristic," describing the assumption of neutrality and absence of bias in technology and algorithms. Carlson (2019) coins a similar phenomenon with "mechanical objectivity," i.e., people assume that algorithms might be better in making unbiased selections as opposed to humans. Thus, while academic and public debates regarding algorithmically curated news often remain negative, the positive perception of people regarding automated news selection stands in contrast to it. Our article, therefore, sets out to look into better understanding the motivations and reasons people might have for preferring algorithms for news selection while public and academic debate rather should lead to a negative image. We draw on two different explanatory factors for diving deeper into the reasons why people might select one gatekeeper over the other: Uses and Gratifications Theory (UGT) and epistemic overconfidence.

2.1. Uses and Gratifications

UGT describes different motives people have for consuming media content (Katz et al., 1974), and has been ample times applied to understand news consumption (see e.g., Diddi & LaRose, 2006; Lee, 2013; Ng & Zhao, 2020). The main question this theory tries to answer is: "Why do people become involved in one particular type of mediated communication or another, and what gratifications do they receive from it?" (Ruggiero, 2000, p. 29). Among the main gratifications people use for news consumption, are entertainment, passing time, escapism, and surveillance. Additionally, answering criticisms over the past few decades, it is also important to include the dimension of habitual usage—accounting for the fact that media selection is not necessarily always based on certain (conscious) goals but often happens as part of a routine (checking the news in the morning) without active intentions (Ruggiero, 2000).

One important aspect of UGT is that it cannot only be used to explain "which media to consume, but also how to consume the media content" (Choi, 2016, p. 250). We draw upon this connection between gratifications sought and the mode or pathway of accessing the news. We expect that the reason why people consume news drives their preferences for gatekeepers of the news. Research on UGT so far mostly focused on looking at the type of media (television, newspaper, online) or the specific outlet that is chosen for news consumption (Lee, 2013). Further, some studies show that, in addition to choosing the content or outlet, the mode of getting access to the content can also be influenced by the gratifications sought. It has been studied which gratifications play a role in using social networking sites in general as a pathway for getting the news (Choi, 2016), which motivations influence the usage of news aggregators (Lee & Chyi, 2015) or using mobile phones as news devices (Li, 2013). However, what we are proposing is to further abstract from different outlets or platforms towards choosing a general mode of gatekeeping for accessing news. This relates less to whether one chooses a newspaper or Twitter but rather to whether the selection of news was made by a human or an algorithm. While we do know that people value algorithms as a selector of the news (Thurman et al., 2019), there are only limited studies that we know of that connect gatekeeping decisions (algorithms vs. humans) with UGT.

We expect that the positive sides of algorithmic news selection—e.g., the ease with which you receive information and reduce information overload (Bozdag, 2013)—exceed the negative sides that are mainly highlighted by scholars—e.g., filter bubbles, polarization, etc.—for specific gratifications more than for others. The negative consequences highlighted by scholars are mainly about the UGT's surveillance domain related to political news—learning more about events happening and keeping up with current events. Gathering political knowledge, seeking information, and shaping attitudes towards topics of societal relevance are the main focus of those concerns. However, a large part of news consumption is not centered around those areas: In the Reuters News Report, around a quarter of the US population indicated to be interested in so-called "soft news" (entertainment, lifestyle, sports) to the same amount or even more than in hard news (politics, economics), especially in the younger age groups (Newman et al., 2015). Studies in the US also show that entertainment motivations play a strong role, especially in the online environment (i.e., being the strongest predictor of the usage of news aggregators, political blogs, or usage of social media for news; Lee, 2013). For escapism, passing time, and entertainment the ease at which articles are obtained might be more important while the stakes for the quality of information are lower than for news sought for surveillance reasons. Especially since recommender systems

are omnipresent and widely used in the entertainment industry for selecting relevant movies, books, and music (i.e., Amazon, Netflix, Spotify), the usage of algorithmic curation in those areas is more widely accepted. As Wölker and Powell (2020) showed, automated sports news content produced by algorithms is perceived as more credible and selected more often compared to human-produced content. We expect that people have more algorithmic appreciation for news when the gratifications sought are escapism, passing time, and entertainment, compared to the surveillance domain since gathering precise information is less important here than fast information selection. Leading to the following expectation:

> H1: People will show less algorithmic appreciation for news selection when the gratification sought for news is surveillance compared to the other gratifications (escapism, passing time, entertainment).

2.2. Epistemic Overconfidence

In addition, we argue that the relationship between gratifications sought for news is conditional upon the confidence in one's own cognitive ability. Typical for algorithmic news selection is that the main information sources to judge the relevance of a news article for the user are (a) the past actions of the user, or (b) the choice of similar people or friends. When expecting less of a negative influence on oneself, the notion of optimistic bias (being less vulnerable to malicious intents) has often been proposed as playing an important role (Salmon et al., 2019; Wei et al., 2007). This has been coined the "third person effect," which states that people tend to believe that media "have a greater effect on others than on themselves" (Davison, 1983, p. 3).

In psychological terms, the third person effect stems from a need to show that one is less gullible to negative effects to bolster a positive self-concept (Kim, 2018). One's self-conception as being superior to others regarding the gullibility of negative effects can be seen as one main driver of the third-person effect. This directly relates to the notion that people overestimate their own capacities and ability to judge the relevance of information (Nisbett & Ross, 1980). The phenomenon has mostly been explored by economists and psychologists as the Dunning-Kruger effect (Dunning, 2011). In this study, we opt for another term, epistemic overconfidence, since it relates to not being able to estimate one's cognitive capabilities correctly. Therefore, we use people's ability for cognitive reflection ("epistemic overconfidence") as a moderator to explore individual overconfidence differences ("third person effect") for the effect of UGT on algorithmic appreciation.

> H2: The more epistemic overconfidence people display, the more algorithmic appreciation they show when the news gratification is surveillance.

3. Data, Measurement, and Methods

To examine how people's expected gratifications of news are connected to the selection of different news gatekeepers, we have fielded an original survey with Amazon Mechanical Turk (MTurk) for the US participants. We have collected data from 268 participants. From the work of Coppock (2019), we know that MTurk samples can be skewed in terms of people's party identities. We had no a priori expectations about people's party identities and their preference for algorithms. Yet, we had 19 respondents (7%) not answering the question. Because this question is asked last in the survey, we will run the statistical tests with and without these respondents, to see if that influences the robustness of the results. In the remainder of our sample, 116 respondents (43%) identified as Democrats, 70 respondents (26%) identified as Republican, 55 respondents (21%) identified as Independent, and 8 respondents (3%) identified as "something else." We grouped the latter two options into the category "Other," comprising of 58 respondents (22%), see Figure 4 for the distribution. In addition, we have fielded the same survey with Pollfish to collect 348 Dutch respondents (Pollfish works similar to MTurk, but has enough Dutch participants). The deviation in number from the pre-registered report (see Supplementary Material) is because of 19 people dropping out of the party ID question. Our research compendium with open materials and more information is available in the Supplementary Material. Our data is fairly balanced regarding gender: 275 respondents (42%) identified as female, 377 respondents (58%) identified as male. The distribution, split up for the two countries, is visualized in the left panel of Figure 4. Regarding age, our sample has a mean age of 35, with a standard deviation of 12.59. This indicates that 95% of our sample is between 18 and 60 years old, with the oldest participant being 74. The distribution of age in our sample is visualized in the middle panel of Figure 4.

In our study, we aim to explain people's appreciation for algorithms when selecting news (exact wording can be found in Appendix A in the Supplementary Material). To do so, we asked people to rank order several gatekeepers: (a) traditional editorial teams, (b) algorithmic selection based on your past reading behavior, (c) algorithmic selection based on the behavior or preferences of your friends or people who are similar to you, and (d) being the gatekeeper themselves. This operationalization made a specific distinction between two forms of algorithms (one based on past-read content, one based on similar users), staying in line with different types of news recommendation algorithms currently being used (i.e., content-based and collaborative). Since most people do only have a vague sense of what algorithms are and based on what information they operate this explanation was added. In the future, it might be good to strike a better balance between the categories by adding similar qualifiers for the traditional

editorial teams to not make it seem as if they pick out news at random compared to algorithms. We asked them to rank them for the four substantive dimensions of news gratifications: (a) to keep up to date with political news (Surveillance); (b) to escape from daily worries (Escapism); (c) to kill time (Pass Time); and (e) to entertain myself (Entertainment). As described in our pre-registered report (see Supplementary Material), we re-coded the ranking variable with 1 if the algorithms options were ranked first, and 0 otherwise. The used scale in the analysis ranges from 0 to 4, where 4 indicates one ranked the algorithms options first for all four news gratifications, 3 indicates one ranked the algorithms options first for three out of four news gratifications, 2 indicates one ranked the algorithms options first for two out of four news gratifications, 1 indicates one ranked the algorithms options first for one out of four news gratifications, 0 indicates one ranked the algorithms options first for none of the four news gratifications. The mean value of this scale is 1.92 (SD 1.43), with 75% of the observations between 1 and 3. This means that on average, people placed an option with algorithmic selection in the first place for two of the gratifications, with some even for three dimensions. Figure 1 shows the distribution. On the x-axis, algorithmic appreciation is approved and on the y-axis the percentage of respondents in each category is displayed.

To explain algorithmic appreciation for news selection, we asked people about their news gratification. We measured this using the scale of Diddi and LaRose (2006). This scale consists of five dimensions: Habit strength, Surveillance, Escapism, Pass Time, and Entertainment. For each of the 23 items, the respondents were asked on a 7-point Likert scale to what extent people thought a statement on news consumption was applicable to them, ranging from 1 (*not at all applicable*) to 7 (*very much applicable*). In the next step, we used a principal components factor analysis using varimax rotation, similar to Diddi and LaRose (2006). Each dimension itself has a high level of reliability: (a) Entertainment consists of 2 items with a Cronbach's α of 0.77; (b) Escapism consists of 5 items with a Cronbach's α of 0.85; (c) Habit Strength consists of 4 items with a Cronbach's α of 0.84; (d) Pass Time consists of 5 items with a Cronbach's α of 0.87; (e) Surveillance consists of 7 items with a Cronbach's α of 0.85.

Figure 2 shows the distribution for each dimension for news gratification. On the x-axis the 7-point scale for each news gratification is displayed, and the y-axis shows the percentage of respondents in each category. All gratifications have means showing that the majority of respondents thought they apply to them, with surveillance being the most sought-after gratification (Entertainment: M = 3.91, SD = 1.73; Escapism: M = 3.59, SD = 1.65; Habit Strength: M = 4.31, SD = 1.60; Pass Time: M = 3.83, SD = 1.65; Surveillance: M = 4.78, SD = 1.37). In general, respondents in the US sample had higher values at the end of the scale, and Dutch samples had higher values at the start of the scale.

Our study explores the mechanism of the third person effect driving the relationship between gratifications of the news and algorithmic appreciation. For that reason, we have used a moderator. As described in the theory section, the third person effect, in this case, is conceptually close to what is also called epistemic overconfidence (e.g., see Kim, 2018; Salmon et al., 2019; Wei et al., 2007). We measured epistemic overconfidence using a Cognitive Reflection Test (CRT) based on the scale developed by Thomson and Oppenheimer (2016) combined with the three standard CRT questions of Toplak et al. (2011). The 7 questions have high face validity, and, in order to address some criticisms of the original CRT, do not require a high degree of mathematical sophistication to generate the correct answer. After the CRT we asked respondents to estimate how many questions they answered correctly, thereby assuming that

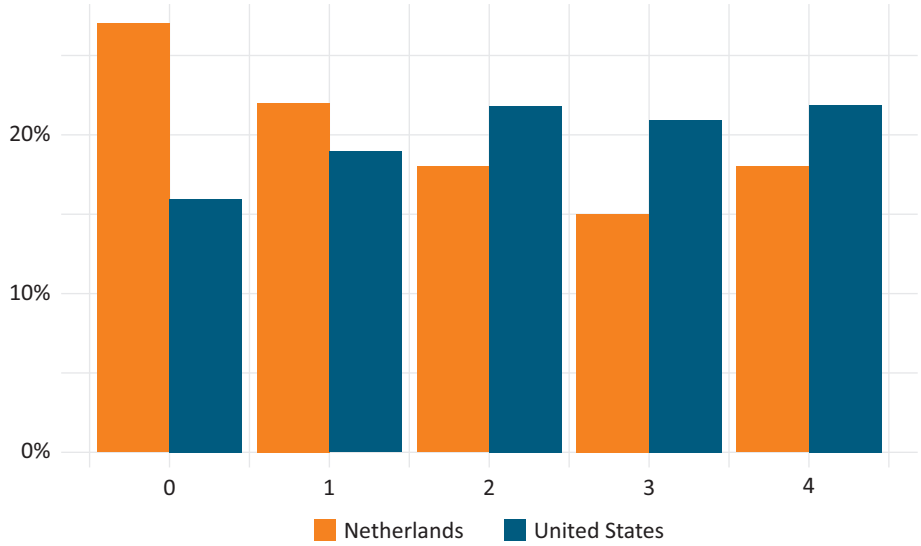


Figure 1. Descriptive information of the dependent variable. Notes: Mean = 1.92; SD = 1.43.

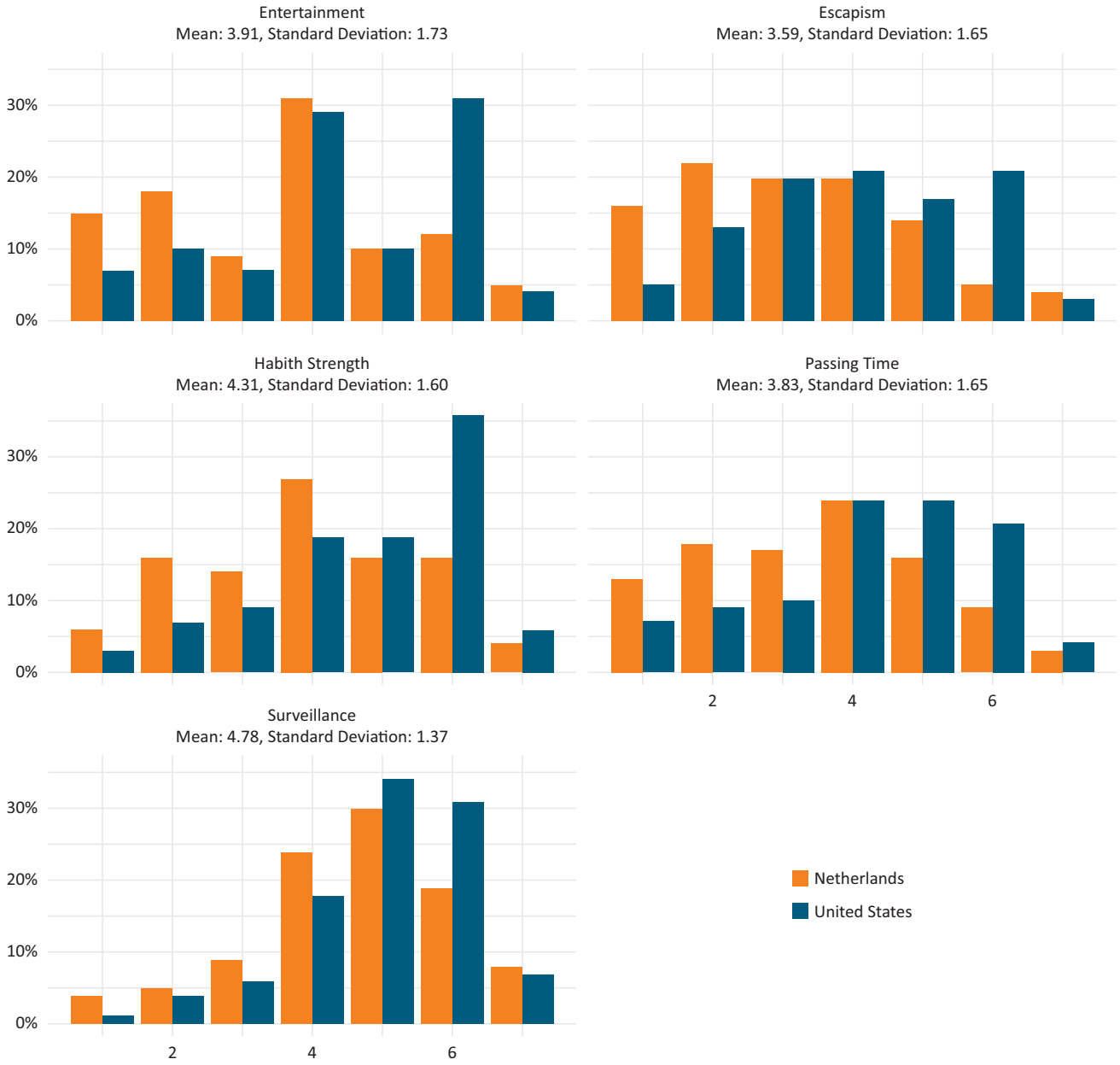


Figure 2. Descriptive information of the independent variables.

those that are overconfident give more intuitive (and thus false) answers while estimating that they have all or the majority of the questions correct. To measure epistemic overconfidence, we subtracted the number of correct answers people think they had given on the CRT to the actual number of correct answers. This could range from −7 (respondent answered all questions correct but estimated that none were correct) to 7 (respondent answered none of the questions correct but estimated that all were correct). Hence a higher number indicates higher levels of epistemic overconfidence, and 0 is the actual middle point (number of questions correctly estimated). On average, people overestimate their capabilities by 2 questions (M = 2.48, SD = 1.96), with 75% of the respondents ranging from slightly overestimating their capability (score of 1) to overestimating their capabilities by 4 questions (score of 4). The sample is slightly skewed towards people being overconfident, especially the majority in the Dutch sample was slightly overconfident (score of 2).

In our analysis, we controlled furthermore for Frequency of News Usage, Political Efficacy, and Trust in Media (see our online compendium for the visualization in the Supplementary Material). First, we measured Frequency of News Usage by asking respondents on an 8-point scale (0 being *never* and 7 being *every day*) how many days of the week they consume news in 5 different ways. The additive scale of the 5 items has a Cronbach's α value of 0.71. On average, people consume news on approximately 4 days of the week (M = 3.79, SD = 1.63). Second, we measured Political Efficacy (lower-left panel in Figure 4) using a combined knowledge and

efficacy seven-point scale, ranging from 1 (*completely disagree*) to 7 (*completely agree*). The additive scale of the 3 items has a Cronbach's *α* value of 0.57. On average, people have a high level of political efficacy (M = 4.41, SD = 1.28)—this is especially driven by the participants of the US sample. Third, we measured Trust in Media by asking respondents how much they (dis)agree with 9 statements regarding various gatekeepers on a seven-point scale, ranging from 1 (*completely disagree*) to 7 (*completely agree*). The additive scale of the 9 items has a Cronbach's *α* value of 0.85 (M = 4.91, SD = 1.10).

To test our hypotheses—i.e., whether algorithmic appreciation is dependent on the gratification of the news (H1) and whether that relation is moderated by epistemic overconfidence (H2)—and to conduct exploratory analyses, we use OLS regression analyses (Bryman, 2016).

4. Results

4.1. Which News Consumers Prefer Algorithmically Curated News?

In this section, we start by exploring the bi-variate relationships between algorithmic appreciation and gratifications of the news. Figure 3 shows the level of algorithmic appreciation (Y-axis)—0 equals *never the preferred gatekeeper*, 4 equals *always preferred gatekeeper*—for

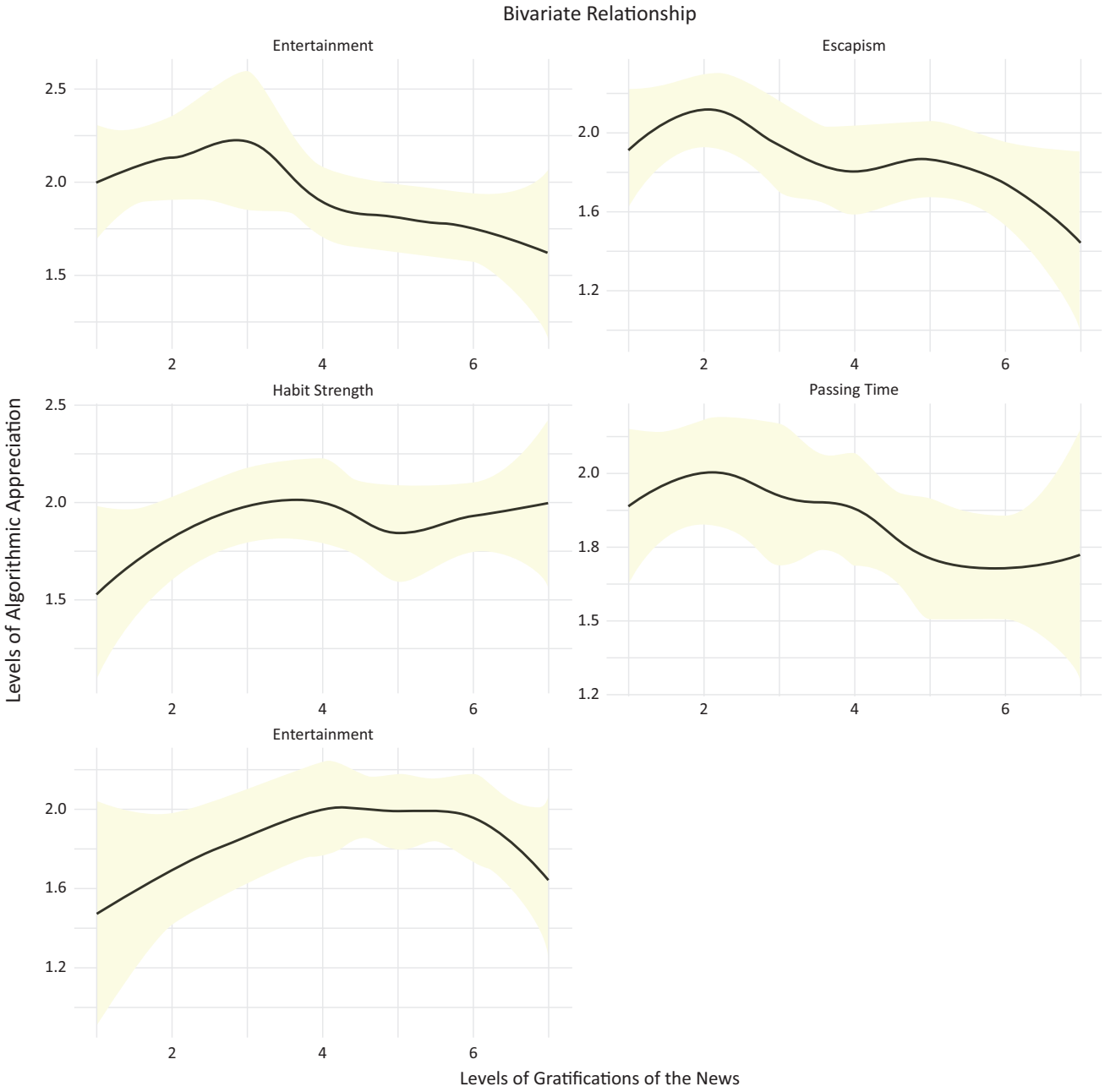


Figure 3. Bi-variate relations between algorithmic appreciation and gratifications of the news.

each gratification (X-axis). The black solid line depicts a LOESS smoothening (locally weighted smoothing) to visualize the relationship between the two variables. The yellow area surrounding the line is the 95% confidence interval. As the panels for Entertainment (upper-left), Escapism (upper-right), and Pass Time (middle-right) of Figure 3 demonstrate, the relationship between these gratifications of the news and algorithmic appreciation is linear and negative. This means that the more you use news to entertain yourself, out of escapism, and/or to pass time, the less likely you are to appreciate an algorithmic gatekeeper (i.e., rank one of the two algorithmic news selections first). The middle-left panel of Figure 3 shows that there is no relationship (or a slightly positive one) between Habit Strength and Algorithmic Appreciation. The lower-right panel of Figure 4 shows that for Surveillance, the relationship is actually positive, but not linear: Meaning that the more you use to keep up-to-date with politics, the more often you rank an algorithmic gate-keeper first, except for when you use surveillance as a gratification all days of the week. This positive relationship is surprising and actually the opposite of what is hypothesized. Based on the theory, we had hypothesized that when the gratification sought for news is surveillance people will show less algorithmic appreciation (see H1).

To see whether this contradictory finding holds while controlling for other variables, we conduct an OLS regression. Figure 4 visualizes the regression effects, the full model is displayed in Model 1 of Table B1 in Appendix B (see Supplementary Material). As Figure 4 below shows, the news gratifications dimensions Escapism, Entertainment, and Passing Time have a negative, yet statistically insignificant, effect on Algorithmic Appreciation after controlling for all the other variables—Entertainment is significant on a 10% α-level. Habit Strength and Surveillance have a positive effect on algorithmic appreciation, with both of them being statistically significant on the 10% α-level. This indicates opposite results for our H1, which stated people show less algorithmic appreciation for news selection when the gratification sought for news is surveillance compared to the other gratifications. None of the control variables have a statistically significant effect on algorithmic appreciation. We do see that on average, the US sample has higher levels of algorithmic appreciation than the Dutch sample.

In a second step, we interact gratifications of the news with epistemic overconfidence. In line with the recommendations of Brambor et al. (2006) and Holbert and Park (2019), we calculate and visualize the predicted effects and standard errors (α-level of 0.05%) in Figure 5. The full models are displayed in Model 2 till Model 6 of Table B1 and B2 in Appendix B (see Supplementary Material). Figure 5 shows that for the dimensions Habit Strength and Surveillance for the overconfident respondents, the more they use these gratifications for news sought, the higher levels of algorithmic appreciation they have. The opposite holds for insecure respondents. These interactions are statistically significant on the 10% α-level. We originally had preregistered the interactions to be considered significant at the 5%, yet the authors' growing awareness of statistical power in interactions (e.g., Franzese & Kam, 2009) led us to deviate from the pre-registered plan and report the 10% α-level as support for our second (H2). For the dimensions of Entertainment, Escapism, and Passing Time, we do not observe a different trend for overconfident and insecure respondents.

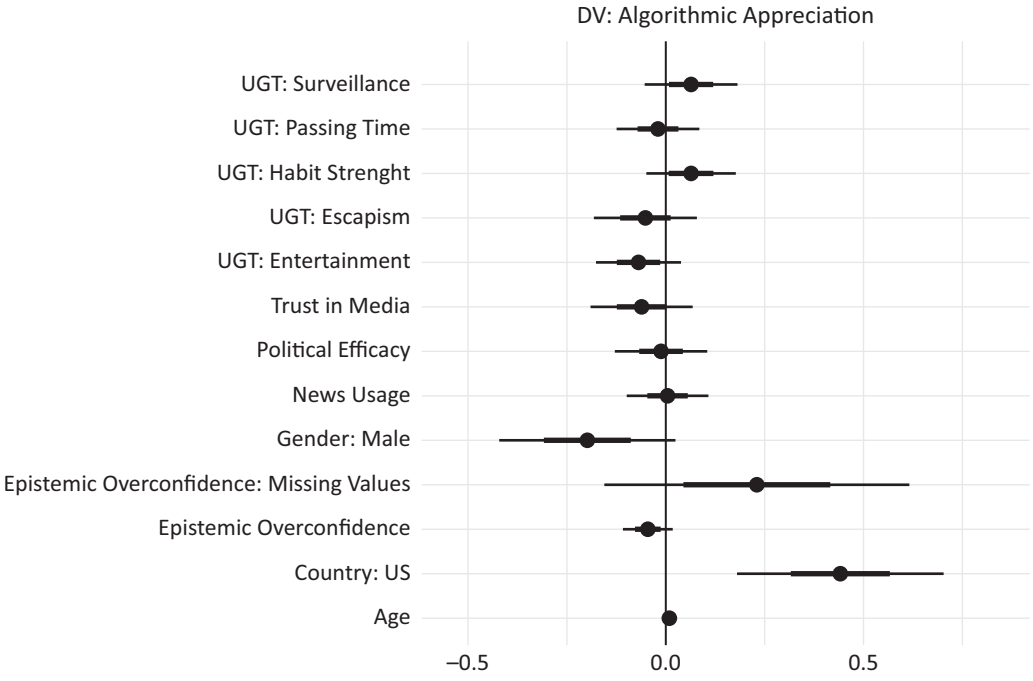


Figure 4. Predicting algorithmic appreciation.

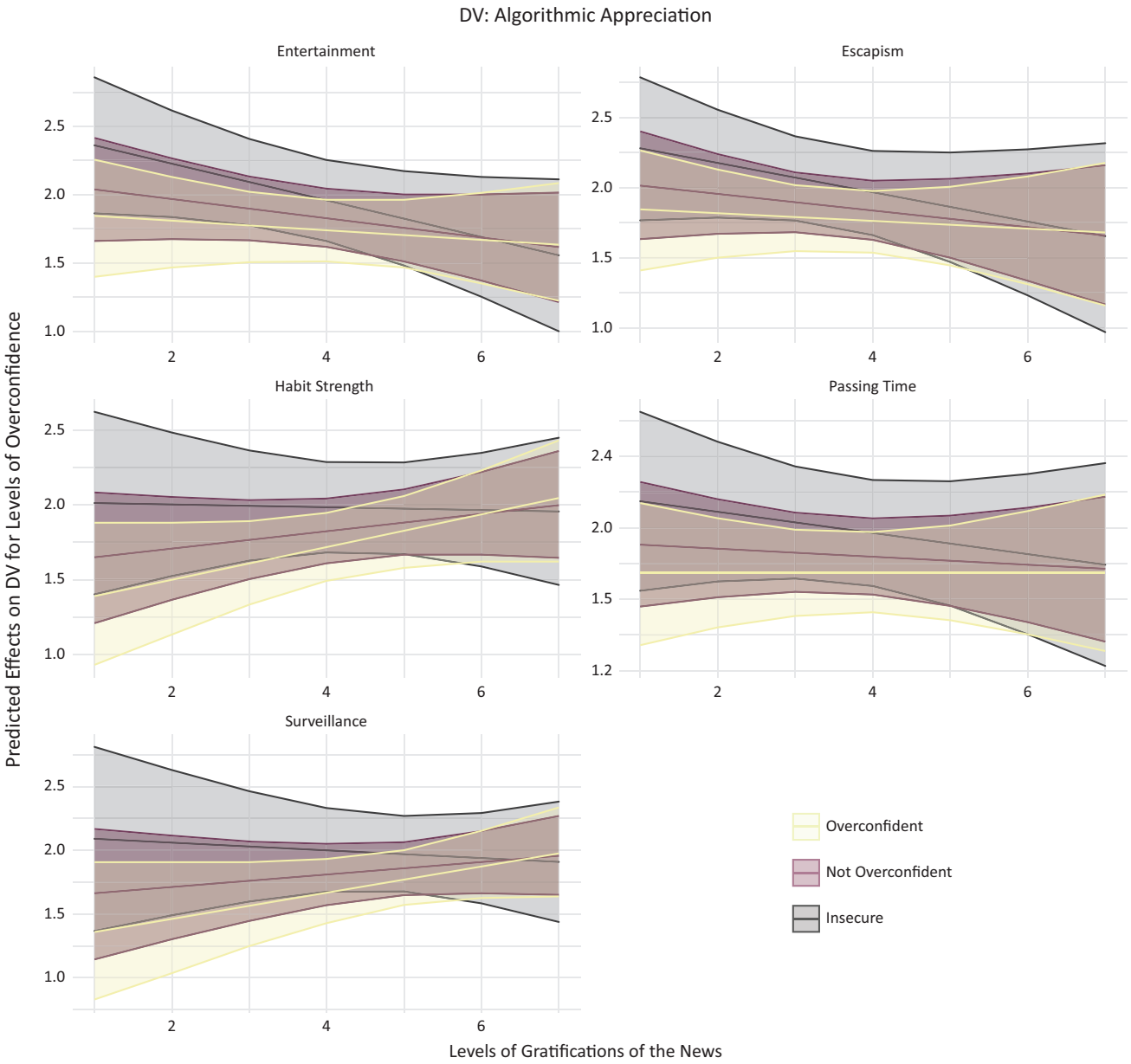


Figure 5. Predicting algorithmic appreciation for different levels of epistemic overconfidence.

4.2. Exploratory Relationships

To explore individual differences for appreciation of news selectors, we look at the role of gender in the relationship between overconfidence and algorithmic appreciation. Moreover, we look at how general trust in media coincides with appreciation for news selectors, and whether our proposed relationships in H1 and H2 also hold for the other types of news selectors.

4.2.1. Gender

While insufficiently powered to slice up the sample once more, Figure C1 in Appendix C (se Supplementary Material) demonstrates that patterns for female (left-hand panel) and male respondents (right-hand panel) are similar. Hence, the effect of H2 is not driven by (fe)males being more overconfident.

4.2.2. Trust in Media

The main analysis reported in Figure 4 demonstrates that the higher levels of generic trust in media, the lower the levels of algorithmic appreciation. Looking at the bi-variate relationships between news selectors (X-axis) and trust in media (Y-axis), Figure 6 demonstrates that for journalistic and algorithmic appreciation (top-left and bottom-left panel) the relationship is curve-linear. People with no and high levels of appreciation have lower trust in media. Most likely people who have high confidence in the journalistic system still prefer experts and journalists as gatekeepers compared to algorithmic

Figure 6. Bi-variate relations between news selectors and trust in media.

solutions. Additionally, there seems to be a group that has low levels of trust in all of the news selector options presented (traditional media or algorithms)—a group that likely proves to be interesting in the context of alternative media and sources.

When this bi-variate relationship is tested while controlling for additional variables in a regression analysis, Figure 7 (see also, Table C1 of Appendix C in the Supplementary Material) shows that for an α-level of 10%, compared to people who appreciate algorithmic news selectors, people who like to select their own news have higher trust in media. The same holds for people with higher levels of overconfidence. This exploratory result demonstrates that the role of overconfidence, or the third person effect, is potentially important to understanding how people think about news and particularly the ways in which the news is selected.

4.3. Other News Selectors

Lastly, we explore the relationship between UGT and the appreciation of other news selectors—the models can be found in Appendix C in the Supplementary Material. Figure 8 demonstrates in the left-hand panel that for the UGT dimension Surveillance, there is a positive relationship between preferring to select your own news

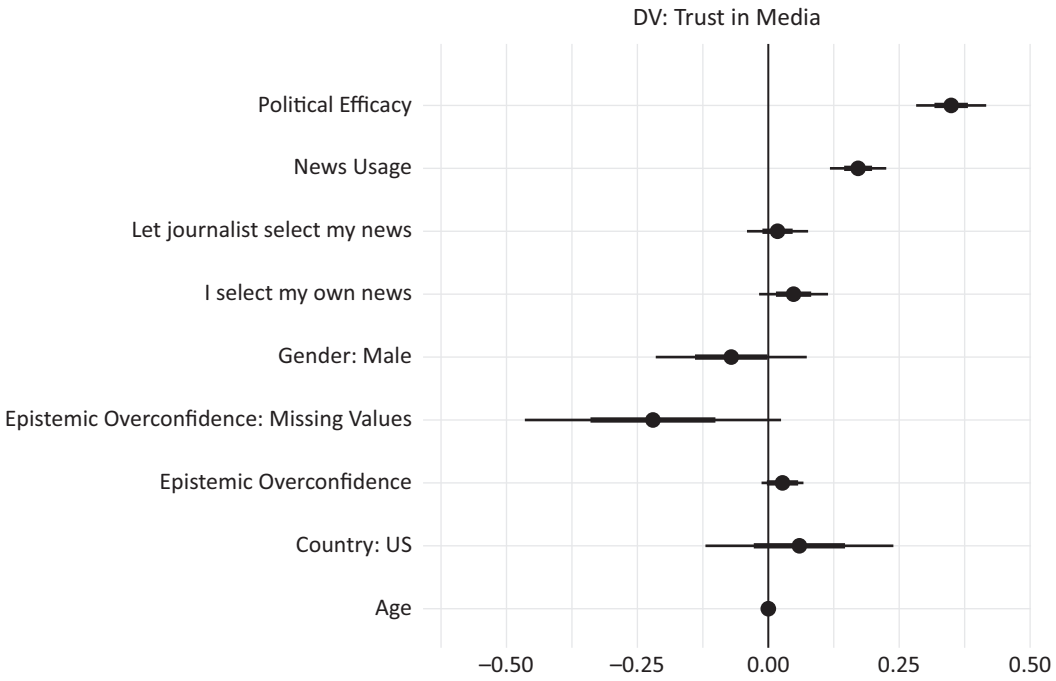


Figure 7. Predicting trust in media.

and using news for keeping up with politics. This indicates that the more you use the news to keep up-to-date with political information, the more you like to select the news yourself (significant at the 10% α-level). The relationship between UGT dimension Escapism and selecting your own news as preferable is negative: The more you use the news for the gratification Escapism the less you prefer to select the news yourself (significant at the 10% α-level). The other dimensions center around 0, indicating no effect. For appreciation of journalistic (human)

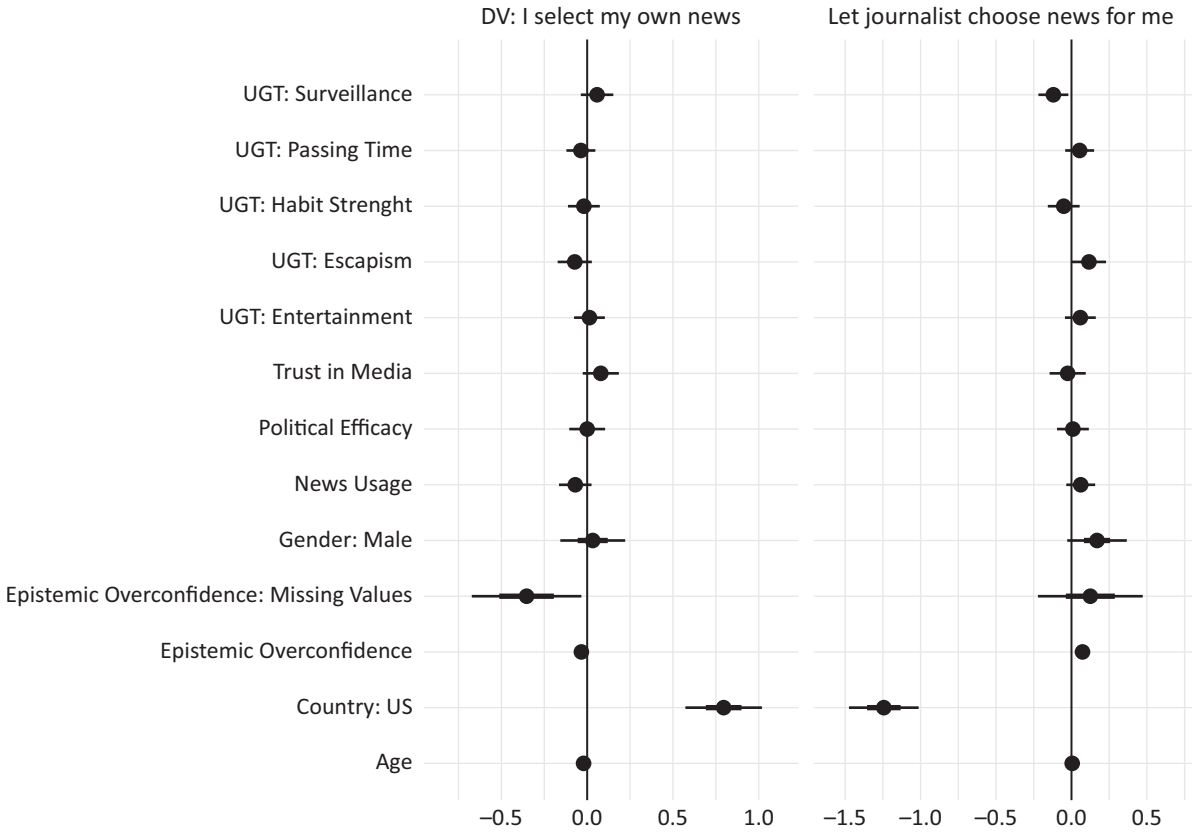


Figure 8. Predicting appreciation for other news selectors.

selectors, the right-hand panel of Figure 8 shows that for the UGT dimension Surveillance and Habit Strength that the more you use the news as a strong habit and for keeping up with politics, the less you appreciate journalistic news selectors. You do have appreciation for these news selectors when the gratifications sought for are Passing Time, Escapism, and Entertainment.

Looking at the moderation of overconfidence for both other news selectors, Figure 9 shows that for respondents with high levels of overconfidence, they appreciate journalistic news selection less when using news more often for surveillance gratifications (bottom-left panel). Figure 10 demonstrates that overconfidence does not play a role for people who appreciate to self-select the news. Across all dimensions of UGT the patterns of insecure and overconfident respondents are similar.

5. Discussion

In this article, we investigated to what extent appreciation for algorithms as news gatekeepers is influenced by gratifications sought. We furthermore proposed that this relation is dependent on people's overconfidence in their cognitive abilities. Our analysis of 652 participants demonstrates that the gratifications the news is sought for matter for which gatekeeper people prefer for the selection of news articles. The main analysis (reported in Figure 6 and Table B1 in the Supplementary Material) demonstrates that the gratification of habitual

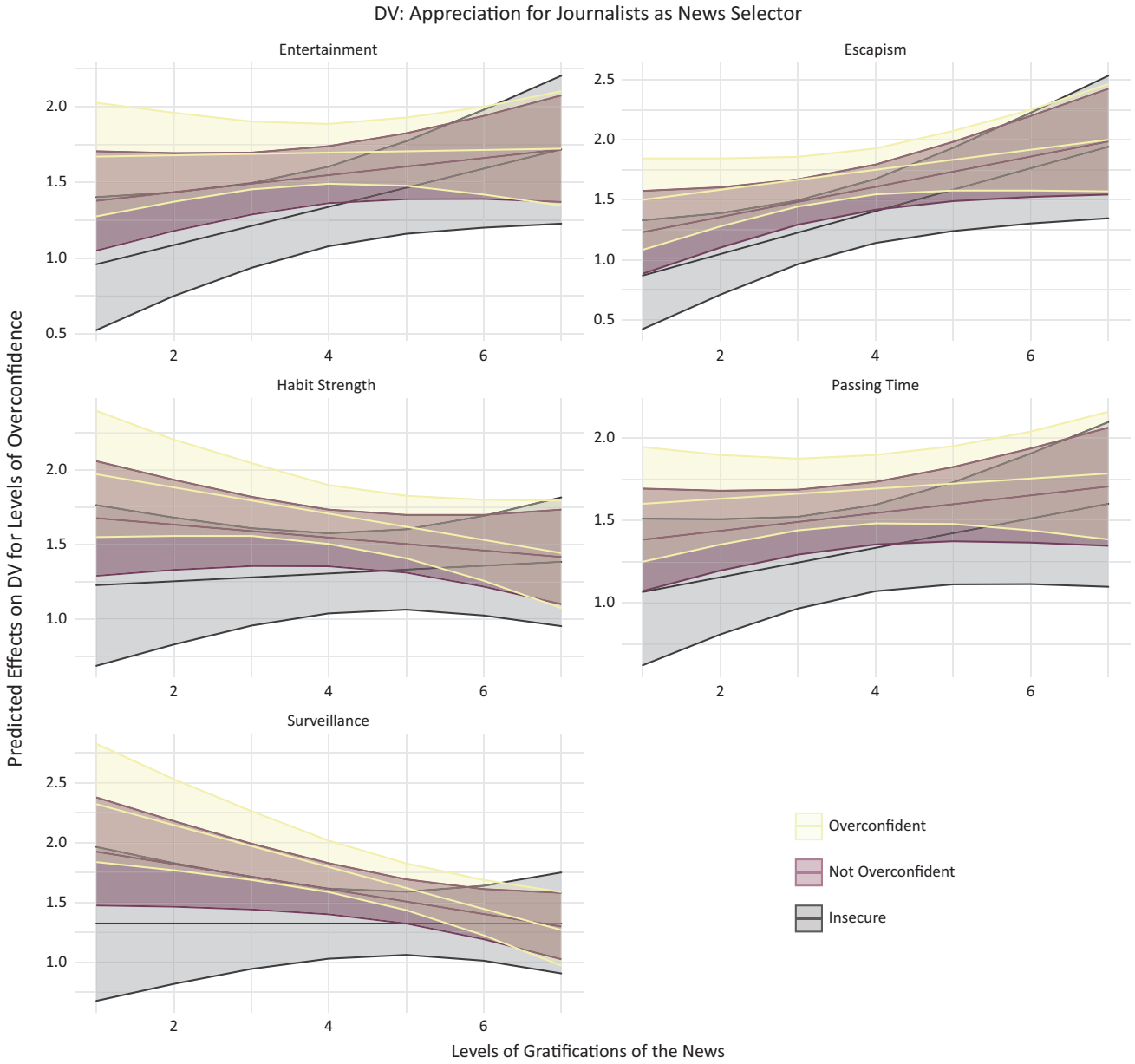


Figure 9. Predicting journalistic appreciation for different levels of epistemic overconfidence.

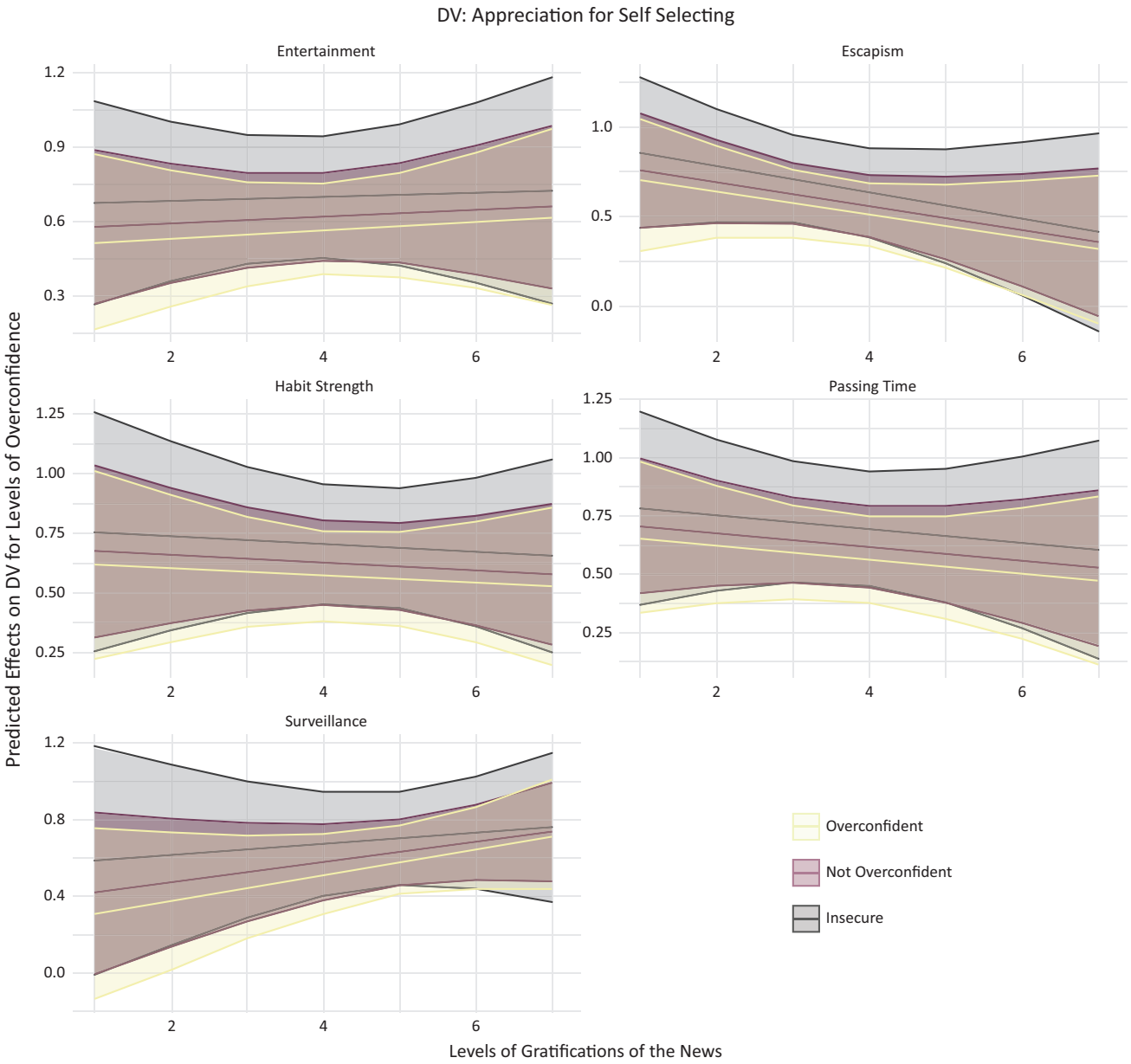


Figure 10. Predicting appreciation for self-selecting news for different levels of epistemic overconfidence.

news usage has a significant positive influence on algorithmic appreciation. This unexpected result might indicate that routine processes and frequent usage of news might depend to a larger extent on algorithmic gatekeepers than when news is consumed for other (more conscious) reasons. Algorithms mostly help consumers to make decisions easier and faster, thus they play a larger role when looking for news with no particular purpose. We hypothesized that when the main aim of news consumption is surveillance, the quality, accuracy, and diversity of information is of rather high importance for the user, calling for relying on expert judgments (i.e., journalists) for determining what is relevant. In contrast, when news consumption is aimed at passing time, entertainment, and escapism, getting information easier, faster, and more specifically targeted might play a larger role. Therefore, the appreciation for expert (human) judgment should be higher when news is consumed for surveillance gratifications while for the other gratifications algorithmic gatekeepers are preferred. The results actually demonstrate the opposite: The more you consume news to pass time, escape from daily worries, or for entertainment, the less likely you are to prefer algorithmic news selection. However, for surveillance gratifications (keeping up with politics), algorithms are rather appreciated than feared. Two alternative explanations for those results come to mind: On the one hand, people get their news for passing time, escaping, or entertainment from platforms where they are less aware of the algorithmically curated processes (e.g., social media) or from preferred websites where the expert curation is key. Especially when it comes to entertainment, getting

new and surprising content that possibly deviates from what was consumed before might be an added bonus. On the other hand, it might be that when the stakes are high (forming your political opinion) algorithms are seen as more "neutral"—in line with the machine heuristics, such that they can be trusted to not be influenced by anything other than the facts. This is of course a grave misconception of the nature and workings of news algorithms—depending on the (biased) data sources the algorithm was trained on and the design of the algorithm (what is it optimized for), algorithms can lead to biased or non-diverse sets of news recommendation. However, this conception might, especially in a polarized media environment, lead to more trust in an algorithm than in a journalist. Indeed we can see in our exploratory results that a negative relationship between using news for surveillance reasons and having a journalist as gatekeeper can be found. A general distrust in the media system and journalists to deliver "unbiased" news might bring people to rely even more on algorithmic gatekeepers when important information is concerned. Journalists can be trusted with the "soft side" of the news, but are seen as skeptical when it comes to providing political information.

Our second finding showcases that users with high levels of confidence in their own abilities are more likely to prefer algorithmic gatekeepers for surveillance gratifications, as we expected in Hypotheses 2. Highly confident—or overconfident—individuals might rather prefer to have a system that learns from their own decisions instead of having other people (i.e., journalists) decide upon their news consumption. The notion of algorithms being more "neutral" (as part of the machine heuristic), seeing them as passive amplifiers of one's own thoughts, might appear especially appealing to those estimating their own cognitive abilities as high. When expecting less of a negative influence on oneself, the notion of optimistic bias (being less vulnerable to malicious intents) has often been proposed as playing an important role (Salmon et al., 2019; Wei et al., 2007). Therefore, one's self-conception as being superior to others regarding the gullibility of negative effects can be seen as one main driver of the third-person effect.

While this small, and low-powered, study has given us some proof of concept for the puzzle of why people might prefer algorithmic news *and* at the same time fear a polarized and "filter bubbled" society, there are several limitations. First of all, the MTurk and Pollfish samples are not representative samples of the US and Dutch populations. In other words, we have to be careful when interpreting results. Moreover, while an interaction effect is able to give some indications for the mechanism underlying algorithmic appreciation, more evidence needs to be brought to the table. This will also involve looking more in detail at good measurements of preferences for gatekeepers—how much information do people get about the kind of gatekeepers (human or algorithmic) they can select from? In this study, we added additional explanations for the algorithmic gatekeepers to show based on which data selections could be made—which already seems to imply a specific reasoning going beyond what human editors are doing. In how far the "motives" of selecting certain news should be included in the items or even systematically varied should be further explored, especially regarding the influences on trustworthiness or objectivity evaluations of the gatekeeper.

This study showed that in the context of gatekeeper preference often the question is less what specific conscious reasons people have to search for news but rather whether they follow habitual patterns instead of searching for particular gratifications. Given the cognitive load of more and more information, leaving the hard work of pre-selection to algorithms will likely gain in importance. However, we also showed that when users have specific goals in mind for their news search, different patterns of gatekeeper preferences occur—being moderated by variables such as confidence in one's own abilities, but also possibly related to trust in the media system. This opens two interesting avenues for future research: Firstly, understanding more about the relation of habitual usage and the use of algorithms as gatekeepers (algorithmic filtering as routine). Secondly, tapping into the gatekeeper decisions users make when they do have a specific purpose for their news search in mind and are thus possibly more likely to pay closer attention to the information they find. The exploratory analyses demonstrate that in order to better understand the preference for news selectors individual characteristics play an important role. Our study gives preliminary insights into this process. We argue that to better understand the relationship, more and higher-powered studies need to be conducted.

Acknowledgments

Earlier versions of this article have been presented at the CommunicatieEtmaal 2020 in Amsterdam, the ICA 2020 online version, and the VU Political Communication Research Group. We thank all participants for their useful comments and suggestions. We would also like to thank the editors of *Media and Communication*, and the journal's anonymous reviewers for the many constructive comments and suggestions. This work was supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek [VI.Veni.191R.006].

Conflict of Interests

The authors declare no conflict of interests.

Supplementary Material

All data and analytical code can be found in the following links: https://github.com/MarikenvdVelden/Epistemic-Overconfidence-in-Algorithmic-News-Selection; https://osf.io/2tqwy

References

Blumler, J. G. (2019). Uses and gratifications research. In T. P. Vos, F. Hanusch, D. Dimitrakopoulou, M. Geertsema-Sligh, & A. Sehl (Eds.), *The international encyclopedia of journalism studies* (pp. 1–8). Wiley. https://doi.org/10.1002/9781118841570.iejs0032

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, *15*(3), 209–227. https://doi.org/10.1007/s10676-013-9321-6

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis*, *14*(1), 63–82.

Bryman, A. (2016). *Social research methods*. Oxford University Press. vCarlson, M. (2019). News algorithms, photojournalism and the assumption of mechanical objectivity in journalism. *Digital Journalism*, *7*(8), 1117–1133.

Choi, J. (2016). Why do people use news differently on snss? An investigation of the role of motivations, media repertoires, and technology cluster on citizens' news-related activities. *Computers in Human Behavior*, *54*, 249–256.

Coppock, A. (2019). Generalizing from survey experiments conducted on mechanical turk: A replication approach. *Political Science Research and Methods*, *7*(3), 613–628.

Davison, W. P. (1983). The third-person effect in communication. *Public Opinion Quarterly*, *47*(1), 1–15.

Diddi, A., & LaRose, R. (2006). Getting hooked on news: Uses and gratifications and the formation of news habits among college students in an internet environment. *Journal of Broadcasting & Electronic Media*, *50*(2), 193–210.

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society*, *21*(5), 729–745.

Dunning, D. (2011). The dunning–kruger effect: On being ignorant of one's own ignorance. In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 247–296). Elsevier.

Franzese, R. J., & Kam, C. (2009). *Modeling and interpreting interactive hypotheses in regression analysis*. University of Michigan Press.

Geiß, S., Magin, M., Jürgens, P., & Stark, B. (2021). Loopholes in the echo chambers: How the echo chamber metaphor oversimplifies the effects of information gateways on opinion expression. *Digital Journalism*, *9*(5), 660–686.

Gil de Zúñiga, H., Weeks, B., & Ardèvol-Abreu, A. (2017). Effects of the news-finds-me perception in communication: Social media use implications for news seeking and learning about politics. *Journal of Computer-Mediated Communication*, *22*(3), 105–123. https://doi.org/10.1111/jcc4.12185

Harari, Y. N. (2020, December 17). Als de wereld één grote samenzwering lijkt [What if the world looks like one big conspiracy]. *de Volkskrant*. https://www.volkskrant.nl/columns-opinie/als-de-wereld-een-grote-samenzwering-lijkt~b061a6b9

Helberger, N. (2020, July 2). Challenging rabbit holes: Towards more diversity in news recommendation systems. *Blogs LSE*. https://blogs.lse.ac.uk/medialse/2020/07/02/challenging-rabbit-holes-towards-more-diversity-in-news-recommendation-systems

Holbert, R. L., & Park, E. (2019). Conceptualizing, organizing, and positing moderation in communication research. *Communication Theory*, *30*(3), 227–246.

Karimi, M., Jannach, D., & Jugovac, M. (2018). News recommender systems: Survey and roads ahead. *Information Processing & Management*, *54*(6), 1203–1227.

Katz, E., Blumler, J. G., & Gurevitch, M. (1974). *The uses of mass communications: Current perspectives on gratifications research*. SAGE.

Kim, J. W. (2018). They liked and shared: Effects of social media virality metrics on perceptions of message influence and behavioral intentions. *Computers in Human Behavior*, *84*, 153–161.

Lee, A. M. (2013). News audiences revisited: Theorizing the link between audience motivations and news consumption. *Journal of Broadcasting & Electronic Media*, *57*(3), 300–317.

Lee, A. M., & Chyi, H. I. (2015). The rise of online news aggregators: Consumption and competition. *International Journal on Media Management*, *17*(1), 3–24.

Li, X. (2013). Innovativeness, personal initiative, news affinity and news utility as predictors of the use of mobile phones as news devices. *Chinese Journal of Communication*, *6*(3), 350–373.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103.

Nechushtai, E., & Lewis, S. C. (2019). What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior*, *90*, 298–307.

Newman, N., Fletcher, R., Levy, D., & Nielsen, R. K. (2015). *The Reuters Institute digital news report 2016*. Reuters Institute for the Study of Journalism.

Ng, Y.-L., & Zhao, X. (2020). The human alarm system for sensational news, online news headlines, and associated generic digital footprints: A uses and gratifications approach. *Communication Research*, *47*(2), 251–275.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgement*. Prentice-Hall.

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

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In F. Ricci, L.

Rokach, & B. Shapira, P. B. Kantor (Eds.), *Recommender systems handbook* (pp. 1–35). Springer.

Rose, K. (2019, June 8). The making of a youtube radical. *The New York Times*. https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html

Ruggiero, T. E. (2000). Uses and gratifications theory in the 21st century. *Mass Communication & Society*, *3*(1), 3–37.

Salmon, C. T., Poorisat, T., & Kim, S.-H. (2019). Third-person effect in the context of public relations and corporate communication. *Public Relations Review*, *45*(2), Article 101823.

Schipper, N. (2020, November 3). Pizzagate, deep state, 9/11 en corona: Waarom complottheorie QAnon zo populair is [Pizzagate, deep state, 9/11 and Covid: Why the conspiracy theory QAnon is so popular]. *Trouw*. https://www.trouw.nl/cultuur-media/pizzagate-deep-state-9-11-en-corona-waarom-complottheorie-qanon-zo-populair-is~bc19b11b

Shoemaker, P. J., & Vos, T. (2009). *Gatekeeping theory*. Routledge.

Sundar, S. S. (2008). The main model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). MIT Press.

Sunstein, C. R. (2009). *Republic.com 2.0*. Princeton University Press.

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99.

Thurman, N., Moeller, J., Helberger, N., & Trilling, D. (2019). My friends, editors, algorithms, and I: Examining audience attitudes to news selection. *Digital Journalism*, *7*(4), 447–469.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275–1289.

Wei, R., Lo, V.-H., & Lu, H.-Y. (2007). Reconsidering the relationship between the third-person perception and optimistic bias. *Communication Research*, *34*(6), 665–684.

Welbers, K., van Atteveldt, W., Kleinnijenhuis, J., & Ruigrok, N. (2018). A gatekeeper among gatekeepers: News agency influence in print and online newspapers in the Netherlands. *Journalism Studies*, *19*(3), 315–333.

Wölker, A., & Powell, T. E. (2020). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism*, *22*(1), 86–103.

Wong, J. C. (2020, June 25). Down the rabbit hole: How QAnon conspiracies thrive on facebook. *The Guardian*. https://www.theguardian.com/technology/2020/jun/25/qanon-facebook-conspiracy-theories-algorithm

About the Authors

Mariken van der Velden is an assistant professor of political communication in the Department of Communication Science at the Vrije Universiteit Amsterdam. She received her PhD in Political Science from the Vrije Universiteit Amsterdam. Her research interests comprise the areas of political communication, political behavior, and computational social science.

Felicia Loecherbach is a PhD candidate in the JEDS Tracking the Filter Bubble project at the Department of Communication Science at the Vrije Universiteit Amsterdam. She received her MA in Communication Science from the University of Amsterdam. She is studying the diversity of issues and perspectives in (online) news and how it is affected by recommender algorithms and selective exposure.

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 4, Pages 198–207
https://doi.org/10.17645/mac.v9i4.4173

Article

When Algorithms Recommend What's New(s): New Dynamics of Decision-Making and Autonomy in Newsgathering

Hannes Cools *, Baldwin Van Gorp and Michaël Opgenhaffen

Institute for Media Studies, KU Leuven, Belgium; E-Mails: hannes.cools@kuleuven.be (H.C.), baldwin.vangorp@kuleuven.be (B.V.G.), michael.opgenhaffen@kuleuven.be (M.O.)

* Corresponding author

Submitted: 16 February 2021 | Accepted: 7 August 2021 | Published: 18 November 2021

Abstract
Newsroom innovation labs have been created over the last ten years to develop algorithmic news recommenders (ANR) that suggest and summarise what news is. Although these ANRs are still in an early stage and have not yet been implemented in the entire newsroom, they have the potential to change how newsworkers fulfil their daily decisions (gatekeeping) and autonomy in setting the agenda (agenda-setting). First, this study focuses on the new dynamics of the ANR and how it potentially influences the newsworkers' role of gatekeeping within the newsgathering process. Second, this study investigates how the dynamics of an ANR could influence the autonomy of the newsworkers' role as media agenda setters. In order to advance our understanding of the changing dynamics of gatekeeping and agenda-setting in the newsroom, this study conducts expert interviews with 16 members of newsroom innovation labs of *The Washington Post*, *The Wall Street Journal*, *Der Spiegel*, the BBC, and the Bayerische Rundfunk (BR) radio station. The results show that when newsworkers interact with ANRs, they rely on suggestions and summaries to evaluate what is newsworthy, especially when there is a "news peak" (elections, a worldwide pandemic, etc.). With regard to the agenda-setting role, the newsworker still has full autonomy, but the ANR creates a "positive acceleration effect" on how certain topics are put on the agenda.

Keywords
agenda-setting; algorithmic news recommenders; gatekeeping; newsroom innovation labs

Issue
This article is part of the issue "Algorithmic Systems in the Digital Society" edited by Sanne Kruikemeier (University of Amsterdam, The Netherlands), Sophie Boerman (University of Amsterdam, The Netherlands) and Nadine Bol (Tilburg University, The Netherlands).

1. Introduction

Newsroom innovation labs have been created over the last ten years to develop algorithmic news recommenders (ANR) that suggest and summarise what news is. With the help of intelligent technologies, these algorithmic recommenders are increasingly being deployed in the news ecosystem, where tools such as Mode (e.g., translating and restructuring stories) and Starfruit (e.g., summarising news stories) are used to make news recommendations based on data (Beam & Kosicki, 2014; Molumby, 2020; Nechushtai & Lewis, 2019; Ricci et al., 2011). Due to the newness of these ANRs, it is unclear whether they will create new dynamics or paths for how news reaches the journalist. On the one hand, some point out that these ANRs and recommendations may lead to a decrease in the quality of the news on offer, or result in more polarisation (Helberger, 2019; Pariser, 2011). On the other hand, others point to the positive repercussions of the implementation of such ANRs, as they can result in finding new angles or cause more interaction with readers (Beckett, 2019).

The 2021 Reuters report concludes that three quarters of the editors and CEOs of news outlets surveyed believe that smart technology such as AI will have the most significant impact on journalism in the next five

years and that this impact will come specifically from news recommender systems (Newman, 2021, p. 30). Although these ANRs are still in an early stage and have not yet been implemented in the entire newsroom, they can create new dynamics within the news ecosystem. By implementing ANRs, these systems have the potential to change both the pace (i.e., the speed of the decision) and the nature (i.e., the choice and selection of relevant articles) of decisions (Bandy & Diakopoulos, 2020; Pavlik, 2000). Indeed, these systems can result in an intrinsic change of crucial work packages—such as deciding what is new(er)—that are usually carried out by newsworkers. It is essential to focus on the concept of (journalistic) autonomy, a theoretical concept often put forward within the field of human-machine interaction and journalism studies. For example, the use of ANRs can result in a loss of journalistic autonomy in the short term because these ANRs can make decisions more accurately and quickly than newsworkers. In the long term, this may result in a change in core journalistic roles such as gatekeeping and agenda-setting (see, for example, Thurman et al., 2019). In addition, the processes in journalism may also change as a result of this new technological application, as ANRs may influence how and where the news moves and goes.

As Shin (2020) notes, research on the role of newsworkers in relation to these ANRs as potential gatekeepers and agenda setters is somewhat limited. In other words, it is less clear how the news employee interacts with the ANR. That is why in this study we focus on the impact of these ANRs by conducting in-depth interviews with newsroom innovation lab employees at *The Washington Post*, *The Wall Street Journal*, *Der Spiegel*, the BBC, and the Bayerische Rundfunk (BR) radio station. These labs are usually part of the overarching newsroom but, unlike the newsroom, they often have a greater affinity with the development and implementation of news tools such as ANRs. The choice to focus on newsroom innovation labs is deliberate, as many of these ANRs are still in a beta phase and are not being used across the newsroom. Lab members—who are, by definition, innovators or at least early adopters—are therefore best placed to assess how these tools are already changing the work and role of the journalist today and how they can and will influence journalism in the near future. By carrying out this study, we want to advance our understanding of the newsworker–ANR interaction, as the issue as to who makes the decisions, selects the news, and sets the digital agenda has become more complex and unclear.

2. Literature Review

2.1. New Actors in the News Ecosystem: Algorithmic News Recommenders

The emergence of new technologies and digital platforms has created a need within newsrooms to innovate continuously and has led to the existence of newsroom innovation labs (Tameling & Broersma, 2013; Thurman et al., 2019). With the further development of AI in the form of machine learning and natural language generation, newsrooms can develop and implement tools that support certain workflows of journalists and in some cases, partially take over specific tasks (Diakopoulos, 2020). One of the tools developed in newsroom innovation labs are ANRs, which are gradually leaving the labs and gradually becoming new links within the news production process.

In order to examine the new dynamics being brought about by these ANRs, it is essential to distinguish between four types of recommender systems. Firstly, some systems make personalised recommendations based on metadata (content-based). Secondly, some ANRs obtain insights based on what other users like to read (collaborative filtering). Thirdly, some algorithms work on data about their users (knowledge-based). Finally, there is a type of ANRs that combine the previous algorithms (see, for example, Helberger, 2019; Karimi et al., 2018). For the scope of this study, we will focus solely on content-based ANRs because they are primarily used in the newsgathering phase. What they actually do is make recommendations to the journalist based on articles, press releases, and other data. In this study, we are not talking about news recommendation systems (such as tools like Chartbeat and SmartOcto), which perform tasks like mapping out the reading habits of the public and, based on that data, making suggestions to the journalist as to whether or not a specific article should be given more prominence on a website or in a newsletter.

These ANRs and the way they collect and analyse data have the potential to serve as a tool for newsworkers, as they are able to present recommendations for news articles to journalists, detect breaking news events and make predictions (Beam & Kosicki, 2014; Diakopoulos, 2019; Marconi, 2020). Due to the novelty of these tools, it is not yet sufficiently clear how these specific ANRs are being used within a news ecosystem, and more specifically, how the newsworker interacts or does not interact with this tool. Since one of the core tasks of media outlets is to provide citizens with accurate information (the so-called watchdog function), it is relevant to investigate whether these ANRs can help or obstruct the newsworker. For example, *The Washington Post* used an algorithmic recommendation system (content-based) called Lead Locator, which displayed the voting results of specific candidates during the November 2020 presidential election. This tool was developed in their dedicated lab that forms part of the general newsroom and uses machine learning to look for outliers in election data by district (Diakopoulos et al., 2020). In doing so, the algorithm compares the data to previous elections in a given locality and, based on that data, writes a short "tip sheet" that the journalist sees in the Lead Locator tool. Based on the available data, the ANR starts making suggestions about what is newsworthy or what could be a possible

lead. The tool and its suggestions may therefore cause the way journalists gather, select, and plan their news to be different than before (i.e., without the ANR). As with Lead Locator, other tools have been developed which, using automated writing, make local stories available to reporters, some of which are then used as leads for additional local coverage.

These ANRs can therefore influence the pace of certain decisions that journalists make on a daily basis and the nature of the choices made, such as selecting certain information or data in order to arrive at an article. The newsworker–ANR interaction is therefore a relevant topic of study, as this interaction has the potential to modify the pace and nature of the work of journalists. As part of such a study, it is also essential to focus on the concept of autonomy that is often put forward within human–machine interaction, as it advances our understanding on how technology can impact journalistic autonomy. Autonomy is defined here as the "freedom that a professional has in performing his or her professional tasks" (Reich & Hanitzsch, 2013, p. 135). As ANRs quietly permeate the news ecosystem newsroom, they have the potential to affect the freedom or autonomy of human newsworkers. Since autonomy has been described as a core value of journalism (Deuze, 2005), this reality could change how autonomy is defined and embedded in the newsroom's daily decision-making. It needs to be rethought in the context of what Splendore (2016, p. 348) called "the increasing intervention of machines." Augmented intervention could mean redefining the autonomy of human newsworkers, as these tools will increasingly take charge of what Diakopoulos (2015, p. 400) called "autonomous decision-making."

This degree of autonomy also relates to McLuhan's (1964) work and is rooted in the theory of technological determinism, a reductionist theory that assumes that a society's technology determines its social structure. Here, the tool used to communicate influences the recipient's mind (Lewis et al., 2019; McLuhan, 1964). In this study, we would therefore like to examine the relationship and interaction between ANRs and newsworkers and evaluate whether these interactions can enhance, or complicate, the performance of their work packages in newsgathering (Milosavljević & Vobič, 2019). As these ANRs assist newsworkers in suggesting, selecting, and summarising news, they may cause new dynamics to emerge in how newsworkers perform their gatekeeping and agenda-setting roles.

2.2. New Dynamics of Decision-Making in Gatekeeping

Nechushtai and Lewis (2019) have pointed out that since the rise of the internet, ANRs can be considered "intervening factors," as they have entered the news ecosystem quite bluntly and influenced the way decisions are made by newsworkers. Indeed, an ANR can help determine which topics to cover and has the potential to influence the news employee's choice about what makes certain information newsworthy. We elaborate on how this gatekeeping role may or may not change if an ANR is able to detect, present, and summarise what news is. After all, if we define gatekeeping as "the process of selecting, writing, editing, posting, scheduling, repeating, and otherwise massaging information to become news" (Shoemaker et al., 2009, p. 73), then it is immediately apparent that the previously described functions of an ANR have a direct link to these journalistic roles and that recommender tools can influence the news production process and the final news output.

Previous studies by Tandoc (2017) and Vu (2014) have demonstrated the impact of metrics tools on editorial choices and pointed to the new dynamics of gatekeeping. Indeed, as these studies have shown, metrics systems (software that links to the news site and indicates, for example, how often an article is clicked on and how long people stay on it on average) can have an impact on news selection and editorial decision making, which can influence news diversity. Algorithmic recommendation systems are also often considered "black boxes," contributing to concerns that news executives will increasingly use them when news outlets are under commercial pressure. This evolution could result in an audience turn, where these recommender systems will be used to distribute personalised content to their online news consumers, leading in some cases to information silos and filter bubbles (Belair-Gagnon & Holton, 2018; Diakopoulos & Koliska, 2017). Tandoc and Thomas (2015, p. 247) concluded that because of algorithmic recommender systems, these filter bubbles could lead to "ghettoizing citizens into bundles based on narrow preferences and predilections rather than drawing them into a community."

If we consider these metrics systems, such as Chartbeat and SmartOcto, as influencing news selection, we can argue that ANRs in particular—which can be considered even more sophisticated as metrics systems—may influence what is "newsworthy." Apart from the abundant research on metrics systems, there has been no research on ANR tools to date, apart from a study by Helberger (2019). She focused on the democratic role of ANRs and argued that these ANRs could create both opportunities and threats when implemented in the news ecosystem. As far as opportunities are concerned, she concluded that ARNs are seen as tools that can reinvent media processes, increase interactivity with readers, and result in news content that is more diverse. As with other algorithmic recommender systems, a potential threat could be the lack of transparency and diversity in disseminating information by those ANRs among journalists and the danger of even greater filter bubbles for online news consumers (Diakopoulos & Koliska, 2017; Nechushtai & Lewis, 2019).

According to Diakopoulos (2019), algorithmic recommender systems can act as gatekeepers because they can process information and data by prioritising, classifying, associating, and filtering it. Through automation, they

can also produce short messages and therefore can suggest or summarise what is newsworthy—think of the "tip sheets" of the Lead Locator—meaning that these ANRs can be seen as even more skilled helpers (or possibly even decision-makers) for the journalist as gatekeepers, compared to the more familiar metrics systems. Since no research has so far been carried out on the interaction between news employees and ANRs, we want to investigate how this interaction takes place within newsgathering. Importantly, we will focus on newsworkers who are members of newsroom innovation labs. These newsworkers are in the cockpit of newsroom innovation and often have different profiles to those of journalists in the broader news ecosystem or newsroom. Because these ANRs are often still in the beta phase, and because the members have a greater affinity with what a tool can and cannot do, they also have a better understanding of how the newsworker–ANR interaction takes place. Because these newsroom innovators are the most knowledgeable about the capabilities, advantages, and disadvantages of these tools, they may also have a better idea of how an ANR might affect a newsworker's daily decisions. This brings us to the first research question: How do members of newsroom innovation labs experience ANRs on the newsworkers' daily decisions in newsgathering?

2.3. New Dynamics of Autonomy

ANRs have the potential to change the role of newsworkers as decision-makers and the way they put something on the "media agenda" (Diakopoulos, 2020). In this study, agenda-setting is seen as the process by which mass media determine what we think and care about (McCombs, 2005). It is how news media come to label topics as more important and present them more prominently, with the result that audiences tend to see these topics as more relevant than others. Like gatekeeping, the agenda-setting role of the media has undergone a new dynamic since the rise of the Internet. New potential agenda setters have appeared on the scene, such as the public and tools that can at least control what goes on the agenda (Denham, 2010; Golan, 2006; McCombs, 2004; Wallsten, 2007). With these new potential agenda setters in mind, scholars have long questioned the agenda-setting power of traditional media. Studies by Brosius et al. (2019) and Tan and Weaver (2007) have analysed the longitudinal evolution of agenda-setting among media and have pointed to the importance of diffusion of agenda-setting in the digital age. In other words, the traditional way of agenda-setting no longer applies online, and the rise of social media platforms has led to a newer process of agenda-setting.

Research by Gleason (2010) and Tandoc and Eng (2017) has shown that the popularity of platforms such as Twitter and Facebook have changed the process of news gathering and how agenda-setting takes place, with the agenda-setting role of traditional media allegedly becoming more diffuse and complex (Weimann & Brosius, 2016). With mass media seemingly struggling to maintain their grip on the public agenda due to increasing selectivity and audience fragmentation, concerns about how news distribution occurs in society have only increased (Feezell, 2018). In short, in a world of ever-evolving digital media, customised news, and fragmentation of online audiences who can choose which news to consume within a high-choice media environment (Van Aelst et al., 2017), it is becoming less and less clear whether a general news agenda still exists and who sets that news agenda.

Therefore, it is essential to look at how ANRs that are developed and implemented to help newsworkers find, select, verify, summarise, and disseminate news can influence the agenda-setting power of news media and journalists. As these ANRs increasingly decide what data and information are shared with newsworkers and assist in the more diffuse and complex process of agenda-setting, these ANRs would, at least partially, determine what is put on the agenda by the news media. Since these ANRs are more sophisticated than metrics systems and can suggest and summarise what is news, it is relevant to look at the news employee's self-perceived decision-making power as an agenda setter.

Although many ANRs are still in their infancy and their success or failure is still uncertain, researchers underscore the importance of newsrooms that adopt a strategy about AI, automation, and computational journalistic tools (Beckett, 2019). Because the relationship of newsworkers and computational journalistic tools is relatively new, journalists will have to learn to "share autonomy," as Deuze (2005) puts it. Building on that development, there could come a time when newsworkers consciously need to outsource their decisions to non-human agents. Helberger (2019) points out that it is crucial for the news employee that an ANR does not encroach on his or her autonomy or decision-making freedom, as this could lead to a breach of trust with the news employee. As a result, the lack of guaranteed autonomy may lead the news employee to decisively reject further interaction with the ANR. This brings us to the second research question: How do members of newsroom innovation labs consider the influence of ANR on the autonomy of setting the agenda for newsworkers?

3. Method

This qualitative study focuses on newsworkers from American, British, and German newsroom innovation labs, as these labs have been working for some time on one or various ANRs. The respondents were selected using the snowball method and focused on the members of the lab that are familiar with ANRs. The sample consists of 16 members from *The Washington Post* (4), the BR radio station (2), *The Wall Street Journal* (2), the BBC (6), and *Der Spiegel* (2). The average age of our respondents is 34.6, with the youngest being 25 and the oldest 61. Almost all members of the sample have

a master's degree and have a background in journalism, computer science, or both.

At the request of some interviewees, we do not mention names in the analyses and use an identifier instead. The interview guide is divided into three parts: During the first part of the questionnaire, we probed for information about the individual news employee, such as job title, background, and responsibilities. In the second part, we focused on their role as gatekeepers and the use of ANR in their daily decision-making in the process of news gathering. The third part of the questionnaire concentrated on how ANRs have a potential impact on the autonomy of the newsworker as an agenda setter. In this series of interviews, the focus is on the interaction with ANRs that are already in use, albeit sometimes still in a beta phase.

A qualitative, descriptive method was used to analyse the interviews (Braun & Clarke, 2006). The 16 interviews lasted one hour on average and were conducted via Skype because of Covid-19. The recordings were transcribed, and the responses were qualitatively coded and analysed to answer the research questions. Particular attention was paid to the statements around ANRs and the role of gatekeeping and agenda-setting. Our main goal during coding was primarily to identify the different ANRs and the variety of potential impacts they have on gatekeeping and agenda-setting, rather than to make a representative estimate of the ANR's impact on the entire news ecosystem.

4. Results

Since we want to map out which ANR are present in the various newsrooms of our sample, and since these ANRs are also related to the results, we will give a brief overview of the ANR here. The BBC uses an ANR called Modus, which helps newsworkers to quickly summarise what is new(s) by displaying the most important highlights of a text or photo in bullet points. At the BR and *Der Spiegel*, they use the same technology for their ANR as they do at BBC. *The Washington Post* has different ANRs but, for the scope of this study, we solely focus on the Lead Locator. This ANR is being used for national and regional elections in the US and suggests potential leads based on clean data. The *The Wall Street Journal* is experimenting with a tool that monitors stock prices and, based on these fluctuations, automatically sends short messages to newsworkers. In this way, newsworkers can be notified more quickly of sudden stock movements. An overview of the results is given below by focusing on both the decision-making and on the autonomy of agenda-setting.

4.1. Algorithmic News Recommendations and Decision-Making in Newsgathering

All newsroom innovation lab members see why an ANR is implemented in the newsgathering phase, but some realise that using an ANR in their daily decisions can be useless too. When they were asked how these ANRs can influence a newsworker's decision-making, they highlighted the fact that they use an ANR especially when there is a so-called "news peak" (e.g., an election night, a worldwide pandemic, other breaking news, etc.). In the case of this peak, respondents are more inclined to delegate part of their decision-making to an ANR. The ANR will help determine what kinds of stories and leads will be used in newsgathering and will lead to potential stories in the phase of news production. The interviews show that the respondents see the ANR as a tool that spurs the decision-making process. Respondents point to a certain level of reluctancy when interacting with the ANR initially, as they have no idea what the features were of these ANRs. As they start to interact with the ANR, this results in a more precise delineation of what the tool can and cannot do. Respondents refer to an ANR as a "shovel" or an "assistant" to uncover potential leads or patterns in newsgathering. Other members of newsroom innovation labs underscore that ANRs can free up time in the newsgathering phase:

> We use the ANR as a shovel to dig up interesting leads from databases and information flows. In the beginning, I was rather hesitant to use the ANR, but when I started interacting with it, I realised it could optimise the decisions I make. (Respondent 12)

> An ANR helped us to cover the riots at the US Capitol as we used several datasets. With the help of this ANR, this journalist went over to three or four editors and said: "Here are the potential storylines we could use….The ANR was definitely good enough to suggest what could be newsworthy. (Respondent 1)

> The process of gatekeeping gets automated or optimised in a way. When journalists interact with ANRs, they will uncover different patterns of what is newsworthy. This will make them more capable of analysing possible leads. The tools could free up time, and the journalists can slow down their newsgathering processes and have a look at where they get their news. (Respondent 16)

When we evaluate the interaction between the newsworker and the ANR in the context of newsgathering, we notice a "trust-distrust dichotomy." In other words, there is a group of respondents who trust ANRs and another group who distrust these tools and how they may impact their decision-making. A group of newsroom innovation lab members points out that newsworkers place great trust in the ANR they use. Without being aware of all the tool's functionalities, they nevertheless start to rely on what the ANR labels as newsworthy and, as one respondent puts it, adopt a kind of "lazy attitude." Respondents in and outside innovation labs place trust in what the tool is selecting, summarising, and suggesting is without

flaws, as it has been presented as a "smart technology." To inform these newsworkers that have "blind faith" in these ANRs, members of specific innovation labs mention that they add a disclaimer on the ANR platform stating the flaws of these systems:

> Some journalists who started using the ANR have become lazy. They think that the algorithm makes no mistakes in summarising or suggesting what news and newsworthy could be. The faith in the tool transcends the criticism of the user. (Respondent 11)

> I was a bit surprised by the fact that other journalists would want to know how these ANRs work, but somehow, they do not care….They use the tool to gather the news and only start complaining when certain features do not work sufficiently….We needed to add a disclaimer to the ANR saying that these tools can make mistakes as well. (Respondent 2)

At the same time, there is a group of newsworkers who do not trust the functionalities of an ANR within newsgathering. According to another group of respondents, this distrust is related to a fear that they "will lose their journalistic autonomy or their editorial control." This group does not believe that the technology and the functionalities of an ANR are sufficient enough to suggest what could be newsworthy. Respondents mention that journalists outside the newsroom innovation labs do not have "the right skill set" to work with these ANRs, resulting in a knowledge gap. This means that the innovation lab members are most familiar with the ins and outs of the ANR but that other newsworkers in the ecosystem do not have that same knowledge to use the ANR responsibly:

> With respect to one ANR project, the algorithm we tested was not sufficient enough, so members of the lab did not want to develop that ANR further. The suggestions sent out to journalists were not interesting enough in the sense that they were like: "Hey, we should write something about this!" At this point, we do not have the right skillset and enough resources for that. (Respondent 8)

> The ANR that we used to suggest what is news is not sufficient enough as the technology for German language models is not advanced today. That is why journalists are not really using it yet….For English, we use the tool to find what could be newsworthy. (Respondent 11)

When we review what can be improved in the ANR-newsworker interaction, one specific leitmotif often recurs: The majority of the respondents point to the importance of the "human in the loop." The members of newsroom innovation labs insist on human control and supervision, even if it turns out that technology in the form of machine learning will ensure that such ANRs will become more complex, advanced, and therefore more autonomous in newsrooms. Some respondents also point out that the lower forms of decision-making (e.g., suggesting what might be newsworthy) can be done by a tool, but the final decision of what is newsworthy remains with the newsworker:

> With all the technology in the world, we still need humans to fight misinformation and disinformation. We aim to contribute to stories that are more personal in the lab….We may have systems that gather the news and show us what could be relevant, but we still need to put a human touch to the story. Technology is just an extra layer to our reporting. (Respondent 3)

> The various labs we created were really there to support the newsroom, to improve the jobs of journalists. If I look at an ANR and a newsworker's interaction, I think that the last one is still in charge….However, the ANR can really be effective in suggesting what news is. (Respondent 10)

> I guess that these ANRs could support the journalist to do the lower forms of decision-making. These tools help the journalist in what is relevant and what could be newsworthy, but the final judgement needs to be made by a journalist. (Respondent 14)

4.2. Algorithmic News Recommendations and Agenda-Setting

If we evaluate the role of ANR on the autonomy of the newsworker as an agenda setter, we can conclude that, according to the members of the innovation labs, ANR does have the ability to influence what "goes on the agenda." We conclude from the analysis of the interviews that a "positive acceleration effect" emerged when it came to the agenda-setting process. Since an ANR can summarise, translate, and label news as newsworthy, the pace and nature of the decisions of what topics are chosen can be accelerated. In doing so, respondents say it is crucial to ensure newsworker autonomy. As some innovation lab members point out, within the process of agenda-setting the newsworker is helped by the ANR in the way it presents different topics to them, potentially increasing the diversity of the news. However, the journalist must have full autonomy to write about topic x or y. If this is not the case, the newsworker will not be inclined to use the ANR:

> We have been really conscientious in the lab about what features to include in the ANR and which not. We want to use those tools as amplifiers for our journalists in order to fulfil their function as agenda setter. (Respondent 13)

> Journalists do not want a tool like an ANR that makes decisions for them. They want autonomy and freedom in how they select the topics and how they put these topics on the agenda. If their freedom is not guaranteed, they will not use the ANR. (Respondent 4)

As with daily decision-making, the majority of the respondents pointed to the technological shortcomings of such ANRs, leading to a "trust–distrust relationship." Thus, apart from the full autonomy that newsworkers want to experience during the process of agenda-setting, another reason why certain ANRs are not used is the fact that this tool does not work sufficiently. We see that this "willingness–reluctance attitude" of newsworkers strongly depends on the type of features of the tool and that this attitude also differs from one editorial office to another. In addition, this "trust–reluctance relationship" has its roots in a more considerable tension between the newsroom innovation lab (which knows how the tool works) and the larger, more conservative newsroom (whose staff do not know how this tool works). In light of this, this full autonomy for the newsworker within the process of agenda-setting will not be safeguarded in the future. This evolution is in turn linked to the reduction of distrust in such ANR as the technology will become increasingly sophisticated and complex:

> As a lab, we are really isolated from the broader newsroom, and that is also why not a lot of journalists used our ANR system….But here is the key thing: When we augment the newsroom with an ANR, we should be able, at all times, to maintain our editorial control. That will give us the autonomy we need to decide which topics we highlight to our audience. (Respondent 7)

> Instead of publishing the news directly produced by an ANR, we have a reporter look at it, supervise it….But in the future, because of technological advancements, journalists will start to trust ANR instead of distrusting it. (Respondent 11)

From the analysis of the interviews, we can infer that the "positive acceleration effect" of ANRs on agenda-setting causes newsworkers to be exposed to more leads, topics, or themes, leading to more news diversity. Respondents speak of "a higher degree of comprehensibility of certain topics," allowing them to make a more informed choice about what to add to the "agenda." Since the algorithm behind these ANRs ensures that specific articles are summarised for the journalist, it is possible for journalists to sift through more information in a short amount of time. Coupled with this, members of the newsroom innovation lab point out that ANRs allow journalists to become better at their jobs:

> Because of the suggestions and the summaries of an ANR, I am able to grasp certain stories better. If I read an interesting lead, I will turn to other reporters to ask them if it would be useful to write an article on that. So, in a way, it smoothens the process of me putting certain topics on the agenda. (Respondent 10)

> Our core problem is that we have to serve everybody as we have this mandate of universality. So, it is essential to have tools like ANR to do our jobs better. If we have tools to cover stories differently and put issues on the agenda via different platforms, it is useful to have more of these ARS around….Amid a global pandemic, there is a lot of data every day, so then we largely rely on our ARS to make our daily coverage more effective and efficient. (Respondent 7)

> Journalism will become more about processing and monitoring information via systems like ANR. Newsworkers will remain the humans in the loop, and they will stay the decision-makers on what to put on the agenda, but the way they weigh their decisions will depend more on ANR and other tools….The news organisations that do not invest in these types of skills and do not train their staff in working with ANR will be left behind. (Respondent 5)

5. Discussion and Conclusion

The main goal of this study was to gain a better understanding of how ANRs are used in newsrooms and how they may influence newsworkers' roles as gatekeepers and agenda setters. The interviews with the 16 members of newsroom innovation labs showed that new dynamics do emerge in how newsworkers make their daily decisions (gatekeeping) and that their (journalistic) autonomy changes when they set the digital agenda (agenda-setting).

The results show that ANRs in newsgathering are mainly used during the "news peak" where, in this situation, more decision-making and agency is attributed to the ANR because it can more quickly make different suggestions about what could be newsworthy. These ANRs are therefore seen as a tool that can help the decision-making process, a tool that can dig up patterns and recognise them quickly. On the other hand, newsworkers are discovering the limitations of the technology that drives ANRs and have stated that (journalistic) autonomy in the interaction with the tool is all-important. The process of decision-making could be further optimised (in terms of accuracy) and automated (in terms of speed) in the future, but that depends on how much journalists want to grant agency to this ANR. This evolution could give a glimpse of the interaction between newsworkers and ANRs in the future, when specific journalistic roles and work packages will be transformed and partially assigned to a tool thanks to these instruments.

From the analysis of the interviews, we can conclude that ANR does create a new dynamic regarding what is "put on the agenda." For example, the ANR is going to

make sure that the leads are more diverse, and this may positively affect the role of the journalist as an agenda setter. One of the key results in terms of agenda-setting is the "positive acceleration effect" that ensures that news leads can offer greater diversity with regard to the topics that are put on the agenda. As with gatekeeping, the autonomy of the newsworker in the process of agenda-setting must be maintained. If the ANR is given partial or complete autonomy over the various topics that appear in the newsroom, the newsroom will begin to distrust the ANR and avoid using the tool. The analysis of the interviews shows that news staff do not want to cede their autonomy to the ANR, which immediately points to the fact that using this ANR is an essential requirement when it comes to agenda-setting. Again, the majority of respondents point to a trust–distrust relationship, as in the process of decision making (gatekeeping). This trust–distrust relationship has its roots in a greater tension between the newsroom innovation lab (which knows how ANRs work) and the larger, more conservative newsroom (which does not know how this tool works). On the one hand, there is a group of news staff that trust ANR almost completely, but at the same time has no affinity for how this ANR works on a technical level (i.e., where the data comes from, how the suggestions are formulated).

This trust can lead to a kind of "laziness" on the part of the newsworker, who assumes that the ANR will make few, if any, mistakes in suggesting and summarising what news is. On the other hand, a group of newsworkers outside the innovation labs are starting to avoid interacting with the ANR precisely because of the technical flaws. The fact that those newsworkers encounter bad news leads or summaries may cause them to distrust the ANR. An analysis of the interviews shows that this sense of distrust is strongly associated with a fear of losing control if the newsworker continues to interact with the ANR. This fear is endorsed by members of the innovation labs and can also be related to the argument that there should always be a "person in the loop" who should act as a final gatekeeper. He or she should continue to act as a gatekeeper to assess what can be labelled newsworthy.

This research has limitations. For example, it is questionable whether the views and opinions of members of newsroom innovation labs can be generalised as representative of the views of the entire editorial staff. Previous research has already shown that members of such labs are more tech-savvy and better informed about the different features of tool x or y (see, for example, Beckett, 2019; Tameling & Broersma, 2013). Therefore, follow-up research could test the differences in views between members of newsroom innovation labs and the broader news ecosystem. This research can start from the concept of autonomy and can therefore advance our understanding of how technology shapes the work processes among newsworkers. In addition, expert interviews have been used to obtain a picture of how this interaction between an ANR on the one hand and a newsworker on the other takes place within various newsroom innovation labs. This method charts the dynamics that occur within specific newsrooms and labs, making the results per news outlet highly context-dependent. Follow-up research could scrutinise a specific newsroom and, through a combination of ethnographic research and expert interviews, examine how such a tool is implemented through the Grounded Theory method (see, for example, Glaser & Strauss, 2017; Urquhart, 2012). This research could contribute to a better understanding of the use and influence of ANRs on the role of newsworkers, both for journalism scholars and journalism professionals.

Acknowledgments

We would like to thank the anonymous reviewers for the useful comments that helped us improve the quality of this article.

Conflict of Interests

The authors declare no conflict of interests.

References

Bandy, J., & Diakopoulos, N. (2020). Auditing news curation systems: A case study examining algorithmic and editorial logic in apple news. In M. De Choudhury (Ed.), *Proceedings of the fourteenth international AAAI conference on web and social media* (pp. 36–47). PKP|PS. https://ojs.aaai.org/index.php/ICWSM/article/view/7277

Beam, M. A., & Kosicki, G. M. (2014). Personalized news portals: Filtering systems and increased news exposure. *Journalism & Mass Communication Quarterly*, *91*(1), 59–77.

Beckett, C. (2019). New powers, new responsibilities. A global survey of journalism and artificial intelligence. *Polis LSE*. https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities

Belair-Gagnon, V., & Holton, A. E. (2018). Boundary work, interloper media, and analytics in newsrooms: An analysis of the roles of web analytics companies in news production. *Digital Journalism*, *6*(4), 492–508.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Brosius, H. B., Haim, M., & Weimann, G. (2019). Diffusion as a future perspective of agenda setting. *The Agenda-Setting Journal*, *3*(2), 123–138.

Denham, B. E. (2010). Toward conceptual consistency in studies of agenda-building processes: A scholarly review. *The Review of Communication*, *10*(4), 306–323.

Deuze, M. (2005). What is journalism? Professional Identity and Ideology of journalists reconsidered. *Journalism*, *6*(4): 442–464.

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, *3*(3), 398–415.

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.

Diakopoulos, N. (2020). Computational news discovery: Towards design considerations for editorial orientation algorithms in journalism. *Digital Journalism*, *8*(7), 945–967.

Diakopoulos, N., Dong, M., & Bronner, L. (2020, March 20–21). Generating location-based news leads for national politics reporting [Paper presentation]. Computation + Journalism Symposium 2020, Boston, MA, US.

Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, *5*(7), 809–828.

Feezell, J. T. (2018). Agenda setting through social media: The importance of incidental news exposure and social filtering in the digital era. *Political Research Quarterly*, *71*(2), 482–494.

Glaser, B. G., & Strauss, A. L. (2017). *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

Gleason, S. (2010). Harnessing social media: News outlets are assigning staffers to focus on networking. *American Journalism Review*, *32*(1), 6–8.

Golan, G. (2006). Inter-media agenda setting and global news coverage: Assessing the influence of the New York Times on three network television evening news programs. *Journalism Studies*, *7*(2), 323–333.

Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, *7*(8), 993–1012.

Karimi, M., Jannach, D., & Jugovac, M. (2018). News recommender systems—Survey and roads ahead. *Information Processing & Management*, *54*(6), 1203–1227.

Lewis, S. C., Guzman, A. L., & Schmidt, T. R. (2019). Automation, journalism, and human–machine communication: Rethinking roles and relationships of humans and machines in news. *Digital Journalism*, *7*(4), 409-427.

Marconi, F. (2020). *Newsmakers: Artificial intelligence and the future of journalism*. Columbia University Press.

McCombs, M. (2004). *Setting the agenda: The mass media and public opinion*. Polity.

McCombs, M. (2005). A look at agenda-setting: Past, present and future. *Journalism Studies*, *6*(4), 543–557.

McLuhan, M. (1964). *The medium is the message*. Routledge.

Milosavljević, M., & Vobič, I. (2019). Human still in the loop: Editors reconsider the ideals of professional journalism through automation. *Digital Journalism*, *7*(8), 1098–1116.

Molumby, C. (2020, May 20). Modus, semi-automatically creating new story modes. *BBC News Labs*. https://bbcnewslabs.co.uk/projects/modus

Nechushtai, E., & Lewis, S. C. (2019). What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior*, *90*, 298–307.

Newman, N. (2021). *Journalism, media and technology trends and predictions 2018*. The Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2021

Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin Press.

Pavlik, J. (2000). The impact of technology on journalism. *Journalism Studies*, *1*(2), 229–237.

Reich, Z., & Hanitzsch, T. (2013). Determinants of journalists' professional autonomy: Individual and national level factors matter more than organizational ones. *Mass Communication and Society*, *16*(1), 133-156.

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), *Recommender systems handbook* (pp. 1–35). Springer.

Shin, D. (2020). Expanding the role of trust in the experience of algorithmic journalism: User sensemaking of algorithmic heuristics in Korean users. *Journalism Practice*, *1*(2), 1–24.

Shoemaker, P. J., Vos, T. P., & Reese, S. D. (2009). Journalists as gatekeepers. In K. Wahl-Jorgensen & T. Hanitzsch (Eds.), *The handbook of journalism studies* (pp. 73–83). Routledge.

Splendore, S. (2016). Quantitatively oriented forms of journalism and their epistemology. *Sociology Compass*, *10*(5), 343–352.

Tameling, K., & Broersma, M. (2013). De-converging the newsroom: Strategies for newsroom change and their influence on journalism practice. *International Communication Gazette*, *75*(1), 19–34.

Tan, Y., & Weaver, D. H. (2007). Agenda-setting effects among the media, the public, and congress, 1946–2004. *Journalism & Mass Communication Quarterly*, *84*(4), 729–744.

Tandoc, E. C., Jr. (2017). Journalistic autonomy and web analytics. In B. Franklin & S. Eldridge (Eds.), *The Routledge companion to digital journalism studies* (pp. 293–301). Routledge.

Tandoc, E. C., Jr., & Eng, N. (2017). Climate change communication on Facebook, Twitter, Sina Weibo, and other social media platforms. *Oxford Research Encyclopedia of Climate Science*, *63*(6), 1011–1031.

Tandoc, E. C., Jr., & Thomas, R. J. (2015). The ethics of web analytics: Implications of using audience metrics in news construction. *Digital Journalism*, *3*(2), 243–258.

Thurman, N., Moeller, J., Helberger, N., & Trilling, D.

(2019). My friends, editors, algorithms, and I: Examining audience attitudes to news selection. *Digital Journalism*, *7*(4), 447–469.

Urquhart, C. (2012). *Grounded theory for qualitative research: A practical guide*. SAGE.

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

Vu, H. T. (2014). The online audience as gatekeeper: The influence of reader metrics on news editorial selection. *Journalism*, *15*(8), 1094–1110.

Wallsten, K. (2007). Agenda setting and the blogosphere: An analysis of the relationship between mainstream media and political blogs. *Review of Policy Research*, *24*(6), 567–587.

Weimann, G., & Brosius, H. B. (2016). A new agenda for agenda-setting research in the digital era. In G. Vowe & P. Henn (Eds.), *Political communication in the online world: Theoretical approaches and research designs* (pp. 26–44). Routledge.

About the Authors

Hannes Cools is a PhD candidate at the Institute for Media Studies, KU Leuven, Belgium. His research interests include computational journalism and news automation.

Baldwin Van Gorp (PhD) is a full professor of journalism and communication management at KU Leuven, Belgium, where is also the coordinator of the Institute for Media Studies. His research interests include framing, journalism, and the social construction of reality.

Michaël Opgenhaffen (PhD) is an associate professor at the Institute for Media Studies, KU Leuven, Belgium, where he is also the director of the master's program in journalism. He is visiting professor at the University of Leiden in the Netherlands. His research focuses on the production and consumption of social media news and the future of (digital) journalism.

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 4, Pages 208–221
https://doi.org/10.17645/mac.v9i4.4241

Article

One Recommender Fits All? An Exploration of User Satisfaction With Text-Based News Recommender Systems

Mareike Wieland *, Gerret von Nordheim and Katharina Kleinen-von Königslöw

Institute of Journalism/Media Research, University of Hamburg, Germany;
E-Mails: mareike.wieland@uni-hamburg.de (M.W.), gerret.vonnordheim@uni-hamburg.de (G.v.N.), katharina.kleinen@uni-hamburg.de (K.K.-v.K.)

* Corresponding author

Submitted: 1 March 2021 | Accepted: 30 April 2021 | Published: 18 November 2021

Abstract
Journalistic media increasingly address changing user behaviour online by implementing algorithmic recommendations on their pages. While social media extensively rely on user data for personalized recommendations, journalistic media may choose to aim to improve the user experience based on textual features such as thematic similarity. From a societal viewpoint, these recommendations should be as diverse as possible. Users, however, tend to prefer recommendations that enable "serendipity"—the perception of an item as a welcome surprise that strikes just the right balance between more similarly useful but still novel content. By conducting a representative online survey with n = 588 respondents, we investigate how users evaluate algorithmic news recommendations (recommendation satisfaction, as well as perceived novelty and unexpectedness) based on different similarity settings and how individual dispositions (news interest, civic information norm, need for cognitive closure, etc.) may affect these evaluations. The core piece of our survey is a self-programmed recommendation system that accesses a database of vectorized news articles. Respondents search for a personally relevant keyword and select a suitable article, after which another article is recommended automatically, at random, using one of three similarity settings. Our findings show that users prefer recommendations of the most similar articles, which are at the same time perceived as novel, but not necessarily unexpected. However, user evaluations will differ depending on personal characteristics such as formal education, the civic information norm, and the need for cognitive closure.

Keywords
algorithm-based recommenders; diversity; news recommender design; recommender field experiment; reliable surprise

Issue
This article is part of the issue "Algorithmic Systems in the Digital Society" edited by Sanne Kruikemeier (University of Amsterdam, The Netherlands), Sophie Boerman (University of Amsterdam, The Netherlands), and Nadine Bol (Tilburg University, The Netherlands).

1. Introduction

News recommendations are widespread, not only on large social media platforms but also in journalistic media (Kunert & Thurman, 2019). In a fragmented and rich information environment, algorithm-based recommender systems help users find relevant content (Bernstein et al., 2020). As social media and news aggregators nowadays have become a common way of accessing news, news organizations face pressure to offer a similar user experience to meet users' expectations (Nielsen, 2016). Implementing news recommendation algorithms on their web pages and mobile applications has thus become an integral part of their revenue strategies (Bodó, 2019; Kunert & Thurman, 2019). At the same time, not all news companies may be able, or want, to employ the "data-hungry" personalization strategies of the recommendation systems developed by large tech platforms. For them, recommendation algorithms based exclusively or primarily on content characteristics might be the more useful

option for satisfying the expectations of their users, their normative goals, and even their economic aims.

Combining a prototype for a text-based news recommender system with an online survey representative of German internet users (n = 588), this article explores how satisfactory a text-based news recommendation algorithm is perceived by news users, and whether certain user dispositions might impact this satisfaction, and make it necessary to optimize the content-based news recommendation engine for specific user groups. Based on our results, we propose optimizing recommender systems by capturing specific user characteristics in a targeted (and explicit) way.

2. Potential and Challenges of Text-Based News Recommenders for News Companies

Recommender systems can be described as a personalization, that is, as "a form of user-to-system interactivity that uses a set of technological features to adapt the content, delivery, and arrangement of a communication to individual users' explicitly and/or implicitly determined preferences" (Thurman & Schifferes, 2012, p. 776). Based on this definition, we can first distinguish recommendations based on explicitly expressed user preferences from systems that draw on data implicitly (in reality, we often find hybrid forms; Spangher, 2015). These, in turn, fall in a continuum between user-data and content-data dependency. Each of these forms entails specific dilemmas for journalistic media that seek to personalize their content: Firstly, users have little motivation to provide explicit information about their preferences to improve recommendations (Thurman et al., 2019). Also, this information (such as interests in certain topics) quickly become obsolete (Kunert & Thurman, 2019, p. 762). Secondly, implicit recommendations are often "data hungry" (Head of Product at BBC News Online, 2016, cited in Kunert & Thurman, 2019, p. 777), i.e., they rely on the extensive collection and sharing of user data. Adams (2020) has pointed out that the "audience has been commodified and therefore instrumentalized" (p. 883)—a practice that threatens to undermine the authority of journalism as an institution committed to democratic norms and that is increasingly addressed by regulation authorities with restrictive legislation (such as the EU's General Data Protection Regulation; Eskens, 2019), even though its effectiveness in protecting consumers' data is debatable (Reviglio, 2020). Thirdly, the technologies that facilitate article recommendations are often provided by third parties such as the content aggregator, for example, Outbrain (Kunert & Thurman, 2019, p. 777). As a result, journalistic media are becoming increasingly dependent on platforms whose recommendation technologies are not transparent. The media themselves are in turn becoming wary of this practice of collecting user data and sharing it with third-party vendors (Kunert & Thurman, 2019, p. 777; von Nordheim & Fuchsloch, 2019, p. 254).

Given these problems, it seems an obvious and forward-looking choice for media to develop their own technologies that require little explicit participation by users and little disclosure of personal data, which rely instead on features of the news content. Today, rapid developments in the field of Natural Language Processing (or Natural Language Understanding) make it possible to compute text similarities based on complex language models. As an example of such technologies, we study article recommendations based on text similarities, operationalized by the BERT language model (Bidirectional Encoder Representations from Transformers), which was introduced by Google researchers (Devlin et al., 2018). The language model, pre-trained on Wikipedia and book texts, is used to compute document similarities (and other Natural Language Understanding tasks such as sentiment classification, natural language inference, and question answering) and achieved state-of-the-art accuracy (Wang et al., 2019). BERT is therefore an obvious algorithm for the development of new content-based recommender systems (Wang & Fu, 2020). Thanks to developments such as BERT, even small publishers (or service providers beyond the big advertising platforms, see Section 4.2.) can now create their own software for news recommendations. Just a few years ago, this level of independence involved huge development costs and was therefore only available to big media—Kunert and Thurman (2019) mention the *Financial Times* (p. 775), other examples include the *New York Times* (Spangher, 2015), or the *Washington Post* (Graff, 2015).

Journalistic media that seek to implement this form of text-based recommendation face the challenge that solely optimizing for text similarity satisfies neither the user's appetite for "news" nor the news media's normative aim to present their users' with a certain level of diversity. Such a similarity-based recommender thus needs to be calibrated against user satisfaction, e.g., a positive assessment of relevance and quality of the recommended article in accordance to personal needs (Bodó et al., 2019). User satisfaction, in turn, is assumed to translate into loyalty and trust, thus increasing the value of a news brand in the long term (Nelson & Kim, 2020).

3. User Satisfaction at the Intersection of Pleasurable Comfort and Valuable Diversity

On the one hand, similarity-based recommendations are likely to be evaluated positively, because users are familiar with the recommended topics, views, or facts. As the mere exposure effect (Zajonc, 1968) suggests, people tend to evaluate objects or people better only because they are more familiar with them (Bornstein, 1989). A similarity-based recommendation could thus encourage positive evaluations of the recommended article through repetition of topics, views, or facts and hence ease of processing. This should particularly hold true if

the original article that the similarity-based recommendation is based on is perceived of as being high quality.

On the other hand, it is obvious from the user's point of view that presenting more of the same makes "a perfectly boring, very foreseeable, very cold and technology-driven product, that doesn't feel like a proper journalistic product" (Bodó, 2019, p. 1068). Indeed, journalistic media are particularly regarded for their skill at providing users with a "reliable surprise" based on a wide range of high-quality content (Schoenbach, 2007), in other words, to let them encounter content that is pleasantly unexpected and new but without seeming accidental and arbitrary. Applied to recommendation engines, users might thus aim for serendipity in news recommendations.

"Serendipity" is a common design goal of recommendation engines not only in the context of news (Reviglio, 2019) and defined as the sweet spot, just the right degree, between novelty and unexpectedness (Maccatrozzo et al., 2017). This balance ensures that, although unexpected and new, the recommendation is still perceived as pleasant, enriching, and thus useful, which reflects in user satisfaction (Chen et al., 2019).

In contrast, optimizing recommendations solely in the direction of maximum content diversity is presumably not rated positively by users and could thus even work against the economic interests of publishers (Bernstein et al., 2020). However, a one-sided optimization in the direction of pleasure and convenience through very similar recommendations may in turn quickly lead to a limited range of content and counteract the societal role of journalistic media. Furthermore, narrowing down the selection of articles by only presenting the news a user likes is considered "the wrong path" (Bodó, 2019, p. 1065): "Our goal as a news organization is to inform people about what is happening and there are things that are not always fun" (Bodó, 2019, p. 1065). Thus, aiming for serendipity based on novelty as well as on unexpectedness might even translate into a more diverse news menu by challenging users' viewpoints from time to time.

Our first research question thus explores the relationship between the recommendation based on text similarity and user satisfaction with the recommendation (while controlling for the quality of the original article): How are text similarity, article quality, and overall recommendation satisfaction related (RQ1)? And how are the evaluations of the recommended article as (a) novel, and (b) unexpected and the overall level of recommendation satisfaction related (RQ2)?

The academic discussion of diversity in news is mostly limited to the supply side (Helberger et al., 2015). Still, there are indications that users have different expectations regarding the diversity of a news offering (Nielsen, 2016) and that they prefer different degrees of diversity (Bodó et al., 2019; Helberger et al., 2018). To perceive serendipity as enrichment, users need a certain "mental readiness" (Lutz et al., 2017, p. 1706) to encounter openly new, unexpected information that is recommended to them. Individuals with a chronic need for cognitive closure (NfcC) generally prefer unambiguous situations and find ambiguity unpleasant (Webster & Kruglanski, 1997). Accordingly, they might benefit more from a recommended news item that is very similar to a previous, already known one.

A preference for algorithmic personalization (Thurman et al., 2019) and the share of algorithmically personalized news on overall news use (Schweiger, et al., 2019), in contrast, could increase satisfaction with article recommendations as both might reflect a higher acceptance of automated news recommendations. Similarly, a high technological affinity (Hampel et al., 2020) might lead to a more playful approach towards interactive online systems, again resulting in a greater mental readiness to encounter recommended news articles (McCay-Peet, 2013).

Differences in civic norms, such as in the duty to keep informed, general news interest, or trust, could influence satisfaction with article recommendations because they express an individually different motivation to engage with the recommended articles. This could lead to a very similar article being perceived as a welcome deepening of the topic. But it could also mean that the recommendation needs to be better, i.e., more tailored to already relatively specific needs and clear expectations of well-informed users who are strongly committed to being informed. Since serendipity does not contribute to satisfaction in the case of a highly purposeful use (Lutz et al., 2017), satisfaction with article recommendations solely based on text similarity might reach its limits here. Finally, sociodemographic characteristics such as age or education might also be influential (Möller et al., 2018).

Conceptualizing serendipity as the central variable of user satisfaction thus incorporates a "liberal-individualistic idea" of diversity (Helberger et al., 2018, p. 195), but it also takes into account the deliberative aspect of being exposed to a variety of different topics, facts, and points of view. Therefore, we assume that, as with diversity expectation, user satisfaction is not a "universal user trait" (Bodó, 2019, p. 208).

We, therefore, ask (RQ3): Which individual dispositions (preference for news personalization, relative share of algorithmically personalized news, technological affinity, NfcC, duty to keep informed, news interest, and trust) influence the relationship between text similarity, article quality, and recommendation satisfaction (RQ3a)? And what is the moderating role of individual dispositions regarding the relationship between recommendation satisfaction and evaluations of the recommendations as novel or unexpected (RQ3b)?

4. Research Design

In the reality of news companies, it is challenging to measure satisfaction with the recommendation beyond the actual click (Bodó, 2019) as additional user surveys are required. In communication research, it is in turn dif-

ficult to simulate realistic article recommendation and study recommendations as to the interaction between algorithms and the user (Loecherbach & Trilling, 2020). Data on user views of real recommendations are accordingly scarce in academia: most are derived from hypothetical instructions.

A unique feature of this study is the integration of a real recommendation engine for news articles into a survey. Even though the advantage of combining web tracking and survey data is clear (Bernstein et al., 2020; Loecherbach & Trilling, 2020), it is mainly the big tech platforms that have taken advantage of it so far (Stray, 2020). It is important to note that the design of the study is exploratory, it is a pilot study. Even though participants were randomly assigned to three different groups for their news recommendations (most similar article, least similar article, or article of random similarity), this is not a classic experimental study. Our analytical strategy aims at exploring and identifying relevant relationships between the different variables as a basis for further study, not at confirming hypotheses. For this reason, we have also retained the similarity score as a metric variable (and not a categorical variable identifying experimental groups).

4.1. Questionnaire Structure

Starting with questions about news usage, an interactive part follows in which the respondents freely search a database of actual news articles. Participants enter a search query that is of interest to them and related to politics, business, or culture in Germany and the world (hereafter depicted as "news"). The search query can consist of any number of terms. We used a search query as a starting point for the news browsing situation rather than a mock-up news webpage with a restricted set of articles as the latter may force participants to select articles on topics they are not interested in. By allowing users to freely select a topic of their choosing, participants are more likely to have a similar baseline of interest in the article on which the recommendation is then based. However, because users have to consciously decide on and type a search query, this overall level of interest in the article presented by the search query is likely to be somewhat higher than in a normal news browsing situation.

Respondents then select one of the multiple search results for further reading. To keep the time requirement reasonable, only articles with a minimum word count of 172 and a maximum of 736 are available. Immediately after reading, participants evaluate the quality of the self-selected article. Afterwards, another article is automatically recommended for further reading, randomly using one of three levels of text similarity (see Section 2). We instructed the participants: "The next click will take you to an article that might be of interest to you as well. This article is recommended to you based on the first article. Please read the article, just as you normally would do." After reading, they again rate the recommended article. In addition, participants indicate their satisfaction with this recommendation. In order to avoid influencing the participants' response behaviour by preceding questions, for example, about their attitude towards personalization, these personal dispositions are surveyed after the interactive part.

4.2. Recommendation Engine

The recommendation engine was developed by the German start-up LakeTech, with whom we cooperated in this study. The start-up offers publishers the opportunity to integrate proprietary recommendation systems into their websites. The article recommendations are based on the content of the previously selected texts aiming to present similar texts. Similarity is calculated based on vector representations of each article—which are in turn based on the average of the vectors of each sentence (sentence embeddings), calculated with the pre-trained language model BERT developed by Google (Devlin et al., 2018). Thus, quantification of a statistical similarity between all articles is possible and represented as a similarity score (values between 0, *no similarity*, and 1, *identical*). For this study described here, three recommendation logics were implemented: (a) the most similar item is recommended; (b) the least similar; (c) a randomly drawn item. The three recommendation logics were randomly assigned to the participants (see Section 4.3).

To generate a representative news corpus, the URLs of relevant texts from ten different media (see Supplementary Files) were first saved via News API (2021) and scraped in the next step. These 194,167 German news texts from the year 2020 (published between 31 January 2020 and 1 January 2021) were then vectorized.

4.3. Measures

The main dependent variable is "satisfaction with the article recommendation," measured as agreement (5-point Likert scale) on items stating that: (a) the topic; (b) viewpoints; and (c) facts of the article are perceived as pleasant and enriching (e.g., "The second article was recommended to you based on the first article. We are interested in your evaluation of the second article compared to the first. Compared to the first article, I perceived the topic [viewpoints, facts] of the second article as pleasant and enriching"). Agreement on these three items is aggregated into a mean index (Cronbach's α = .89).

As possible independent variables related to the recommendation, we looked at text similarity, article quality, and evaluation of the recommendation. "Similarity score" is calculated using the vectorized articles (values between 0, *minimum*, and 1, *maximum similarity*). For each article in the corpus, the IDs and similarity scores of three other articles were stored as meta-data (the most similar, the least similar, and a randomly drawn article)

as a basis for the random assignment of recommended articles (as described in Section 4.2).

We operationalized the "evaluation of the recommendation" using the two dimensions of serendipity (see Section 3): novelty and unexpectedness. In analogy to the satisfaction measurement, we surveyed perception of how new and how unexpected the topics, viewpoints, and facts in the recommended article were (Haim et al., 2018) in comparison to the first article. Again, we aggregated agreement on these items into mean indices for "novelty" (Cronbach's α = .59) respectively "unexpectedness" (Cronbach's α = .73).

As we assume that recommendation satisfaction will be higher if the original article is rated as being of good quality, we also control for perceived "article quality." It is rated by the respondents, applying journalistic quality criteria previously used by Jungnickel (2011) using pairs of opposites (e.g., balanced, illustrative, comprehensible, trustworthy) on a 7-step scale and aggregated into a mean index (Cronbach's α = .89).

As possible independent variables relating to the individual dispositions of the users, we included the following: "Attitude towards news personalization" is measured with items applied by Bodó et al. (2019), Thurman et al. (2019), and Schweiger et al. (2019), in some cases with slight adjustments. The items relate both to perceived usefulness of news personalization, e.g., "when a news website highlights content that is particularly important to me," and to concerns about possibly (un)balanced ("I worry that personalized news will cause me to miss articles that contradict my views") or incomplete information ("I worry/fear that personalized news will cause me to miss important information") and privacy ("I worry that personalized news will make my privacy more vulnerable"). The mean index calculated from these six items shows good reliability (Cronbach's α = .75).

The "relative share of algorithmically personalized news" was calculated following the measurement of news usage proposed by Schweiger et al. (2019), it can take values from 0 (*no algorithmically personalized news used*) to 1 (*all used news are algorithmically personalized*). For technology affinity, we used three statements from the annual survey on technology attitudes among the German population (Hampel et al., 2020) aggregated into a sufficiently reliable mean index (Cronbach's α = .65).

"Duty to keep informed" was measured using four items proposed by McCombs and Poindexter (1983; α = .65). For "need for cognitive closure," we shortened the scale proposed by Schlink and Walther (2007) to five items as did Schweiger et al. (2019), but we replaced two of the items to provide a more specific reference to diversity (α = .62). All items were measured using five-point Likert scales and aggregated into mean indices. "News interest and news trust" are each single item measurements as used by Thurman et al. (2019). The full questionnaire is available in the Supplementary Files.

4.4. Sample

Findings are based on a sample representing all German-speaking internet users aged 18 and over. Participants were recruited by the online access panel provider Norstat, and cross-sampled according to education and age, as well as by place of residence (federal state), and gender. At the end of the ten-day field period in January 2021, 1,027 finished questionnaires resulted. After correcting for respondents who did not meet our pre-defined quality criteria (non-plausible answers, unrealistic response times, straightlining, incomplete cases), 588 valid cases remained for further analyses.

On average, participants are 48.2 years old. 51.5% identify with the male gender. 53.9% have a low level of formal education (no degree, secondary school diploma), 46.1 are higher educated (A-Levels, bachelor, master, doctorate). Accordingly, the sample offers a good representation of the German online population.

5. Results

To explore the relationships between the different variables as outlined in our research questions, we chose a regression analysis approach with recommendation satisfaction as the dependent variable. The predictors are included block-wise (forced entry), starting with sociodemographic variables (Model 1), evaluations of the recommendation and the original article (Model 2), other individual characteristics (Model 3), and selected interaction effects.

Regression assumptions were tested using plots (Luhmann, 2015). These plots (see the Supplementary Files) are used to verify the correct model specification (nonsystematic distribution, Lowess line parallel to the x-axis in the residuals vs. fitted diagram), to check the normal distribution of the residuals (in the Q-Q plot comparison with the diagonal), the homoscedasticity assumption (in the scale-location diagram unsystematic distribution of the residuals), and to diagnose outliers and influential values (with the residuals vs. leverage diagram using Cook's distance).

5.1. How Are Text Similarity, Article Quality, and Satisfaction With the Recommended Article Related?

Among the sociodemographic variables (Table 1, Model 1), only gender is initially influential. Those who identify themselves as male are a little more satisfied with the article recommendation ($b = .17$, $p < .05$). However, education and age do not correlate substantially with recommendation satisfaction. In total, sociodemographic characteristics explain only 1.6% of the variance ($F(3,584) = 3.22$, $p = .02$), opening up great explanatory potential for other predictors.

For this reason, all variables evaluating the recommendation or the article were included as the model's second block (Table 1, Model 2), increasing the explained

Table 1. Hierarchical regression predicting recommendation satisfaction.

	Model 1				Model 2				Model 3				Model 4			
	b	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]
Independent variables																
Intercept	2.71**	[2.57, 2.84]			2.40**	[2.09, 2.72]			2.61**	[2.28, 2.95]			2.64**	[2.31, 2.97]		
Block 1: Demographic variables																
Gender (1 = male)	.17*	[.01, .34]	.01	[–.01, .02]	.19**	[.05, .32]	.01	[–.00, .02]	.14*	[.01, .28]	.00	[–.00, .01]	.14*	[.00, .27]	.00	[–.00, .01]
Age	–.01*	[–.1, –.00]	.01	[–.01, .02]	–.00	[–.01, .00]	.00	[–.00, .01]	–.00	[–.01, .00]	.00	[–.00, .01]	–.00	[–.01, .00]	.00	[–.00, .01]
Education (1 = high)	–.14	[–.31, .02]	.00	[–.01, .02]	–.23**	[–.36, –.09]	.01	[–.00, .03]	–.24**	[–.38, –.11]	.01	[–.00, .03]	–.25**	[–.39, –.11]	.01	[–.00, .03]
Block 2: Recommendation measures																
Similarity score					.70**	[.28, 1.11]	.01	[–.00, .03]	.72**	[.31, 1.13]	.01	[–.00, .03]	.70**	[.29, 1.10]	.01	[–.00, .03]
Quality article 1					.07*	[.01, .13]	.01	[–.00, .02]	.03	[–.03, .09]	.00	[–.00, .01]	.03	[–.03, .09]	.00	[–.00, .00]
Novelty of recommendation					.67**	[.58, .75]	.28	[.22, .34]	.63**	[.55, .71]	.24	[.18, .29]	.63**	[.54, .71]	.24	[.18, .29]
Unexpectedness of recommendation					–.22**	[–.30, –.14]	.03	[.01, .06]	–.20**	[–.28, –.13]	.03	[.01, .05]	–.20**	[–.27, –.12]	.03	[.01, .05]

Table 1. (Cont.) Hierarchical regression predicting recommendation satisfaction.

	Model 1				Model 2				Model 3				Model 4			
	b	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]
Block 3: Individual dispositions																
Attitude towards personalization									.20**	[.11, .29]	.02	[.00, .04]	.20**	[.11, .29]	.02	[.00, .04]
Relative share of algorithmically personalized news									−.19	[−.49, .11]	.00	[−.00, .01]	−.19	[−.49, .10]	.00	[−.00, .01]
Duty to keep informed (DTKI)									−.06	[−.21, .08]	.00	[−.00, .00]	−.08	[−.23, .06]	.00	[−.00, .01]
Need for Cognitive Closure									.12*	[.02, .22]	.01	[−.00, .02]	.12*	[.02, .22]	.01	[−.00, .01]
News interest									.14*	[.02, .26]	.01	[−.00, .02]	.13*	[.01, .25]	.00	[−.00, .01]
News trust									.01	[−.07, .10]	.00	[−.00, .00]	.02	[−.06, .10]	.00	[−.00, .00]
Technical affinity									−.03	[−.11, .06]	.00	[−.00, .00]	−.02	[−.11, .06]	.00	[−.00, .00]
Block 4: Moderating effects																
Unexpectedness of recommendation * DTKI													−.18**	[−.30, −.06]	.01	[−.00, .02]
R2		R2 = .016*				R2 = .351**				R2 = .387**				R2 = .396**		
95% CI		[.00, .04]				[.28, .40]				[.31, .43]				[.32, .44]		
ΔR2						ΔR2 = .335**				ΔR2 = .036**				ΔR2 = .009**		
95% CI						[.27, .40]				[.01, .06]				[−.00, .02]		

Notes: A significant b-weight indicates the semi-partial correlation is also significant; *b* represents unstandardized regression weights; sr^2 represents the semi-partial correlation squared; LL and UL indicate the lower and upper limits of a confidence interval, respectively; * $p < .05$; ** $p < .01$.

variance rises by 33.5 percentage points up to 35.1% for Model 2 (ΔR2 = .335, p < .01; F(7,580) = 44.79, p = .00).

Among the predictors, text similarity has the largest impact (b = .7, p < .01) on recommendation satisfaction, with higher text similarity leading to better ratings. However, since the recommendation is based on similarity, this relationship should prove especially true if the first article is already rated as high quality. Although the evaluation of the first article itself only weakly contributes to recommendation satisfaction (b = .07, p < .05), Figure 1 might indicate a possible moderating effect. For those who already rate the first article better (mean + 1SD; all variables are mean-centred), the visualization suggests a stronger positive correlation between the text similarity used for recommending an article and recommendation satisfaction. A moderation analysis was run to determine whether the interaction between the evaluation of the first article and text similarity significantly predicts recommendation satisfaction. For this, the interacting variables were centred at their mean (using gscale from the jtools package; Long, 2021), then the linear model was fitted and plotted using the interactions package (Long, 2020). Since this interaction (b = .18, p = .4) does not become statistically significant (ΔR2 = .39, F(15, 572) = 24,14, p < .01), the interaction term was not added to the final model as suggested by Hayes and Little (2018, p. 236).

5.2. How are the Evaluations of Article Recommendation and the Level of Recommendation Satisfaction Related?

Recommendation satisfaction is positively related to evaluating the recommended article as novel: If the recommended article presents new topics, perspectives, and/or facts, then the recommendation is perceived as more pleasant and enriching (b = .67, p < .01). By contrast, unexpected topics, viewpoints, and/or facts lead to the recommendation being experienced somewhat less as pleasant and enriching (b = −.22, p < .01)—even when, as in this model, all other factors such as text similarity and perceived quality of the first article are held constant. Apparently, the novelty and the unexpectedness of topics, facts, and/or viewpoints in a recommended article have a different impact on readers' recommendation satisfaction (see also Section 5.3, where we explore this relationship further).

By including the predictors related to article recommendation, formal education now also becomes significant. Users with a higher level of formal education are apparently less satisfied with the article recommendation (b = −.23, p < .01) holding all other predictors constant, which may be explained by their having higher or more specific content expectations (Helberger et al., 2018), and this will be further explored in the following sections.

5.3. Do Individual Dispositions Influence the Relationship Between Text Similarity, Article Quality, and Recommendation Satisfaction?

In a further analytical step, we consider individual dispositions as possible predictors (Table 1, Model 3), complementing the article and recommendation related measures and socio-demographics explored above. Contrary to our expectation, including individual dispositions improves model fit only minimally but still significantly (ΔR2 = .036, p < .01) leading to Model 3

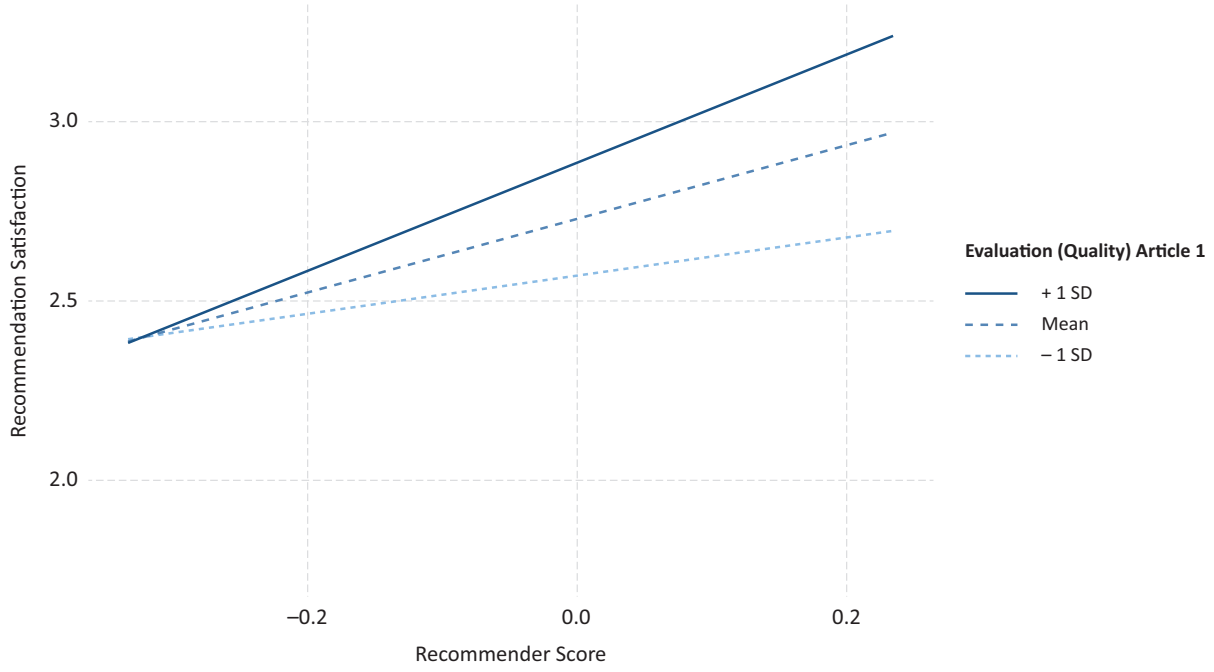


Figure 1. Moderating effect (n.s.) of quality article 1 on the conditional effect of text similarity on recommendation satisfaction.

(F(14,573) = 25.83, p = .00). Moreover, the inclusion of the individual variables does not lead to any significant change in the results already found.

Interest in news is positively but only moderately related to recommendation satisfaction (b = .14, p < .05). News trust (b = .01, n.s.) and a perceived duty to keep informed (b = −.06, n.s.), on the other hand, as well as a general affinity for technology (b = −.03, n.s.), show no correlation with recommendation satisfaction.

By contrast, the general attitude towards news personalization makes a more pronounced difference regarding the level of recommendation satisfaction. Those who report preferring news recommendations, as in the present case, also rate the recommendation as more pleasant and enriching (b = .2, p < .01), controlling for all other predictors. A further graphical exploration (Figure 2) reveals that for these news personalization endorsers, their positive evaluations of recommended articles are almost independent of the similarity to the original article. The (self-reported) personalization sceptics (mean −1SD), however, appear to prefer more similar article recommendations.

An additional moderation analysis did not show that opinion on personalization moderates the effect between text similarity and recommendation satisfaction was statistically significant (b = .27, p = .16; ΔR2 = .39, F(15, 572) = 24,28, p < .01). Again, the interaction term was dropped from the model, resulting in the simple effects only model (which is identical to Model 3). Similarly, the relative share of algorithmic news media in total news consumption becomes ineffective for recommendation satisfaction.

The significant, albeit weak, correlation of recommendation satisfaction with a NfcC (b = .12, p < .05) is surprising at first but plausible given the fact that the recommendation here is based on text similarity. People who avoid ambiguities and prefer closed-world views are more likely to perceive an article that matches their first, self-selected article, and thus the recommendation itself, as pleasant and enriching than those who enjoy challenging perspectives. Here, there might be a connection to the above finding that a differentiation of evaluation dimensions is apparently needed when measuring recommendation satisfaction, as shown by the opposing influence of articles that are evaluated as new compared to articles that are evaluated as unexpected.

The analysis of RQ3 shows that rating the recommended item as novel versus unexpected has an opposite effect on recommendation satisfaction: Recommendations of articles rated as novel are evaluated more positively, while at the same time recommendations of articles rated as unexpected are evaluated negatively. We explore the interrelations of these three measures further by analysing how individual dispositions might interact with rating the recommended article as unexpected and as novel, respectively.

Including the corresponding interaction terms again only leads to minimal further variance explanation (Model 4; ΔR2 = .009, p < .01; F(15,572) = 25.00, p = .00). In the direct comparison of the novel vs. unexpected dimension, it is noticeable that none of the interactions with novelty contributes significantly to the model, also underlined by a visual exploration (see the Supplementary Files). Again, the moderation analysis did not find any of the assumed moderating effects.

However, for the correlation between the recommendation satisfaction and its unexpectedness, the impact of different individual dispositions can be visually detected,

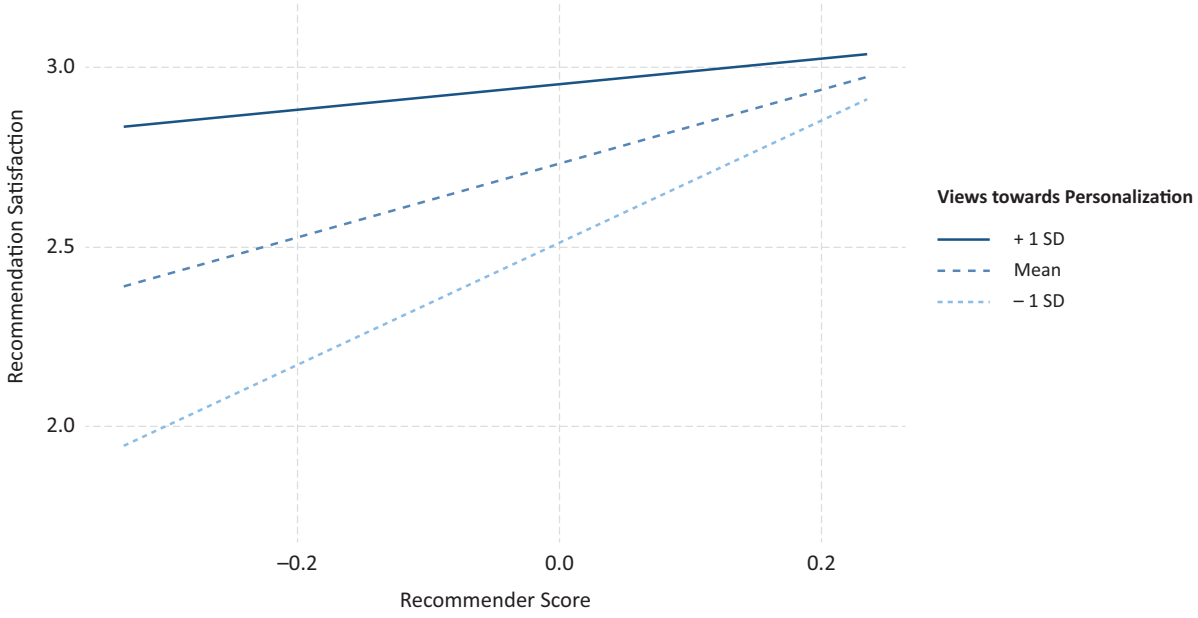


Figure 2. Moderating effect (n.s.) of view towards personalization on the conditional effect of text similarity on recommendation satisfaction.

each showing different strengths in their correlations. Accordingly, a moderation analysis was conducted to determine whether the interaction between the personal characteristics and the rating of the recommended article as unexpected significantly predicts satisfaction. Results show that duty to keep informed (DTKI) moderated the effect between unexpectedness and recommendation satisfaction significantly, $F(15,572) = 25.00$, $p < .001$.

To investigate this effect in more detail, a Johnson-Neyman diagram was plotted (Figure 3). For average (i.e., around the mean) values of DTKI, there is no significant moderation effect. For below-average values of the DTKI (from about .8SD below the mean), on the other hand, we see a positive effect. Thus, for people with a low DTKI, recommendations perceived as pleasant and enriching are also more likely to be perceived as unexpected. For people with a greater sense of duty to inform themselves, on the other hand, there is a negative effect, i.e., here the recommendations not perceived as pleasant and enriching are more likely to be perceived as "not unexpected," in other words, as expected and unsurprising.

This refers to the challenge of finding the right balance between being novel and unexpected, while

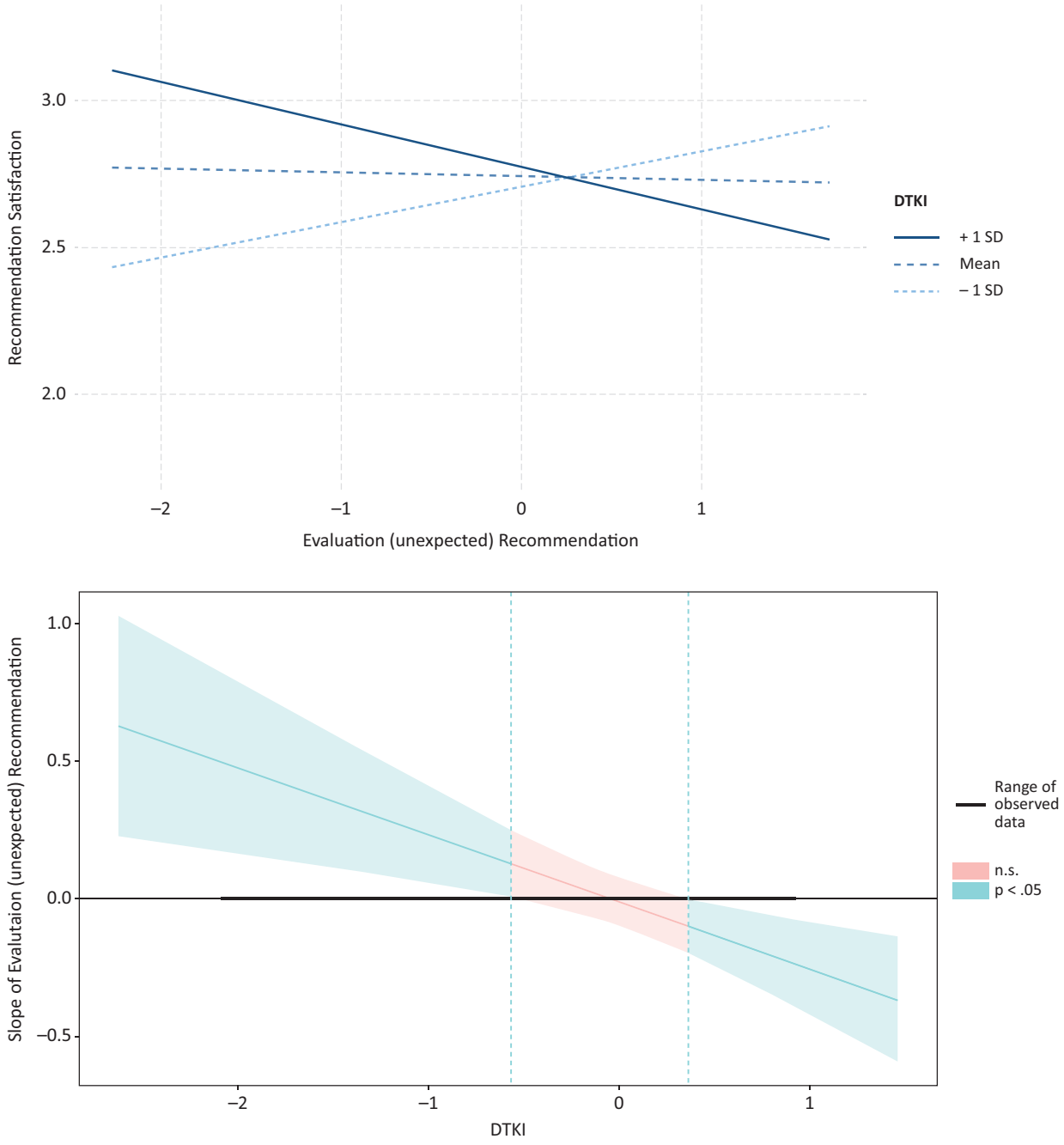


Figure 3. Moderating effect of duty to keep informed on the conditional effect of evaluation as unexpected on recommendation satisfaction.

still being perceived as pleasant and enriching in the overall experience and thus contributing to satisfaction. Optimizing toward pleasant and enriching through novelty thereby represents the simpler strategy of satisfying users through more familiar ways of recommendation. Optimizing satisfaction through unexpected items is much more challenging and obviously also depends on how strong the civic norm to keep informed is. Nevertheless, from a normative point of view, this strategy has more potential to create diversity in article recommendations.

6. Discussion

In summary, the calibration of the recommender, i.e., the degree of similarity between the original and the recommended article, turns out to be the strongest predictor of the entire model. The stronger the recommendation is based on article similarity, the more pleasant and enriching it is perceived to be (RQ1). Moreover, this satisfaction strongly depends on whether the recommended article is evaluated as novel (RQ2a). By contrast, if the topic, facts, and/or viewpoints of the recommended article are perceived as unexpected, this decreases satisfaction with the recommendation (RQ2b). This would confirm previous research on the preference of news users for "reliable surprises" (Schoenbach, 2007), news media should aim to recommend—and produce—news content that adds novel positions or facts, but still falls within the expectations of users for the topic.

In all models, the more educated rate article recommendations as less pleasant and enriching (RQ3a). Yet, higher news interest leads to slightly higher recommendation satisfaction. Users with a high NfcC are also more likely to be satisfied, which stands to reason given that the recommendation is based on text similarity. For general attitudes toward news personalization, a significant simple effect emerges—endorsing algorithmic personalization leads to greater recommendation satisfaction. Even if the moderation analysis is not significant, it can still be deduced based on the visual analysis that, in particular, people with a greater scepticism towards such news recommendations are satisfied precisely with a more similar recommendation, thus preferring less diverse recommendations (RQ3b). This finding deserves further exploration in the future.

For the fascinating opposing relationship between perceiving the recommended article as novel or unexpected and recommendation satisfaction, a significant moderating effect of the civic information norm emerges. For people with a strong sense of duty to inform themselves about current events, rating the recommended article as unexpected nevertheless goes along with recommendation satisfaction, i.e., perceiving the article recommendation as pleasant and enriching. For those users, news media should aim to recommend even more diverse news and thus fulfil their democratic role in the best possible way. By contrast, people who are less concerned about informing themselves about current events are also less satisfied with article recommendations on unexpected topics, facts, or points of view. Here, news media should aim to recommend a comparatively homogeneous news diet to avoid alienating them during these current times when more and more citizens are losing the connection that news media provides them to the public sphere.

7. Conclusion

So, do our results indicate that there is a possible calibration of the recommender that satisfies all user groups equally? This seems not to be the case. It seems certain that news organizations apparently cannot go far wrong with a text-based recommendation algorithm, as indicated by the strength of the predictor text similarity on recommendation satisfaction. At the same time, we find several individual dispositions for which this relationship is less strong. For example, higher educated people (a key target group of many news organizations) are generally less satisfied with the article recommendation if it is based on text similarity. And if we avoid only focusing on the "liberal-individualist" goal of satisfaction (Helberger et al., 2018), and also take into account unexpected and thus potentially challenging content (which is important from a deliberative point of view), a trade-off becomes apparent. The moderator effect of the duty to keep informed make clear that a single, standardized recommender solution will be difficult to achieve. This is especially true if normative goals beyond user satisfaction are to be met.

This leaves media organizations with the data dilemma in that content-based algorithms alone can hardly meet the individual requirements of different target groups. A mixed strategy of implicit and content-based recommendations could remedy this, as Loecherbach and Trilling (2020) also argue. However, the fact that we have already been able to identify different user segments with a standardized survey should also encourage media organizations to meet user expectations of personalized recommendations sufficiently well with comparatively simple means. For example, on-site surveys segmenting one's own target group on the basis of social science concepts such as the duty to keep informed, personalization preference, or the NfcC used here could already be informative enough to increase recommendation satisfaction in the future and thus contribute to greater trust and customer loyalty.

This brings us to the limitations of our study. To achieve a similar level of interest in the recommended articles for all participants, we asked participants to formulate an active search query in the first step. Only in the subsequent step did they receive an article by automated recommendation. However, the instruction to enter a real information need may have had an effect on the participants' expectations regarding the second, recommended article. Within the context of a goal-oriented

search, users tend to hold a fairly specific set of expectations regarding the characteristics of the article (Lutz et al., 2017). This clear set of expectations may have carried over to the second article. Purposeful searches are also possible on news websites, but open browsing to pass time, or at least quite undirected behaviours in which people just update themselves with current events, are even more prevalent. We, therefore, assume that our design influenced the findings especially in terms of the negative correlation between unexpected topics, viewpoints, and facts and article recommendation satisfaction. In a further study, our exploratory findings need to be investigated in a confirmatory design and under systematic variation of the instructions (search task vs. browsing).

Furthermore, though individual dispositions such as the NfcC or news interest are considered to be relatively stable (and our items aimed to identify the more general, not situational attitudes), there is the possibility that the topics selected by our participants had an impact on these attitudes. A more "emotional" topic such as the Covid-19 pandemic might have temporarily increased the NfcC, whereas a "safer" or very familiar topic may have decreased it. Future studies might consider including more items on the user interest in the selected topic and the level of attention while reading it, to control for these possible priming effects.

Even if the inclusion of individual dispositions already means a shift away from short-term engagement metrics, our study was only able to provide a snapshot of the interrelations between user characteristics and article recommendation evaluation. Especially against the background of building brand loyalty through satisfaction (Nelson & Kim, 2020), the mid- and long-term development of these interrelations need greater attention in the future as does the question of which other individual dispositions might be relevant. Informational self-efficacy, for example, contributes to a higher mental readiness to value serendipity (Lutz et al., 2017). Serendipity could be a valuable link between the sometimes challenging diversity in news and a pleasant user experience. Self-efficacy could also increase the sense of control and agency in news use which might in the long term contribute to the willingness to actively engage with the personalization settings and thus to implicitly or explicitly provide personal data (Monzer et al., 2020).

Despite these limitations, the major advantage of our study lies in the practical relevance and transferability of the recommendation algorithm used. With BERT, we not only simulated a realistic content-based personalization based on genuine articles (an approach that is used cost-effectively by smaller news organizations) but also embedded it in a survey interface that enables authentic recommendations tailored to user interests. Here, typical news portal features such as headlines and images were omitted, as was the media brand, which has undoubtedly reduced the ecological validity of our research design. However, this allowed us to avoid confounding in our exploratory setting. Further studies should aim to gradually include these parameters as well. The simplicity of the design, however, made it possible to achieve a sample that is representative of the German internet population.

Acknowledgments

The authors would like to thank the three anonymous reviewers for their valuable comments and suggestions.

Conflict of Interests

The authors declare no conflict of interests.

Supplementary Material

All raw data and scripts necessary to reproduce the results can be found on https://osf.io/kh5va/?view_only=aff4561db93d434cbb14d7b2b3f2de0c. All attachments are available in this repository.

References

Adams, P. C. (2020). Agreeing to surveillance: Digital news privacy policies. *Journalism & Mass Communication Quarterly*, *97*(4), 868–889. https://doi.org/10.1177/1077699020934197

Bernstein, A., Vreese, C. d., Helberger, N., Schulz, W., Zweig, K., Baden, C., Beam, M. A., Hauer, M. P., Heitz, L., Jürgens, P., Katzenbach, C., Kille, B., Klimkiewicz, B., Loosen, W., Moeller, J., Radanovic, G., Shani, G., Tintarev, N., Tolmeijer, S., . . . Zueger, T. (2020). *Diversity in news recommendations*. ArXiv. http://arxiv.org/pdf/2005.09495v1

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

Bodó, B., Helberger, N., Eskens, S., & Möller, J. (2019). Interested in diversity. *Digital Journalism*, *7*(2), 206–229. https://doi.org/10.1080/21670811.2018.1521292

Bornstein, R. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, *106*, 265–289.

Chen, L., Yang, Y., Wang, N., Yang, K., & Yuan, Q. (2019). How serendipity improves user satisfaction with recommendations? A large-scale user evaluation. In L. Liu & R. White (Eds.), *The world wide web conference on—WWW '19* (pp. 240–250). ACM. https://doi.org/10.1145/3308558.3313469

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of deep bidirectional transformers for language understanding*. ArXiv. http://arxiv.org/pdf/1810.04805v2

Eskens, S. (2019). A right to reset your user pro-

file and more: GDPR-rights for personalized news consumers. *International Data Privacy Law*, *9*(3), 153–172. https://doi.org/10.1093/idpl/ipz007

Graff, R. (2015, June 3). How the *Washington Post* used data and natural language processing to get people to read more news. *Knight Lab*. https://knightlab.northwestern.edu/2015/06/03/how-the-washington-posts-clavis-tool-helps-to-make-news-personal

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

Hampel, J., Zwick, M., & Störk-Biber, C. (2020). *Technik Radar 2020: Was die Deutschen über Technik denken* [Technology radar 2020: What Germans think about technology]. Acatech; Koerber-Stiftung. https://www.koerber-stiftung.de/fileadmin/user_upload/koerber-stiftung/redaktion/technikradar/pdf/2020/TechnikRadar-2020_Langfassung.pdf

Hayes, A. F., & Little, T. D. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Methodology in the social sciences*. The Guilford Press.

Helberger, N., Karppinen, K., & D'Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, *21*(2), 191–207. https://doi.org/10.1080/1369118X.2016.1271900

Helberger, N., Kleinen-von Königslöw, K., & van der Noll, R. (2015). Regulating the new information intermediaries as gatekeepers of information diversity. *Info*, *17*(6), 50–71. https://doi.org/10.1108/info-05-2015-0034

Jungnickel, K. (2011). Nachrichtenqualität aus Nutzersicht. Ein Vergleich zwischen Leserurteilen und wissenschaftlich-normativen Qualitätsansprüchen [News quality from the perspective of the users. A comparison between user evaluations and scientific-normative quality requirements]. *M&K Medien & Kommunikationswissenschaft*, *59*(3), 360–378. https://doi.org/10.5771/1615-634x-2011-3-360

Kunert, J., & Thurman, N. (2019). The form of content personalisation at mainstream, transatlantic news outlets: 2010–2016. *Journalism Practice*, *13*(7), 759–780. https://doi.org/10.1080/17512786.2019.1567271

Loecherbach, F., & Trilling, D. (2020). 3bij3: Developing a framework for researching recommender systems and their effects. *Computational Communication Research*, *2*(1), 53–79. https://doi.org/10.5117/CCR2020.1.003.LOEC

Long, J. A. (2020). Interactions: Comprehensive, user-friendly toolkit for probing interactions (1.1.3) [Computer software]. https://cran.r-project.org/package=interactions

Long, J. A. (2021). Jtools: Analysis and presentation of social scientific data (2.1.3) [Computer software]. https://cran.r-project.org/package=jtools

Luhmann, M. (2015). *R für Einsteiger: Einführung in die Statistik-Software für die Sozialwissenschaften* [R for beginners: An introduction to the statistics software for social scientists]. Beltz.

Lutz, C., Pieter Hoffmann, C., & Meckel, M. (2017). Online serendipity: A contextual differentiation of antecedents and outcomes. *Journal of the Association for Information Science and Technology*, *68*(7), 1698–1710.

Maccatrozzo, V., Terstall, M., Aroyo, L., & Schreiber, G. (2017). SIRUP: Serendipity in recommendations via user perceptions. In G. A. Papadopoulos, T. Kuflik, F. Chen, C. Duarte, & W.-T. Fu (Eds.), *Proceedings of the 22nd international conference on intelligent user interfaces—IUI '17* (pp. 35–44). ACM. https://doi.org/10.1145/3025171.3025185

McCay-Peet, L. (2013). *Investigating work-related serendipity, what influences it, and how it may be facilitated in digital environments.* [Doctoral dissertation, Dalhousie University]. DalSpaceg. http://dalspace.library.dal.ca/handle/10222/42727

McCombs, M., & Poindexter, P. (1983). The duty to keep informed: News exposure and civic obligation. *Journal of Communication*, *33*(2), 88–96. https://doi.org/10.1111/j.1460-2466.1983.tb02391.x

Möller, J., Trilling, D., Helberger, N., & van Es, B. (2018). Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society*, *21*(7), 959–977. https://doi.org/10.1080/1369118X.2018.1444076

Monzer, C., Moeller, J., Helberger, N., & Eskens, S. (2020). User perspectives on the news personalisation process: Agency, trust, and utility as building blocks. *Digital Journalism*, *8*(9), 1142–1162. https://doi.org/10.1080/21670811

Nelson, J. L., & Kim, S. J. (2020). Improve trust, increase loyalty? Analyzing the relationship between news credibility and consumption. *Journalism Practice*, *15*(3), 348–365. https://doi.org/10.1080/17512786.2020.1719874

News API. (2021). *Documentation*. https://newsapi.org/docs

Nielsen, R. K. (2016). People want personalised recommendations (even as they worry about the consequences). In N. Newman, R. Fletcher, D. A. L. Levy, & R. K. Nielsen (Eds.), *Reuters institute digital news report 2016* (pp. 112–114). Reuters Institute.

Nordheim, G. v., & Fuchsloch, S. (2019). Der gemeinwohlorientierte Intermediär [The common good orientated intermediary]. In J. Krone & A. Gebesmair (Eds.), *Reihe Medienstrukturen: Band 14, Zur Ökonomie gemeinwohlorientierter Medien: Massenkommunikation in Deutschland, Österreich und der Schweiz* [Series media structures: Volume 14, On the economics of public interest media: Mass communication in Germany, Austria and Switzerland] (pp. 245–260). Nomos.

Reviglio, U. (2019). Serendipity as an emerging design

principle of the infosphere: Challenges and opportunities. *Ethics and Information Technology*, *21*(2), 151–166. https://doi.org/10.1007/s10676-018-9496-y

Reviglio, U. (2020). Towards a right not to be deceived? An interdisciplinary analysis of media personalization in the light of the GDPR. In I. O. Pappas, P. Mikalef, Y. K. Dwivedi, L. Jaccheri, J. Krogstie, & M. Mäntymäki (Eds.), *Digital transformation for a sustainable society in the 21st century* (pp. 47–59). Springer.

Schlink, S., & Walther, E. (2007). Kurz und gut: Eine deutsche Kurzskala zur Erfassung des Bedürfnisses nach kognitiver Geschlossenheit [Short and good: A German short scale for assessing the need for cognitive closure]. *Zeitschrift Für Sozialpsychologie*, *38*(3), 153–161. https://doi.org/10.1024/0044-3514.38.3.153

Schoenbach, K. (2007). "The own in the foreign": Reliable surprise—An important function of the media? *Media, Culture & Society*, *29*(2), 344–353. https://doi.org/10.1177/0163443707074269

Schweiger, W., Weber, P., Prochazka, F., & Brückner, L. (2019). *Algorithmisch personalisierte Nachrichtenkanäle: Begriffe, Nutzung, Wirkung* [Algorithmic personalized news channels: Terminology, usage, effects]. Springer.

Spangher, A. (2015, August 11). Building the next *New York Times* recommendation engine. *The New York Times*. https://open.blogs.nytimes.com/2015/08/11/building-the-next-new-york-times-recommendation-engine

Stray, J. (2020). Aligning AI optimization to community well-being. *International Journal of Community Well-Being*, *3*(4), 443–463. https://doi.org/10.1007/s42413-020-00086-3

Thurman, N., Moeller, J., Helberger, N., & Trilling, D. (2019). My friends, editors, algorithms, and I. *Digital Journalism*, *7*(4), 447–469. https://doi.org/10.1080/21670811.2018.1493936

Thurman, N., & Schifferes, S. (2012). The future of personalization at news websites. *Journalism Studies*, *13*(5/6), 775–790. https://doi.org/10.1080/1461670X.2012.664341

Wang, T., & Fu, Y. (2020). Item-based collaborative filtering with BERT. In S. Malmasi, S. Kallumadi, N. Ueffing, O. Rokhlenko, E. Agichtein, & I. Guy (Eds.), *Proceedings of the 3rd workshop on e-Commerce and NLP* (pp. 54–58). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.ecnlp-1.8

Wang, W., Bi, B., Yan, M., Wu, C., Bao, Z., Xia, J., Peng, L., & Si, L. (2019). *StructBERT: Incorporating language structures into pre-training for deep language understanding*. ArXiv. http://arxiv.org/pdf/1908.04577v3

Webster, D. M., & Kruglanski, A. W. (1997). Cognitive and social consequences of the need for cognitive closure. *European Review of Social Psychology*, *8*(1), 133–173. https://doi.org/10.1080/14792779643000100

Zajonc, R. B. (1968). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*, 117–123.

About the Authors

Mareike Wieland is a research assistant and doctoral student in journalism and communication studies at the University of Hamburg. Her research focuses on the use, perception, and processing of political information in automated news environments.

Gerret von Nordheim is a postdoctoral researcher in the field of communication science at the University of Hamburg. His research focuses on intermedia effects in the hybrid media system, especially between journalism, social media, and populist actors. He is specialized in the field of computational methods.

Katharina Kleinen-von Königslöw is a professor of journalism and communication studies, especially digital communication and sustainability at the University of Hamburg. Her research focuses on the impact of digitalisation and technological innovations on political communication and, in particular, the role of social network platforms and their use by citizens and political actors.

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 4, Pages 222–233
https://doi.org/10.17645/mac.v9i4.4062

Article

Automated Trouble: The Role of Algorithmic Selection in Harms on Social Media Platforms

Florian Saurwein [1,*] and Charlotte Spencer-Smith [2]

[1] Institute for Comparative Media and Communication Studies, Austrian Academy of Sciences, University of Klagenfurt, Austria; E-Mail: florian.saurwein@oeaw.ac.at
[2] Department of Communication Studies, Paris Lodron University of Salzburg, Austria;
E-Mail: charlotte.spencer-smith@sbg.ac.at

* Corresponding author

Submitted: 25 January 2021 | Accepted: 3 June 2021 | Published: in press

Abstract
Social media platforms like Facebook, YouTube, and Twitter have become major objects of criticism for reasons such as privacy violations, anticompetitive practices, and interference in public elections. Some of these problems have been associated with algorithms, but the roles that algorithms play in the emergence of different harms have not yet been systematically explored. This article contributes to closing this research gap with an investigation of the link between algorithms and harms on social media platforms. Evidence of harms involving social media algorithms was collected from media reports and academic papers within a two-year timeframe from 2018 to 2019, covering Facebook, YouTube, Instagram, and Twitter. Harms with similar casual mechanisms were grouped together to inductively develop a typology of algorithmic harm based on the mechanisms involved in their emergence: (1) algorithmic errors, undesirable, or disturbing selections; (2) manipulation by users to achieve algorithmic outputs to harass other users or disrupt public discourse; (3) algorithmic reinforcement of pre-existing harms and inequalities in society; (4) enablement of harmful practices that are opaque and discriminatory; and (5) strengthening of platform power over users, markets, and society. Although the analysis emphasizes the role of algorithms as a cause of online harms, it also demonstrates that harms do not arise from the application of algorithms alone. Instead, harms can be best conceived of as socio-technical assemblages, composed of the use and design of algorithms, platform design, commercial interests, social practices, and context. The article concludes with reflections on possible governance interventions in response to identified socio-technical mechanisms of harm. Notably, while algorithmic errors may be fixed by platforms themselves, growing platform power calls for external oversight.

Keywords
algorithmic content curation; algorithmic harm; algorithms; behavioural advertising; content moderation; internet; social media

Issue
This article is part of the issue “Algorithmic Systems in the Digital Society” edited by Sanne Kruikemeier (University of Amsterdam, The Netherlands), Sophie Boerman (University of Amsterdam, The Netherlands) and Nadine Bol (Tilburg University, The Netherlands).

1. Introduction

In recent years, internet platforms have gained enormously in reach and influence among the broader population. Scholars have therefore pointed to a trend towards platformisation and the development of a platform society (Helmond, 2015; van Dijck et al., 2018). Among platform services, social media enjoy particular popularity: 2,7 billion people now use at least one of the applications owned by Facebook (Facebook, Instagram, WhatsApp, Messenger), of which Facebook alone has 1,84 billion daily active users (Facebook

Investor Relations, 2021). However, the rise of social media platforms has also attracted strong criticism due to a range of social and economic harms. This includes privacy violations through collection and processing of user data, the potential for social sorting and discrimination, the growth of surveillance capitalism (Zuboff, 2019), and the promotion of intense and addictive user behaviour. Critics point to the consequences of growing platform power (e.g., Helberger, 2020), as internet platforms exert significant influence over societal communication and can strategically use their technologies to direct user attention and shape reality. Platforms stand accused of harming public discourse and democracy by fuelling social fragmentation, political bias, and polarisation, and by contributing to the spread of problematic content such as hate speech and disinformation (Persily & Tucker, 2020). Moreover, platforms are criticised for perpetuating economic harms, such as increasing the dependence of sectors like retail and publishing on platform intermediaries, evading tax, and abusing dominant market positions, which has already led to antitrust cases and fines in Europe and the US (US House Judiciary Subcommittee on Antitrust, Commercial, and Administrative Law, 2020). In addition to this, a long-standing debate continues over whether platforms damage economic rights by enabling copyright violations or damage freedom of expression by enforcing copyright too heavy-handedly.

In particular, algorithms have been identified as a contributing factor in a number of these harms. Indeed, several applications on social media platforms are based on algorithms. Algorithms are used to perform functions such as monitoring, scoring, recommendation, forecasting, and automated transactions. Platforms use algorithms to provide personalised news feeds, features on "trending topics," search and autocomplete functions, computational advertising, contact and group recommendations, as well as to identify and filter unwanted content (e.g., pornography, spam, disinformation). In addition, third parties use their own algorithmic tools on social media platforms, such as chatbots and clickbots. Third parties also deploy algorithms for the purposes of social scoring to offer credit or insurance based on the analysis of the customer's social media posts, and the People's Republic of China is developing a social credit system. In sum, algorithms play a wide range of roles on social media platforms and are believed to contribute to several harms that have emerged with the development of the internet.

When harms arise, it may be tempting to "blame it" on the algorithm, but the link itself between algorithms and harm is often unclear. In addressing problems caused by algorithms, the literature describes harms that range from damage to democratic processes (Tufekci, 2015) to economic harm, which may be genuinely unintentional or purposefully abusive (Muller, 2020). In the context of this article, we use the term "algorithmic harm" to describe harmful or negative effects upon individuals, markets, and society caused in part or in full by the use of algorithms. Based on this definition, the aim of the article is to contribute to the understanding of algorithmic harm by exploring the roles that algorithms play in the emergence of harms on social media platforms. To analyse these roles, the article provides an introductory primer on the use of algorithms on social media platforms, describing their application and purpose in recommendation, content moderation, and advertising. After a brief description of the method, the article goes on to present the results. From a broad collection of case studies of harms, the article develops a typology that distinguishes five areas of harm according to the role that algorithms play in their emergence: Algorithms are (1) deficient tools that lead to errors, (2) instruments that serve manipulation, (3) amplifiers of problematic content, (4) enabling structures for problematic behaviour, and (5) instruments of platform power. This typology contributes to a nuanced understanding of the role of algorithms in the emergence of harms and offers a basis to draw preliminary conclusions for governance strategies to combat algorithmic harms.

2. Areas of Application of Algorithms on Social Media Platforms

In recent years, attention has been increasingly paid to algorithms as they enter more and more areas of public life. On the Internet, algorithms are abstract procedures implemented in software programmes that transform input through specified computational procedures (throughput) into output. Many of these programmes are developed to handle the massive data and information available online (input). They therefore screen and assign relevance to data, select information, and put it into order (throughput). The output may take on different forms to be used in functions such as rankings, recommendations, price setting, or text. Latzer et al. (2016, p. 397) suggest the term "algorithmic selection" to describe these operations, defined as "a process that assigns relevance to information elements of a data set by an automated, statistical assessment of decentrally-generated data signals." The centrepiece of this process model is the throughput stage at which the algorithms operate that define the input–output relationship. Although input, throughput, and output vary for different services, algorithmic selection builds the techno-functional core of a number of internet applications in a broad range of fields and for diverse functions, such as search (e.g., search engines), aggregation (e.g., news aggregators), observation and surveillance (e.g., government surveillance), forecasting (e.g., predictive policing), recommendation (e.g., music platforms), scoring (e.g., credit scoring), content production (e.g., robot journalism), and allocation (e.g., computational advertising). Social media platforms too rely heavily on algorithmic selection. For the purposes of orientation, this article highlights three key areas of application

central to the day-to-day operations of major social media platforms.

2.1. Curation, Recommendation, and Discovery

As users upload content to social media platforms at an incredible rate (Hale, 2019), algorithms help sort through the flood of information to show the most relevant content to each user. Such algorithms take into account the interests, preferences, past behaviour, and predicted behaviour of a particular user to recommend content that might interest them (Cobbe & Singh, 2019). They may also recommend content that is popular among other, similar users, or among all users on the platform. Well-known examples are the Facebook News Feed algorithm (DeVito, 2017) and the YouTube algorithm that selects the next video to play automatically. Such personalised recommendation algorithms help users by showing them relevant content, but they are also engineered in the interests of the platform to maximise user time and engagement on the site (Bergen, 2019). Beyond the feed, algorithms also suggest other users to connect with, pages to follow, or groups to join. Search engines autocomplete functions operate while a user is typing a term into the search engine of a platform, making suggestions by automatically completing the search term. This may reveal terms that other users have searched for, or other combinations that the user might not otherwise have thought of. In summary, curation, recommendation, and discovery systems offer personalisation of services across millions of users and allow users to find relevant content among the ocean of information online, including not just "more of the same" but also new content they might be interested in (McKelvey & Hunt, 2019).

2.2. Content Moderation

Major social media platforms use filtering mechanisms to identify problematic content and remove or hide it either automatically or after human review. As well as deleting child abuse imagery, terrorist propaganda, copyright infringement, and other illegal content as mandated by national laws, platforms have developed their own community guidelines, enforced by a combination of algorithms and human content moderators. Such rules affect content such as nudity, bullying and harassment, toxic language and hate speech, spam, and deceptive or "fake" accounts (Gillespie, 2018; Saurwein & Spencer-Smith, 2019). Due to the sheer volume of content, this task would be impossible with human labour alone, so platforms use algorithms to deal with this problem of scale (Gillespie, 2018). Pattern-matching content moderation algorithms identify patterns in text, images, video, audio, and user behaviour. These algorithms are continually updated with new information and indicators, known as classifiers, and retrained, so variables and results vary continually. At a high certainty, the algorithms might delete content automatically, and at a lower certainty, the content is sent to a human content moderator (Bradford et al., 2019). In cases where illegal content such as child sexual abuse imagery, terrorist propaganda, and copyright violations have already been identified, the content is provided with a unique identifier called a hash, and can be automatically identified and blocked if users attempt to re-upload it to platforms (Gorwa et al., 2020). While hotly debated due to potential consequences for freedom of expression, the use of algorithms in content moderation enables platforms to quickly remove the most abhorrent kinds of content and helps provide a safer environment for users.

2.3. Allocation of Advertising

As social media platforms do not charge fees to their users, targeted advertising plays a central role in their business models. In contrast to television or print advertising, in which advertisers choose the context in which their advertising is shown, on social media platforms, advertisers can directly select the audience. Social media platforms are able to offer detailed target group definition due to the large quantities of data they hold about users (Busch, 2016). Data is gathered from users' profile information, user behaviour, and their connections. In the US, Facebook previously embellished this with household income and financial data from third party data brokers, although it has since discontinued this practice (Williams & Gebhart, 2018). Platforms are also known to make algorithmic inferences about users from existing data to create new advertising categories. For example, according to information provided by a Facebook spokesperson, "multicultural affinity" is not a category that users assign themselves but is automatically inferred according to pages and posts users have engaged with (Angwin & Parris, 2016). For advertisers, targeted advertising has several advantages over traditional advertising, in terms of automation, accuracy, efficiency, and control. Digital technologies offer more data points to profile individual consumers and allow advertisers to target audiences more precisely. Better profiling and targeting are intended to provide consumers with more relevant information, with which they are more likely to engage (Bodó et al., 2017). Digital formats give advertisers better feedback and control of the process, and allow for experimentation at comparatively lower costs. However, the extensive data collection and processing involved has given rise to concerns about the development of "surveillance capitalism" that comes at the expense of user privacy (Zuboff, 2019).

3. Areas of Algorithmic Harm

While the use of algorithms on social media platforms provides several benefits in terms of user experience and business optimization, it is also accompanied by harms that are subject to increasing public concern. This

section explores these harms and develops a typology that distinguishes harms according to the role that algorithms play in their emergence. In this analysis, evidence of harms involving social media algorithms was collected from media reports and academic papers within a two-year timeframe (from 2018 to 2019), covering Facebook, YouTube, Instagram, and Twitter. The reports were collected through internet searches for the term "social media algorithms" and related search terms in the English and German languages. In a first step, the reports were screened for descriptions of harms and mentions of algorithms in association with harms. As soon as unfamiliar harms were identified, these were investigated further through literature research. In a second step, the harms identified were analysed regarding the causal mechanism of their emergence and the particular role played by algorithms. Harms with similar casual mechanisms were grouped together to inductively develop a typology of algorithmic harms. For example, one kind of harm was caused by users consciously manipulating algorithms with malicious intent. All instances of harm of this nature across different platforms were thus grouped in this harm area. The harm areas were developed inductively until all the instances of harm identified in the study could be assigned to a group. This procedure led to a differentiation of five areas of algorithmic harm:

1. Errors: Algorithms make unsuitable, undesirable or disturbing selections.
2. Manipulation: Algorithms are manipulated by users to produce algorithmic outputs that harass other users or disrupt public discourse.
3. Reinforcement effects: Algorithms strengthen pre-existing harms and inequalities in society.
4. Enablement of harmful practices: Algorithms provide the infrastructure that enables harmful behaviour, e.g., targeted advertising that is opaque and discriminatory.
5. Platform power: Algorithms establish or strengthen platform power over users, markets, and society, and thus pose a challenge to competition, consumers, and individual rights.

3.1. Errors

The first category of algorithmic harm is errors and unsuitable selections by algorithms. Here, algorithms can be seen as the "wrong tool for the job" that make selections lacking human judgement and sensitivity. From a technical standpoint, an algorithm cannot "make a mistake": When a content moderation algorithm deletes a photo of a nude statue, it is carrying out its programmed instructions. From the standpoint of the platform's policy, however, this action is erroneous, because photos of nude statues are not banned. Thus, an algorithmic error refers to an algorithmic decision that produces an outcome at odds with rules, policy, or intention of the algorithm's proprietor.

A well-known example of algorithmic error is the problem of "overblocking" in content moderation. As algorithms are unable to understand the context of a post, this can lead to content being flagged and/or removed when it should not have been. At various points, algorithms have mistaken nudity in art for pornography because they are able to detect patterns that indicate nudity but cannot differentiate between the contexts of art and pornography (Gillespie, 2018). Similarly, algorithms may identify certain keywords or speech patterns as hate speech without being able to evaluate context or intent (Gorwa et al., 2020). As well as impacting freedom of expression, algorithmic errors can have consequences for economic rights, particularly in automated copyright enforcement (Lessig, 1999; Rugnetta, 2018) and on platforms such as YouTube, where users can be sanctioned by losing the ability to earn income from their content (Caplan & Gillespie, 2020).

Alongside enforcement errors, inappropriate recommendations can also be considered as a form of algorithmic selection at odds with the intentions of the platform. Dubbed by Bucher (2016) as "cruel connections," a well-known example of this occurred on Facebook when a user was automatically shown an algorithmically-generated "year in review" album of his posts, which featured a picture of his recently deceased child (Meyer, 2014). Such examples underscore algorithms' lack of understanding for context that a human would have. To summarise, algorithms make unsuitable selections lacking human judgement and sensitivity. However, this harm does not arise from algorithms alone; it can rather be considered an assemblage that encompasses component parts including data that is imbued with social meaning and context, and platforms that seek to automate potentially sensitive tasks using such data.

3.2. Manipulation

Algorithmic selections can be manipulated by users for commercial or abusive purposes, with the outcome that they harass others, disrupt public discourse, and cause harm. This can be seen as a form of "manipulation of institutions or systems," which has at its goal the attainment of covert influence over the people using the platform (Susser et al., 2019, p. 13). Computational propaganda, for instance, refers to "the ways in which the use of algorithms, automation (most often in the form of political bots), and human curation are used over social media to purposefully distribute misleading information" (Woolley, 2020, p. 90). The Russian troll factory Internet Research Agency, for example, allegedly used bots to like and share social media posts from certain accounts so that social media algorithms would consider them popular and be more likely to share them (Osipova & Byrd, 2017). As well as using bots to make content appear more popular than it really is, groups of users can act together in coordinated campaigns to make their content more likely to be recommended by algorithms (Gillespie,

2014). For example, the German far-right internet hate group Reconquista Germanica coordinated their members to post the same hashtags on Twitter at the same time, so that the hashtags would be selected by the algorithm to appear in the top Twitter trends (Kreißel et al., 2018).

It is important to note that the line between "genuine'' behaviours, "legitimate optimization," and "illegitimate, manipulative" behaviours that "game the algorithm" is a thin one (Gillespie, 2017), particularly as users pursuing such strategies, to whatever end, are all ultimately incentivised by the algorithmic logics of the platforms. Indeed, users engage in behaviours designed to manipulate social media algorithms without intending or causing harm to others. On Instagram, users form "engagement pods" in a mutually beneficial arrangement to boost each other's content algorithmically (O'Meara, 2020). However, attempts to exploit the algorithmic logics of platforms can also lead to grotesque consequences, as is evidenced by the "Elsagate" controversy on YouTube, in which inappropriate content, e.g., showing popular children's characters in disturbing situations, are recommended to young audiences on YouTube, who are too young to enter search terms and are thus wholly reliant on recommender systems (Jaakola, 2019). This has resulted in YouTube channels creating increasingly bizarre and troubling content for children by orienting themselves towards the algorithm for commercial purposes (Bridle, 2017). To summarise this kind of harm, the role of the algorithm is as a means that can be manipulated to produce a harmful outcome. This harm does not emerge from the algorithm alone, but from an assemblage that encompasses platforms that offer content recommendations, and thus promise publicity or commercial gain, and users who employ tactics to exploit the logics of algorithms (O'Meara, 2020).

3.3. Reinforcement Effects

Algorithms reinforce, strengthen, or amplify pre-existing phenomena that pose a threat to public discourse and democracy, such as spreading hate speech and disinformation, and entrenching polarisation and radicalisation. Here, algorithms act as a strengthener of, or catalyst for, pre-existing harms that have been present in communication since pre-internet times, but that have been accelerated by the introduction of algorithms. One example of this is the amplification of hate speech and disinformation online. Especially posts that generate strong emotions attract high levels of engagement, which signals high relevance to recommendation algorithms and leads to further recommendation to other users (e.g., Stark et al., 2020, p. 40). Observers claim that algorithms such as the Facebook News Feed algorithm play a role in how hate speech posts go viral, inspiring real-life violence in Sri Lanka (Taub & Fisher, 2018a) and Myanmar (McLaughlin, 2018). When it comes to disinformation, it has been hypothesised that disinformation content achieves amplification by provoking curiosity through novelty, as well as anger through outrage (Vosoughi et al., 2018). In one example, shortly after Facebook shifted its "trending topics" feature from human to algorithmic curation, a number of disinformation stories appeared on it, including a fake story about US journalist Megyn Kelly being fired from Fox News, as the algorithms boosted popular stories without being able to sift out false information (Ohlheiser, 2016).

Another facet of reinforcement is the concept of the "filter bubble" (Pariser, 2011), a personalised media environment that develops when algorithms select content personally tailored to user preferences. While amplification is a phenomenon that occurs across a platform, the filter bubble is generated at the level of the individual user, as algorithms recommend content that fit the algorithmically-assigned interests of that user and the user's activity on the platform provides further feedback to the algorithm. It is argued that algorithms reinforce confirmation bias because they predominantly deliver opinions that affirm pre-existing beliefs and mislead users into believing that everyone else holds the same opinions as them, creating an echo chamber. Echo chambers can also induce people to believe that hatred of a particular group is the social norm (Taub & Fisher, 2018b). It is hypothesised that algorithmic personalisation reduces exposure to different content and new ideas, with potentially negative outcomes for innovation and the development of new ideas (Sunstein, 2001). However, empirical studies have suggested that the impact of filter bubbles is limited (Zuiderveen Borgesius et al., 2016) and moderated by an environment in which a variety of different media continue to be consumed (Dubois & Blank, 2018). Recently, however, the phenomenon of "rabbit holes" has come to attention in the media, in which recommendation algorithms contribute to radicalisation of users by recommending more and more extreme content, such as conspiracy theories (Lewis, 2018). Furthermore, research suggests that algorithmically-mediated advertising on social media reinforces gender and age stereotyping by showing ads to users that fit stereotypes, such as showing advertising about beauty to women or fashion to younger people (Bol et al., 2020), and that not just advertiser choice or user preferences play a role, but also algorithmic selection (Ali et al., 2019). By charging more per click for advertising to audiences that are not in the perceived core market for an ad, platforms may also be making it more difficult for political parties to break through the "filter bubble" to reach users outside their traditional voter base (Ali et al., 2021).

To summarise, the role of the algorithm is as a technology that reinforces problematic content and harmful conduct. Here, algorithms are part of the interplay between content, platform logics and user behaviour. In particular, the algorithms in question operate in the context of recommender systems and are thus engineered to recommend content with high levels of

engagement. The harm arises when it promotes content that has little impact on society in small quantities but becomes problematic when it is amplified across many users or reinforces problematic worldviews in individual users (Cobbe & Singh, 2019). In addition, a role is played by the large numbers of users who engage with such content by liking, commenting, sharing, and clicking on it.

3.4. Enabling Harmful Practices

Algorithms can also enable actors to carry out discriminatory practices, particularly through online advertising. Here, algorithms are used as infrastructure to target or exclude certain groups of users, with harmful effects. For example, Facebook uses data and algorithms to determine if users in the US belong to an ethnic minority for the purpose of advertising to those ethnic groups. However, the same functions have been used for manipulative purposes. The Trump 2016 presidential campaign disclosed that it targeted Facebook ads to African Americans to discourage them from voting (Green & Issenberg, 2016).

The same functions also made it possible to exclude ethnic minorities from seeing certain ads, as journalists have found instances where it was possible to exclude ethnic minorities from seeing ads for housing and accommodation (Angwin et al., 2017; Cotter et al., 2021). Facebook disabled advertisers' ability to exclude ethnic minorities at the end of 2017, but the incident nonetheless shows how platforms have not carefully considered how automated, targeted advertising can be used to suppress and discriminate against marginalised groups. In addition, targeted advertising ensures that it is only seen by the target audience and not by others. This is particularly troubling in political microtargeting, as a political advertiser can send different voters different, contradictory information while avoiding broader public scrutiny. This decreases the transparency of campaigns, political positions, and electoral promises and could lead to a skewed perception of priorities of political parties among voters (Zuiderveen Borgesius et al., 2018, pp. 87–89). To summarise, the role of the algorithm is as an infrastructure that enables harmful practices, such as discrimination. That said, not the algorithm alone is at fault: It is rather part of an assemblage of the infrastructure of online advertising that is intentionally designed to include and exclude segments of the audience to optimize targeting, as well as the social ills that can be strengthened by such techniques.

3.5. Platform Power

Algorithms may strengthen platform power, particularly over competitors, markets, and users. Here, the role of algorithms is as a tool of influence and surveillance over other actors. The use of big data and algorithms can enable a "God view," using "big data and big analytics for a clearer overview of the marketplace at any given moment" (Ezrachi & Stucke, 2016, p. 72). For example, the Facebook app Onavo Protect offered users a VPN service while also collecting data on how users use competitor apps. The information is alleged to have informed Facebook's decisions about which app features to imitate, including stories from Snapchat, and which companies to acquire, including WhatsApp and Instagram (Seetharaman & Morris, 2017). Monitoring through Onavo Protect may be one of a number of anticompetitive practices in the technology market ("American tech giants," 2018).

Algorithms have also contributed to the unequal relationship between platforms and certain markets, particularly in fields of publishing that are particularly dependent on social media platforms for distribution. The rise of algorithm-based targeted advertising on internet platforms has contributed to the disruption of traditional funding models for journalism (Lobigs, 2016, pp. 103–104), and publishers have become increasingly dependent on social media to the extent that Facebook has been described as a "kingmaker" (Pasquale, 2015). The dependence of publishers on social media algorithms is exemplified by Facebook's shifting priorities when it comes to video. When Facebook increased the importance of video content in the News Feed algorithm in 2015, publishers active on social media responded by moving resources to video production (Griffith, 2015). This, however, proved short-lived as Facebook decided to assign less priority to video in the algorithm three years later (Vogelstein, 2018) and the same publishers then made social video employees redundant (Bilton, 2018). The pivot to video can be seen as an example of how algorithms are used to impose the changing commercial interests of a social platform on sectors of the publishing industry that are particularly vulnerable to algorithmic change (Oremus, 2018). Indeed, social media creators who are commercially active on platforms are particularly impacted by changes in recommendation and content moderation algorithms, leading to "algorithmic precarity" (Duffy, 2020). Social media creators carry a higher level of exposure to algorithmic change, and thus experience heightened algorithmic precarity, due to their particular dependence on platforms for distribution.

Finally, algorithms strengthen platform power over users by promoting addictive behaviour and eroding privacy. Although more empirical research is needed, it is believed, for example, that social media platforms use algorithms to withhold and distribute likes and notifications so that users keep checking the app (Peitz, 2017). The use of algorithms also raises complex questions about personal privacy and informational self-determination, especially regarding the use of inferential analytics, in which algorithms make inferences about users, often without their consent (Wachter & Mittelstadt, 2019). A further concern is facial recognition technology (Wolfangel, 2018). The application for a patent for the use of facial recognition for payments

(Moore Davis, 2016), as well as a patent for eye-tracking technology (San Agustin Lopez et al., 2014), both by Facebook, has generated speculation about the depth of observation and data gathering that users may be subjected to in the future. In addition, the US news website *The Intercept* claims to have seen a confidential document in which Facebook outlines a new advertising service that will be able to predict users' future consumer behaviour using machine learning (Biddle, 2018). Such applications fuel fears about the potential for surveillance and social scoring, as well as about consumers' continuing ability to make purchasing choices without covert psychological influence.

To summarise, algorithms are used to strengthen influence and surveillance over other actors, and increase platform power over competitors, markets, and users. The assemblage that produces this harm encompasses commercial platforms upon which user interactions and economic activity take place, all mediated by the respective platform company. The concentration of such power by platforms requires significant quantities of data and algorithms that are able to process them. In turn, these data and algorithms enable platforms to conduct surveillance, as well as to intervene and exert influence to pursue their own goals (Zuboff, 2019).

4. Summary and Conclusion

The aim of this article was to contribute to the understanding of algorithmic harm by exploring the roles that algorithms play in the emergence of harms on social media platforms. The article therefore developed a typology of five areas of algorithmic harm based on the mechanisms of their causation. This analysis demonstrated that algorithms contribute to the emergence of harm in manifold ways. Algorithms can be deficient tools that lead to errors, instruments that serve manipulation, technologies that reinforce and amplify problematic content, enabling infrastructure for problematic behaviour and instruments that serve to establish or strengthen platform power.

However, the analysis also found that harms do not arise from the application of algorithms alone. Instead, harms can be best conceived of as socio-technical assemblages that encompass the use and design of algorithms, platform design, commercial interests, social practices, and context. Altogether, these findings support the suggestion that algorithms are not isolated technical artefacts, but "assemblages of institutionally situated code, human practices and normative logics" (Ananny, 2016, p. 108). It is thus useful to understand how they "work within socio-technical assemblages and how they perform actions and make a difference in particular domains" (Kitchin, 2017, p. 26). This is particularly evident on social media platforms, where algorithms and their implications are inseparable from platform architectures, normative logics, and commercial interests of platform companies (van Dijck, 2013).

In addition, it should be considered that types of harm are not isolated from one another but can interact and intersect. Boundaries between types of harm are porous and permeable. For example, users manipulate content moderation algorithms to produce errors, in an event in which algorithms are both deficient tools that lead to errors and instruments of manipulation. After the Christchurch terror attacks in 2019, social media platforms struggled to prevent users from uploading footage filmed by the shooter, in part because users employed techniques designed to bypass content moderation algorithms, such as superimposing footage of YouTube personalities to make the upload look like video game footage (Timberg et al., 2019). A further technique was to upload the footage as a live stream, preventing the video from being analysed by content moderation algorithms as a fully uploaded file (McDonald, 2019). User evasion of content moderation algorithms is an example of interlocking algorithmic harms that may provide further avenues for research.

The analysis of algorithmic harms inevitably leads to questions as to whether and how to deal with them. From an institutional perspective, options range from market solutions and users' own strategies for countering harms, via voluntary industry self-regulation to command-and-control regulation by state authorities (Latzer et al., 2006). Some algorithmic harm could be reduced by consumers' self-help strategies (opting out of services, switching to other providers and technical self-protection, such as privacy tools; Saurwein et al., 2015). However, there are several barriers to effective self-help, and the potential of user self-protection should not be overestimated. Users may not be able to avoid using services or switch to other providers because of network effects and other barriers. Privacy tools may be able to limit the use of cookies, but do not prevent platforms from gathering data on user behaviour on their services. Moreover, because of the opaque nature of algorithmic selection and low levels of awareness about algorithms among users, algorithmic harm is often barely noticeable to consumers. For example, an average internet user can hardly detect errors, reinforcement of problematic phenomena, or manipulation. Consequently, it is argued that if harms and risks are not visible, then there is no reason to consider self-protection strategies (Saurwein et al., 2015). In practice, however, some countries (e.g., Switzerland) report a considerable level of awareness of algorithms and algorithmic harms (Latzer et al., 2020) while in other countries (e.g., Norway) awareness of algorithms is rather low (Gran et al., 2020). For Germany, Fischer and Petersen (2018) report a widespread unawareness of algorithms, strong indecision about risks and opportunities, discomfort over algorithmic decision-making, and a strong desire for more control.

Regarding control, the typology of harms allows us to reflect upon suitable governance responses (Latzer et al., 2019) by exploring incentives for social media

platforms to reduce harms by means of platform self-regulation. Industry self-regulation is unsuitable in cases where harms are also indistinguishable from the commercial interests of industry players. This is particularly evident in growing platform power, where social media platforms have the least motivation to reduce algorithmic harms. Thus, the current focus of statutory regulation on data protection and antitrust is well justified. In the case of algorithmic errors, however, there are clear incentives for platforms to reduce errors and increase accuracy because functioning automation is key to the further scaling up of services. In content moderation, for instance, platforms have been making efforts to improve the accuracy of automated moderation systems and regularly report performance indicators to demonstrate progress. In the case of manipulation by third parties there are some incentives for platform providers to combat manipulation and maintain the integrity and reputation of their services. Platforms make use of their terms of service to define unwanted behaviour and have made efforts to identify and sanction inauthentic behaviour and block bot accounts. A continual challenge is drawing the line between legitimate optimization and illegitimate gaming.

Compared to errors and manipulation, incentives to counter the reinforcement of problematic phenomena are less clear-cut. It can be argued that amplification of problematic content contributes to profitability, which reduces incentives to curb it. On the other hand, platforms may be motivated to control amplification when it starts to impair user experience and discourage users from spending time on the platforms. Indeed, Facebook now deprioritises "borderline content" in its News Feed algorithm (Zuckerberg, 2018) and the major platforms have proven more willing to act against the spread of problematic content in the context of the Covid-19 pandemic and threats to US democracy that culminated in the storming of the Capitol in January 2021. Similarly, when it comes to harmful advertising practices, platforms have been motivated to disable some problematic features and improve transparency through a political advertising database only after the issue became a public relations problem. This is illustrative of a broader pattern of platform governance as a cycle of "shocks and exceptions" (Gillespie & Ananny, 2016).

Moreover, platforms may be more or less motivated to address harms depending on who is affected by them. A reliance on public relations shocks means a reliance on journalism as a mechanism for uncovering algorithmic harms, as well as other online harms (Diakopoulos, 2015). Considering that most users do not have access to this kind of publicity, relying on journalism as a principal accountability mechanism is not a sustainable means of reducing harm. Furthermore, bias and insufficient employee diversity within platform companies create a blind spot towards algorithmic harms that affect groups who are commonly discriminated against and marginalised in society (Benjamin, 2019; Noble, 2018). These factors could slow the response of companies in addressing harms that affect users who do not have access to publicity or are structurally oppressed in society. The failures of social media companies in addressing algorithmic harm have led to a growing call to increase statutory regulation and oversight. Most recently, the European Commission published a legislative initiative for a Digital Services Act to enhance platform accountability (European Commission, 2020). The proposed regulations also concern algorithm-based services such as recommendation systems, content moderation, and advertising. The regulations shall force very large online platforms to increase the transparency of their algorithmic systems, to provide opportunities to opt-out from profiling and personalisation, to protect services from manipulation, and to carry out risk assessments to avoid the spread of illegal content, restrictions of fundamental rights, and manipulation. The proposal suggests the establishment of external and independent auditing procedures and "technical assistance at EU level, for inspecting and auditing content moderation systems, recommender systems and online advertising" (European Commission, 2020, p. 12). The discussion of a Digital Services Act is at an early stage, but the legislative initiative clearly indicates that algorithmic harms have become a prominent issue on the internet governance agenda, which may lead to stronger control of internet platforms and their algorithm-based modes of operation.

The limitations of our study regarding its scope provide potential impulses for future research. While the article has analysed algorithms on social media platforms, further research could investigate algorithmic harm across other kinds of platforms, such as Amazon and Uber, building upon existing critiques of individual platforms (see Khan, 2017; Muller, 2020). Furthermore, this research focused on platforms popular in North America and Europe, and used sources in the English and German languages, limiting its geographical and cultural scope. Future avenues of research could include investigations of algorithmic harm across non-Western cultural contexts, in particular in areas such as algorithmic content moderation on large global platforms, where implementation across languages and geographic regions is uneven. Finally, analyses of algorithmic harms lead to questions of suitable governance responses. The article provides a set of theoretical reflections upon the incentives for social media platforms to reduce harms by means of platform self-regulation. Future research should verify if the governance of algorithms in fact coincides with the proposed patterns.

Acknowledgments

The article presents results from the project "The Automation of the Social: Algorithmic Selection in Social Online Networks" funded by the Vienna Anniversary Fund for the Austrian Academy of Sciences.

Conflict of Interests

The authors declare no conflict of interests.

References

Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., & Rieke, A. (2019). Discrimination through optimization: How Facebook's ad delivery can lead to biased outcomes. *Proceedings of the ACM on Human-Computer Interaction*, *3*. https://doi.org/10.1145/3359301

Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., & Rieke, A. (2021). Ad delivery algorithms: The hidden arbiters of political messaging. In *WSDM '21: Proceedings of the 14th ACM international conference on web search and data mining* (pp. 13–21). Association for Computing Machinery.

American tech giants are making life tough for startups. (2018, June 2). *The Economist*. https://www.economist.com/business/2018/06/02/american-tech-giants-are-making-life-tough-for-startups

Ananny, M. (2016). Toward an ethics of algorithms: Convening, observation, probability, and timeliness. *Science, Technology, & Human Values*, *41*(1), 93–117.

Angwin, J., & Parris, T. (2016, October 28). Facebook lets advertisers exclude users by race. *ProPublica*. https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race

Angwin, J., Tobin, A., & Varner, M. (2017, November 21). Facebook (still) letting housing advertisers exclude users by race. *ProPublica*. https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the New Jim Code*. Polity Press.

Bergen, M. (2019, April 2). YouTube executives ignored warnings, letting toxic videos run rampant. *Bloomberg*. https://www.bloomberg.com/news/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant

Biddle, S. (2018, April 13). Facebook uses artificial intelligence to predict your future actions for advertisers, says confidential document. *The Intercept*. https://theintercept.com/2018/04/13/facebook-advertising-data-artificial-intelligence-ai

Bilton, R. (2018, February 21). Post-Facebook News Feed tweaks, Vox Media lays off 50 employees. *Nieman Lab*. http://www.niemanlab.org/2018/02/post-facebook-news-feed-tweaks-vox-media-lays-off-50-employees

Bodó, B., Helberger, N., & de Vreese, C. H. (2017). Political micro-targeting: a Manchurian candidate or just a dark horse? *Internet Policy Review*, *6*(4), 1–13.

Bol, N., Strycharz, J., Helberger, N., van de Velde, B., & de Vreese, C. H. (2020). Vulnerability in a tracked society: Combining tracking and survey data to understand who gets targeted with what content. *New Media & Society*, *22*(11), 1996–2017.

Bradford, B., Grisel, F., Meares, T., Owens, E., Pineda, B., Shapiro, J., Tyler, T., & Evans Peterman, D. (2019). *Report of the Facebook data transparency advisory group.* Yale Law School. https://law.yale.edu/sites/default/files/area/center/justice/document/dtag_report_5.22.2019.pdf

Bridle, J. (2017). *Something is wrong on the internet*. Medium. https://medium.com/@jamesbridle/something-is-wrong-on-the-internet-c39c471271d2

Bucher, T. (2016). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, *20*(1), 30–44.

Busch, O. (2016). The programmatic advertising principle. In O. Busch (Ed.), *Programmatic advertising. The successful transformation to automated, data-driven marketing in real-time* (pp. 3–15). Springer.

Caplan, R., & Gillespie, T. (2020). Tiered governance and demonetization: The shifting terms of labor and compensation in the platform economy. *Social Media + Society*, *6*(2), 1–13.

Cobbe, J., & Singh, J. (2019). Regulating recommending: Motivations, considerations, and principles. *European Journal of Law and Technology*, *10*(3), 1–37.

Cotter, K., Medeiros, M., Pak, C., & Thorson, K. (2021). "Reach the right people": The politics of "interests" in Facebook's classification system for ad targeting. *Big Data & Society*, *8*(1), 1–16.

DeVito, M. (2017). From editors to algorithms. *Digital Journalism*, *5*(6), 753–773.

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, *3*(3), 398–415.

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, *21*(5), 729–745.

Duffy, E. B. (2020). Algorithmic precarity in cultural work. *Communication and the Public*, 5(3/4), 103–107.

European Commission. (2020). *Proposal for a regulation of the European Parliament and the Council on a single market for digital services (Digital Services Act) and amending directive 2000/31/EC* (COM(2020)825). https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2020%3A825%3AFIN

Ezrachi, A., & Stucke, M. (2016). *Virtual competition*. Harvard University Press.

Facebook Investor Relations. (2021). *Facebook reports fourth quarter and full year 2020 results*. https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-Fourth-Quarter-and-Full-Year-2020-Results/default.aspx

Fischer, S., & Petersen, T. (2018*). Was Deutschland über Algorithmen weiß und denkt. Ergebnisse einer repräsentativen Bevölkerungsumfrage* [What Germany knows and thinks about algorithms. Results from a representative population survey]. Gütersloh.

Gillespie, T. (2014). The relevance of algorithms. In T.

Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–194). MIT Press.

Gillespie, T. (2017). Algorithmically recognizable: Santorum's Google problem, and Google's Santorum problem. *Information, Communication & Society*, *20*(1), 63–80.

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Gillespie, T., & Ananny, M. (2016). Public platforms: Beyond the cycle of shocks and exceptions. *Oxford Internet Institute*. Retrieved from http://blogs.oii.ox.ac.uk/ipp-conference/2016/programme-2016/track-b-governance/platform-studies/tarleton-gillespie-mike-ananny.html

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, *7*(1), 1–15.

Gran, A.-B., Booth, P., & Bucher, T. (2020). To be or not to be algorithm aware: a question of a new digital divide? *Information, Communication & Society*. Advance online publication. https://doi.org/10.1080/1369118X.2020.1736124

Green, J., & Issenberg, S. (2016, October 26). Why the Trump machine is built to last beyond the election. *Bloomberg*. https://www.bloomberg.com/news/articles/2016-10-27/inside-the-trump-bunker-with-12-days-to-go

Griffith, E. (2015, June 3). How Facebook's video-traffic explosion is shaking up the advertising world. *Fortune*. http://fortune.com/2015/06/03/facebook-video-traffic

Hale, J. (2019, May 7). More than 500 hours of content are now being uploaded to YouTube every minute. *Tubefilter*. https://www.tubefilter.com/2019/05/07/number-hours-video-uploaded-to-youtube-per-minute

Helberger, N. (2020). The political power of platforms: How current attempts to regulate misinformation amplify opinion power. *Digital Journalism*, *8*(6), 842–854.

Helmond, A. (2015). The platformization of the web: Making web data platform ready. *Social Media + Society*, *1*(2), 1–11.

Jaakola, M. (2019). From vernacularized commercialism to kidbait: Toy review videos on YouTube and the problematics of the mash-up genre. *Journal of Children and Media*, *14*(2), 237–254.

Khan, L. M. (2017). Amazon's antitrust paradox. *Yale Law Journal*, *126*(3), 710–805.

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), 14–29.

Kreißel, P., Ebner, J., Urban, A., & Jakob, G. (2018). *Hass auf Knopdruck. Rechtsextreme Trollfabriken und das Ökosystem koordinierter Hasskampagnen im Netz* [Hate at the touch of a button. Far-right troll factories and the ecosystem of coordinated hate campaigns on the Internet]. Institute for Strategic Dialogue.

Latzer, M., Festic, N., & Kappeler, K. (2020). *Awareness of risks related to algorithmic selection in Switzerland.* University of Zurich. https://mediachange.ch/media//pdf/publications/Report_3_Risks.pdf

Latzer, M., Hollnbuchner, K., Just, N., & Saurwein, F. (2016). The economics of algorithmic selection on the Internet. In J. Bauer & M. Latzer (Eds.), *Handbook on the economics of the internet* (pp. 395–425). Edward Elgar.

Latzer, M., Just, N., Saurwein, F., & Slominski, P. (2006). Institutional variety in communications regulation. Classification scheme and empirical evidence from Austria. *Telecommunications Policy*, *30*(3/4), 152–170.

Latzer, M., Saurwein, F., & Just, N. (2019). Assessing policy II: Governance-choice method. In H. van den Bulck, M. Puppis, K. Donders, & L. van Audenhove (Eds.), *The Palgrave handbook of methods for media policy research* (pp. 557–574). Palgrave Macmillan.

Lessig, L. (1999). *Code and other laws of cyberspace*. Basic Books.

Lewis, P. (2018, February 2). 'Fiction is outperforming reality': How YouTube's algorithm distorts truth. *The Guardian*. https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth

Lobigs, F. (2016). Finanzierung des Journalismus—Von langsamen und schnellen Disruptionen [Financing of journalism—Of slow and fast disruptions]. In K. Meier & C. Neuberger, C. (Eds.), *Journalismusforschung. Stand und Perspektiven* [Journalism research. Current state and perspectives] (pp. 69–137). Nomos.

McDonald, S. (2019, March 15). Google AI has trouble keeping NZ massacre video off YouTube. *Newsweek*. https://www.newsweek.com/google-ai-has-trouble-keeping-nz-massacre-video-youtube-1365375

McKelvey, F., & Hunt, R. (2019). Discoverability: Toward a definition of content discovery through platforms. *Social Media + Society*, *5*(1), 1–15.

McLaughlin, T. (2018, July 6). How Facebook's rise fuelled chaos and confusion in Myanmar. *Wired*. https://www.wired.com/story/how-facebooks-rise-fueled-chaos-and-confusion-in-myanmar

Meyer, E. (2014). *Inadvertent algorithmic cruelty*. Meyerweb. http://meyerweb.com/eric/thoughts/2014/12/24/inadvertent-algorithmic-cruelty

Moore Davis, S. (2016). *Facial recognition identification for in-store payment transactions* (US Patent No. US20170323299). Patent and Trademark Office.

Muller, Z. (2020). Algorithmic harms to workers in the platform economy: The case of Uber. *Columbia Journal of Law and Social Problems*, *53*(2), 167–210.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Ohlheiser, A. (2016, August 29). Three days after removing human editors, Facebook is already trending fake news. *Washington Post*. https://www.washingtonpost.com/news/the-intersect/wp/2016/08/29/a-fake-headline-about-megyn-kelly-was-trending-on-facebook

O'Meara, V. (2020). Weapons of the chic: Instagram influencer engagement pods as practices of resistance to Instagram platform labor. *Social Media + Society*, *5*(4), 1–11.

Oremus, W. (2018, October 18). The big lie behind the 'pivot to video.' *Slate Magazine*. https://slate.com/technology/2018/10/facebook-online-video-pivot-metrics-false.html

Osipova, N. V., & Byrd, A. (2017, October 31). Inside Russia's network of bots and trolls. *New York Times*. https://www.nytimes.com/video/us/politics/100000005414346/how-russian-bots-and-trolls-invade-our-lives-and-elections.html

Pariser, E. (2011). *The filter bubble*. Penguin Press.

Pasquale, F. (2015). *The black box society*. Harvard University Press.

Peitz, D. (2017, July 17). Erstmals geben Tech-Leute zu: Wir haben ein echtes Problem [Tech people admit for the first time: We have a real problem]. *Die Zeit*. https://www.zeit.de/digital/2018-07/smartphone nutzung-sucht-david-levy-computerwissenschaftler

Persily, N., & Tucker, J. A. (Eds.). (2020). *Social media and democracy. The state of the field, prospects for reform*. Cambridge University Press.

Rugnetta, M. (2018). *Automated copywrongs*. Reasonably Sound. http://reasonablysound.com/2018/01/15/automated-copywrongs

San Agustin Lopez, J., Sztuk, S., & Henrik Tall, M. (2014). *Systems and methods of eye tracking control* (US Patent No. US9829971B2). Patent and Trademark Office. https://patents.google.com/patent/US9829971B2/en

Saurwein, F., Just, N., & Latzer, M. (2015). Governance of algorithms: Options and limitations. *Info: The Journal of Policy, Regulation and Strategy for Telecommunications, Information and Media*, *17*(6), 35–49.

Saurwein, F., & Spencer-Smith, C. (2019). *Inhaltsregulierung auf Internet-Plattformen. Optionen für verantwortungsvolle Governance auf nationaler Ebene* [Content moderation on internet platforms. Options for accountability-oriented governance at national level] (Research Report). CMC.

Seetharaman, D., & Morris, B. (2017, August 13). Facebook's Onavo gives social-media firm inside peek at rivals' users. *Wall Street Journal*. https://www.wsj.com/articles/facebooks-onavo-gives-social-media-firm-inside-peek-at-rivals-users-1502622003

Stark, B., Stegmann, D., Magin, M., & Jürgens, P. (2020). *Are algorithms a threat to democracy? The rise of intermediaries: A challenge for public discourse*. AlgorithmWatch.

Sunstein, C. (2001). *Echo chambers: Bush v. Gore, impeachment, and beyond*. Princeton University Press.

Susser, D., Roessler, B., & Nissenbaum, H. (2019). Online manipulation: Hidden influences in a digital world. *Georgetown Law Technology Review*, *4*(1), 1–45.

Taub, A., & Fisher, M. (2018a, April 21). Where countries are tinderboxes and Facebook is a match. *New York Times*. https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html

Taub, A., & Fisher, M. (2018b, August 21). Facebook fueled anti-refugee attacks in Germany, new research suggests. *New York Times*. https://www.nytimes.com/2018/08/21/world/europe/facebook-refugee-attacks-germany.html

Timberg, C., Harwell, D., Shaban, H., Ba Tran, A., & Fung, B. (2019, March 15). The New Zealand shooting shows how YouTube and Facebook spread hate and violent images—yet again. *Washington Post*. https://www.washingtonpost.com/technology/2019/03/15/facebook-youtube-twitter-amplified-video-christchurch-mosque-shooting

Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colorado Technology Law Journal*, *13*(2), 203–218.

US House Judiciary Subcommittee on Antitrust, Commercial, and Administrative Law. (2020). *Investigation of competition in digital markets*.

van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford University Press.

van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society*. Oxford University Press.

Vogelstein, F. (2018, January 13). Facebook's Adam Mosseri on why you'll see less video, more from friends. *Wired*. https://www.wired.com/story/facebooks-adam-mosseri-on-why-youll-see-less-video-more-from-friends

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151.

Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review*, *2019*(2), 494–620.

Williams, J., & Gebhart, G. (2018). *Facebook isn't telling the whole story about its decision to stop partnering with data brokers.* Electronic Frontier Foundation. https://www.eff.org/de/deeplinks/2018/04/facebook-isnt-telling-whole-story-about-its-decision-stop-partnering-data-brokers

Wolfangel, E. (2018, March 5). Facebook: Gesichtserkennung lässt sich eben nicht abschalten [Facebook: Facial recognition cannot be turned off]. *Spektrum*. https://www.spektrum.de/kolumne/gesichtserkennung-laesst-sich-eben-nicht-abschalten/1548879

Woolley, S. C. (2020). Bots and computational propaganda: Automation for communication and control.

In N. Persily & J. A. Tucker (Eds.), *Social media and democracy. The state of the field, prospects for reform* (pp. 89–110). Cambridge University Press.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Hachette.

Zuckerberg, M. (2018). *A blueprint for content governance and enforcement*. Facebook. https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634

Zuiderveen Borgesius, F. J., Moeller, J., Kruikemeier, S., Fathaigh, R., Irion, K., Dobber, T., Bodó, B., & de Vreese, C. H. (2018). Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review*, *14*(1), 82–89.

Zuiderveen Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., de Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, *5*(1), 1–16.

About the Authors

Florian Saurwein is a senior scientist at the Institute for Comparative Media and Communication Studies (CMC) of the Austrian Academy of Sciences and the University of Klagenfurt. He studied communication science and political science, and holds a PhD from the University of Zurich. His research centres around interrelations of media, society, and governance, with a current focus on risks and governance of content moderation and algorithmic selection on internet platforms.

Charlotte Spencer-Smith is a doctoral candidate at the Department of Communication Studies of the Paris Lodron University of Salzburg, Austria. She was previously a senior scientist without doctorate at the Institute for Comparative Media and Communication Studies of the Austrian Academy of Sciences and the University of Klagenfurt.

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 4, Pages 234–249
https://doi.org/10.17645/mac.v9i4.4184

Article

What's "Up Next"? Investigating Algorithmic Recommendations on YouTube Across Issues and Over Time

Ariadna Matamoros-Fernández [1,2,*], Joanne E. Gray [1,2], Louisa Bartolo [1,2], Jean Burgess [1,2], and Nicolas Suzor [1,3]

[1] Digital Media Research Centre, Queensland University of Technology, Australia
[2] School of Communication, Queensland University of Technology, Australia;
E-Mails: ariadna.matamorosfernandez@qut.edu.au (A.M.-F.), joanne.e.gray@qut.edu.au (J.E.G.), l.bartolo@qut.edu.au (L.B.), jean.burgess@qut.edu.au (J.B.)
[3] School of Law, Queensland University of Technology, Australia; E-Mail: n.suzor@qut.edu.au

* Corresponding author

Submitted: 13 February 2021 | Accepted: 26 April 2021 | Published: 18 November 2021

Abstract
YouTube's "up next" feature algorithmically selects, suggests, and displays videos to watch after the one that is currently playing. This feature has been criticized for limiting users' exposure to a range of diverse media content and information sources; meanwhile, YouTube has reported that they have implemented various technical and policy changes to address these concerns. However, there is little publicly available data to support either the existing concerns or YouTube's claims of having addressed them. Drawing on the idea of "platform observability," this article combines computational and qualitative methods to investigate the types of content that the algorithms underpinning YouTube's "up next" feature amplify over time, using three keyword search terms associated with sociocultural issues where concerns have been raised about YouTube's role: "coronavirus," "feminism," and "beauty." Over six weeks, we collected the videos (and their metadata, including channel IDs) that were highly ranked in the search results for each keyword, as well as the highly ranked recommendations associated with the videos. We repeated this exercise for three steps in the recommendation chain and then examined patterns in the recommended videos (and the channels that uploaded the videos) for each query and their variation over time. We found evidence of YouTube's stated efforts to boost "authoritative" media outlets, but at the same time, misleading and controversial content continues to be recommended. We also found that while algorithmic recommendations offer diversity in videos over time, there are clear "winners" at the channel level that are given a visibility boost in YouTube's "up next" feature. However, these impacts are attenuated differently depending on the nature of the issue.

Keywords
algorithms; automation; content moderation; digital methods; platform governance; YouTube

Issue
This article is part of the issue "Algorithmic Systems in the Digital Society" edited by Sanne Kruikemeier (University of Amsterdam, The Netherlands), Sophie Boerman (University of Amsterdam, The Netherlands) and Nadine Bol (Tilburg University, The Netherlands).

1. Introduction

YouTube is a dominant platform for news consumption, self-education, and opinion formation via video (Burgess & Green, 2018). A large proportion of YouTube content is suggested or automatically delivered to users via the platform's automated recommendation systems (Solsman, 2018), which have been criticized for amplifying misinformation, harmful content, and extreme views (Bergen, 2019; Roose, 2020). In particular, the platform's "up next" or "suggested videos" feature, which displays and automatically plays a sequence of videos following the one that is currently playing in the main window, has been criticized for leading people down

recommendation chains ("rabbit holes") of disturbing content (O'Callaghan et al., 2015), and contributing to political radicalization (Lewis, 2018).

The recommender system behind YouTube's "up next" feature has evolved over time and comprises multiple components including: the "related videos" algorithm (in use in various iterations for more than a decade; see Davidson et al., 2010); personalized "recommended videos" related to the user's watch history; and videos drawn from the same channel as the currently playing video. It typically prioritizes those videos that have been recently uploaded which have a high number of views and long average watch times, and it considers the popularity of a video by including viewer satisfaction measures such as likes and dislikes (Covington et al., 2016). Increasingly, the system relies on deep learning approaches to improve the "quality" of recommendations and to increase user engagement (Zhao et al., 2019).

A central concern regarding YouTube's "up next" feature is that, in service of the goal of increasing "engagement," it may tend to select videos that are highly evocative or provocative, including radical, alarmist, or otherwise extreme content (see for example a study by Mozilla Foundation and UC Berkeley scholars—Faddoul et al., 2020). By way of responding to these concerns, in 2019, the company announced it had made three significant "improvements" to its recommendation systems. First, they were updated to promote more diverse content by suggesting videos from a wider range of topics to avoid suggesting "too many similar recommendations, like seeing endless cookie videos after watching just one recipe for snickerdoodles" (The YouTube Team, 2019a). Second, changes were made so that "borderline content" would be demoted by the recommendation algorithms, so as to "reduce the spread of content that comes right up to [but does not cross] the line" of violating the platform's community guidelines (The YouTube Team, 2019b). Third, changes were made to increase the amplification of "authoritative voices" (The YouTube Team, 2019c), for example, for some breaking news events, YouTube's algorithms will prioritize videos published by trusted news outlets (The YouTube Team, 2019c).

In this article, we demonstrate a new method and generate new empirical evidence that contributes to public oversight of the operations of YouTube's suggested video feature, especially regarding potentially controversial sociocultural issues. We explore patterns in the recommendations made by YouTube's suggested videos feature over time for keyword search terms connected to sociocultural issues: "coronavirus," "feminism," and "beauty." By studying the algorithmic amplification of content connected to these terms we are able to provide empirical evidence for evaluating the claims made by critics and the counterclaims made by YouTube about the role of its "up next" feature in the amplification (or lack thereof) of problematic, authoritative, and diverse media content.

2. Social Media Recommender Systems, Exposure Diversity, and Platform Observability

Over the last decade, social media has emerged as important elements of a "hybrid media system" (Chadwick, 2017) that continues to reconfigure how information is created, distributed, and consumed. While there is technically more content available to audiences than ever before, in practice algorithms play a pivotal role in influencing users' exposure to a range of diverse media content and information sources, which is an important element of a media environment supportive of deliberative democracy (Glasser, 1984; Helberger, 2012; Horwitz, 2005; Napoli, 1999). Scholars and public commentators have argued that platforms' focus on maximizing "engagement" (giving users more of what they seem to want) can limit users' exposure to different points of view (Pariser, 2011; Sunstein, 2001), which in turn may lead to the hardening of extreme views, and political radicalization (Helberger, 2019). However, the supporting evidence is mixed. While some studies indicate that the algorithmic promotion of extremist and far-right content can lead users through recommendation chains of increasingly extreme content (e.g., Mozilla Foundation, 2020; Ribeiro et al., 2020), others suggest that actors exploit recommender systems by creating content to fill "voids," thereby gaining outsized attention for extreme content (Golebiewski & boyd, 2019, p. 29). Still others conclude that users encounter far-right content mostly through their own searches, indicating a level of pre-existing demand for extreme or radicalizing content, and that recommendation systems (including search engines) play a subsidiary role in its delivery (see e.g., Ledwich & Zaitsev, 2020). A number of scholars have suggested that excessive concern about algorithmic recommendation and associated personalization limiting users' exposure to diverse content may not be warranted (e.g., Haim et al., 2018; Möller et al., 2018), and more broadly, that additional work needs to be conducted to conceptualize standards for "diversity" in critiques of recommender systems' outputs (Loecherbach et al., 2020; Nechushtai & Lewis, 2019; Vrijenhoek et al., 2021). For example, questions of social diversity (i.e., the representation in both content and production of a range of class- and identity-based communities) are increasingly relevant to policy and practice, and platforms' use of "diversity" without definition (e.g., Zhao et al., 2019) illustrates the limits of corporate attempts to provide transparency relating to complex sociocultural and policy issues.

Despite these diverging views, there is consensus around two important aspects of platforms' recommender systems and how to hold them accountable. First, it is widely accepted that algorithms' opacity (Diakopoulos & Koliska, 2017)—or what Pasquale (2015) calls the "black box" of algorithmic decision making—makes it difficult to curtail platform power, which has motivated a growing body of empirical research

interested in studying algorithms *from the outside*. This includes methods such as "reverse engineering" (Diakopoulos, 2015), "scraping audits" (Sandvig et al., 2014), "everyday algorithm auditing" (Shen et al., 2021), small-scale observation (Bucher, 2012), and systematic large-scale observation (Rieder et al., 2018). Second, debates around how to hold the media accountable in general, and social media in particular, tend to focus on calls for greater transparency for regulatory inspection (Diakopoulos, 2016; Pasquale, 2015). However, the technical complexities of digital platforms pose unique challenges that complicate the effectiveness of transparency as a tool for generating knowledge about "what is hidden" (Ananny & Crawford, 2018; Rieder & Hofmann, 2020, p. 5).

"Observability" has been proposed as a path "to deal more systematically with the problem of studying complex algorithmic systems" (Rieder & Hofmann, 2020, p. 1). Whereas transparency invokes the idea of the algorithm as a mathematical formula which, if revealed for oversight, could lead to a better understanding of platforms' roles in the realization of media diversity, for example, observability as a tool for better regulation recognizes platform algorithms as complex socio-technical systems. The performance of platform algorithms that use deep learning models is influenced by multiple factors: developers' design choices, built-in randomness, business practices, content creators' optimization tactics, and audience viewing and engagement patterns. The idea of "algorithmic cultures" has been proposed to describe the variety of factors and agencies involved in generating algorithmic outcomes (McKelvey & Hunt, 2019; Rieder et al., 2018; Seyfert & Roberge, 2016) and to tackle the difficult task of assessing the social impacts of platformization. Rieder and Hofmann (2020, p. 22) advocate for "regulating for observability." They stress the need to observe platform behaviour over time and to institutionalize "processes of collective learning" to develop "the skills that are required to observe platforms" (p. 24).

This article aligns with the idea of "platform observability" and presents a method for observing and studying YouTube's recommendation "algorithmic cultures" over time. The following research questions inform our study: What kind of media does YouTube frequently recommend over time in relation to specific sociocultural topics? Are there patterns in these recommendations that can help answer longstanding questions about media diversity? Are there patterns in these recommendations that can improve our understanding of how YouTube operationalizes "media authority" in relation to different sociocultural issues? Drawing on Rieder et al.'s (2018) method for studying "ranking cultures" on YouTube—which they define as "unfolding processes of hierarchization and modulation of visibility that call on users, content creators and a platform that intervenes and circumscribes in various ways" (p. 52)—we use a combination of computational and qualitative methods to investigate the different factors that converge in producing recommendations in the "up next" section. We are attentive to the "moments of choice" that find their form in algorithmic operations (Rieder, 2017, p. 113); that is, since the platform can only show a limited number of videos in the "up next" interface (between 4 and 60), there is a complex process of selection that factors a wide range of features to provide "more quality information to users" (YouTube, n.d.). These processes of selection rely on sophisticated deep learning approaches that learn from user feedback to assess the "quality" of content (e.g., popularity and "freshness" of videos; see Covington et al., 2016), refine for personalization, improve the diversity of recommendations (Zhao et al., 2019), raise "authoritative voices," and demote "borderline content" (The YouTube Team, 2019c).

In this study, we were also interested in understanding how the platform's cultures of use influence YouTube's "up next" feature in practice and for different topics. We paid attention to "platform vernaculars": that is, the specific practices emerging from platforms' unique technical affordances and how users appropriate them in practice (Gibbs et al., 2015). In the case of YouTube, examples of platform vernaculars are users' tactics to gain algorithmic visibility: from word choices in titles and thumbnails, to other optimization techniques such as being an active content creator and building a network on the platform via featuring and subscribing to other channels (Bishop, 2019). Following Rieder et al.'s (2018, p. 54) suggestion that YouTube's ranking practices "cannot be easily detached from the specificities of concrete issues," we also consider the role of "issue vernaculars", by which we mean the ways that platform vernaculars are articulated and given form in the context of specific social, cultural, and political issues. For example, for topics such as Islam, highly active and controversial Islamophobic "niche entrepreneurs" gain exceptional levels of visibility on YouTube (Rieder et al., 2018, p. 64). Our focus on platform and issue vernaculars complements existing literature that has focused on platform design as a central requisite to facilitate or constrain exposure to media diversity (Helberger, 2011).

Our aim in selecting the topics Covid-19 ("coronavirus"), feminism ("feminism"), and beauty ("beauty") was to focus on contemporary issues of public concern that have been at the center of controversies on YouTube, and where YouTube recommendations potentially play a role in shaping public perception and understanding of these topics. "Coronavirus" was selected due to the relevance of Covid-19 as a news topic during the time period studied, one beset by misinformation which therefore might trigger YouTube's amplification of "authoritative sources." We selected "feminism" as a highly political and contentious issue on YouTube (Burgess & Matamoros-Fernández, 2016; Siddiqui, 2008), which might therefore provide indications of content and perspectival diversity. In contrast, "beauty" was selected as a contested issue that is less frequently the subject of

mainstream political discourse, and so could provide a comparison to topics more strongly associated with controversial political issues.

3. Methods

The methods we use in this article provide the basis for a crucial intervention in the space between technology press speculation and folk theories about algorithms on the one hand, and abstract critical theory on the other. To observe what the algorithms underpinning YouTube's "up next" section "do," we follow Rieder et al.'s (2018) approach to studying algorithmic outcomes through description instead of aiming at identifying " 'hard' moments of causality" (p. 53). Along with Rieder et al. (2018), we are inspired by Savage's (2009) idea of descriptive assemblage—"where processes of creativity, conceptual innovation, and observation can be used to mobilize novel insights" (p. 170)—and use it as a strategy to make sense of broader forms of agency involved in algorithmic power.

Our method provides two main vantage points from which to study algorithmic cultures in general, and recommendations on YouTube specifically. First, we consider time as a crucial aspect of "platform observability" and hence examine YouTube's "up next" feature over time to move away from the "snapshot logic" (Rieder & Hofmann, 2020, p. 7) underlying many studies on algorithmic accountability (see Airoldi et al., 2016; O'Callaghan et al., 2015; Ribeiro et al., 2020; Schmitt et al., 2018). Second, we study YouTube recommendations across different sociocultural topics because we consider that "good" recommendations can only be envisioned and operationalized in relation to specific issue domains (Rieder, 2020, p. 334). A significant limitation in our study, however, is that we are unable to account for the effects of user preferences in how YouTube suggests what videos to watch next.

3.1. Data Collection

We designed two separate gatherers to collect the data for this research. Gatherer 1 used the "search: list" endpoint of the YouTube API to collect recommended videos and their rankings for each of our three keywords. We set the "order" parameter to "upload date," which is one of the user-facing search settings in YouTube's website and mobile applications. Our rationale for selecting "upload date" as a ranking criterion responded to our interest in gathering videos from channels that were active in creating content during the period studied. We collected the 20 top results (globally) for each of our three queries from 7 March to 22 April 2020, once per day at approximately the same time each day. Our cut-off date represents the last date we were able to extract reliable data. During the study period, YouTube made changes to its API that prevented us from searching for newly uploaded videos in real-time.

Gatherer 2 was used to collect the recommendation chains (the sequence of suggested videos) that followed each of the videos gathered daily by Gatherer 1. Drawing on research suggesting that users pay more attention to items ranked at the top of lists (Jugovac & Jannach, 2017), we collected the top five recommendations for each video, going three levels deep (see Figure 1). This yielded a daily collection of up to 3,100 recommended videos per query (the sum of related videos collected every day for step 1, step 2, and step 3), and a total collection of up to 145,700 recommended videos (up to 3,100 recommended videos per 47 days of data collection). Gatherer 2 sent requests from an Australian-based IP address, without any identifying cookies. This means that we did not collect "recommended for you" videos, but we were able to receive localized suggestions in the "up next" feature. Our web-interface scraping method likely explains why English-language sources were so heavily present in our data and why Australian channels were recommended for queries.

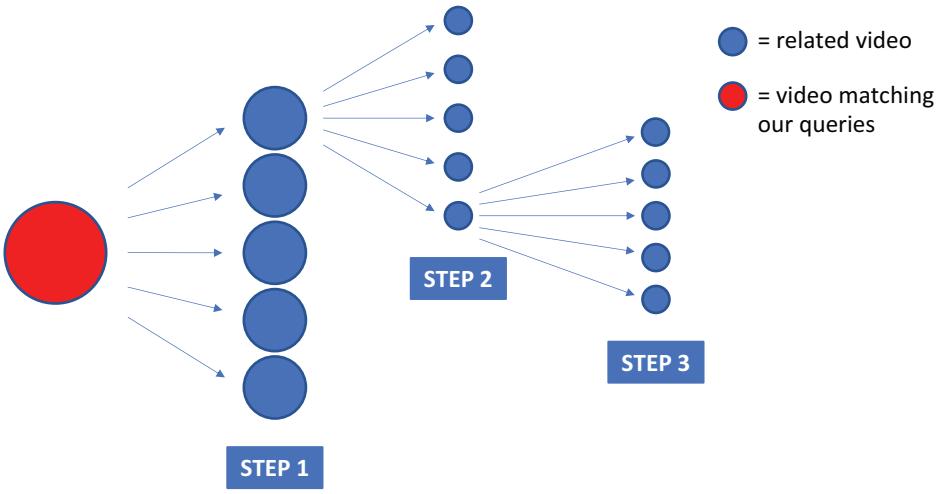


Figure 1. A schematic overview of our scraping method.

3.2. Data Analysis

For the data analysis, we combined data visualizations and qualitative analysis to identify patterns in how YouTube preferences certain content in the "up next" feature. Using the RankFlow tool (Rieder, 2016), we created flow diagrams that made changes in the top 10 most recommended videos and channels observable for analysis. All channels and videos included in the rankflows were recommended at least twice per day. Our aim with this approach was to understand the video and channel "winners" for each query; that is, those videos and channels most recommended over time and across steps in the chain. The visual inspection presented in Figures 2, 3, and 4 helps to identify patterns and operates as a starting point for in-depth investigation. The flow diagrams allow us to answer specific questions, such as whether certain unique channels are frequently recommended across days and steps. As seen in Figure 1, this is indeed the case (the flow diagrams exhibit a high number of blue/purple/red "waves" for the channel diagrams, which the RankFlow tool displays when it identifies that a unique video or channel appears on different days over time).

For the qualitative analysis, we looked for patterns in terms of media authority or popularity (proxies for quality), and we looked for patterns in perspectives that might give indications as to content diversity. To make sense of recommendation patterns for each of our topics, we were also attentive to platform and issue vernaculars. We took similar approaches for both channels and videos, but a greater emphasis was placed on channels in order to assess how "authoritativeness" is operationalized by YouTube's "up next" feature in relation to each of our queries. In our qualitative investigation, we also privileged patterns observed in step 1 of the recommendation chain. We considered those channels and videos surfaced in step 1 as the clear "winners" in terms of visibility—their position in the chain means they are most likely to be watched via autoplay or selected for play by a user.

For channels, we focused on the top 20 most recommended media sources over time and across steps for each of our queries (see Tables 1, 3, and 5). Since YouTube mentions "authoritative voices" in its policies but does not define the term, we looked at channel subscriber count, relevancy of the channel topic in relation to our queries, and frequency of upload at the time of our data collection as proxies for "media authority." For example, we considered subscriber count as an indicator for professionalization (see Rieder et al., 2020) and, hence, a metric potentially linked to a channel's authority, at least within YouTube's attention economy. We drew on YouTube's own "benefit levels" classification for channels to account for professionalization: "graphite" status (channels with less than 1,000 subscribers) gives content creators access to basic tools; surpassing the threshold of 1,000 subscribers, "opal" status, gives channels access to monetization through advertisements; "bronze" status (>10,000 subscribers) allows channels to access professional production tools; and "silver and up" (>100,000 subscribers) gives content creators the ability to have their own YouTube partner manager and receive Creator Awards (YouTube Creators, n.d.). To break down the rather broad "silver and up" tier, we added "gold" (>1,000,000–<10,000,000 subscribers) and "diamond" status (>100,000,000 subscribers) to YouTube's official channel classification system. In terms of media diversity signals, we considered channels' geographic regions and paid attention to questions of representation among the content creators most recommended for the "beauty" and "feminism" queries.

For videos, we focused on the top five most recommended videos over time and across steps for each of our queries (see Tables 2, 4, and 6) and we assessed their popularity, relevancy, and recency, through an analysis of their view counts, user engagement metrics, and upload dates, respectively. In terms of diversity of viewpoints, for "coronavirus," we focused on media frames (e.g., the use of militaristic language to describe the pandemic); for "feminism," we paid attention to whether the most recommended videos had a feminist or an anti-feminist stance; and for "beauty," drawing on the work of Bishop (2018), we were interested in examining how gendered and commercial logics influenced the content recommended for this query.

4. Findings

4.1. General Patterns Observed

Our visual analysis reveals two clear patterns in recommendations in the "up next" section over time, across queries, and across steps in the recommendation chain: (a) there is a higher level of variation in recommended videos than in recommended channels, (b) there is always more variation in suggested content at step 1 than at steps 2 and 3 (see Figures 2, 3, and 4). For each of our queries, there are clear "winners" at the channel level (media source) that are given a visibility boost by YouTube's recommendation algorithms—some channels are recommended repeatedly each day and consistently over time and down the chain (see Tables 2, 4, and 6).

4.2. Coronavirus

For "coronavirus," the platform prioritizes US news media outlets in the "up next" section as "authoritative" media in relation to Covid-19 (see Figure 2 and Table 1). Especially in step 1, while only 5.7% of videos (n = 16) were recommended on two or more days during the period studied, 49.4% of channels (n = 39) were recommended on more than one day. Mainstream news media channels falling within the "gold" or "diamond" tier dominate at each step, with NBC News the clear "winner" across the entire recommendation chain (see

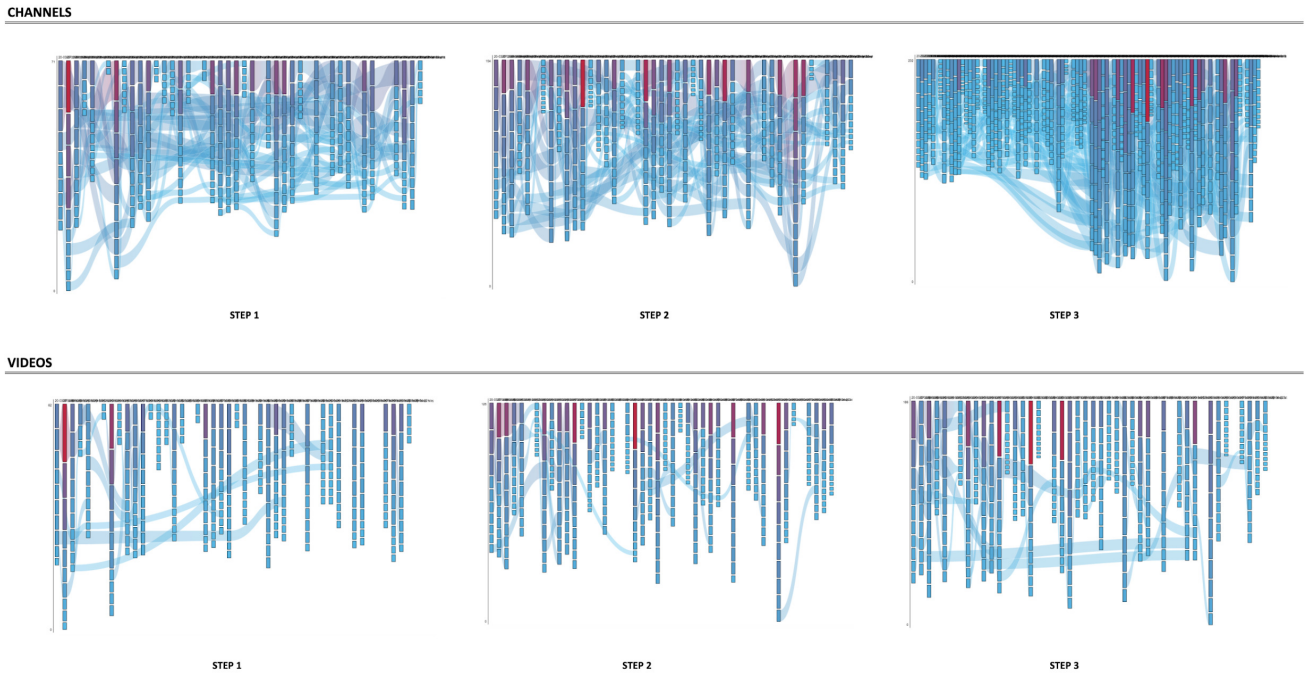


Figure 2. RankFlow visualization of "coronavirus" for the top 10 most frequently recommended channels at step 1, step 2, and step 3 (top), and for the top 10 most recommended videos at step 1, step 2, and step 3 (bottom). Notes: In the RankFlow charts, for all queries, columns represent days, and blocks individual videos or channels, ranked by number of recommendations, with the top-ranking video or channel on top of the column; color (blue to red) and bar height both indicate the number of times a channel or video was recommended on a single day; blue, purple, and red "waves" are created when the tool identifies that a particular video or channel appears on multiple days, that is, the tool creates a flow to trace the recurrence of unique videos or channels over time; morphologies with many "waves" (see channel rank flows, step 3) indicate the presence of certain videos and channels recurrently recommended over time.

Table 1). In terms of source diversity, US channels make up between 70% (n = 14) and 75% (n = 15) of the top 20 channels at each step, while UK channels make up between 10% (n = 2) and 15% (n = 3). Fox News, which has played an important role in spreading Covid-19 misinformation (Li et al., 2020), does progressively better as we go down the chain: ranking 17th at step 1, 10th at step 2, and 7th at step 3, being recommended over 6, 8, and 17 days, respectively.

The channels of health authorities such as the World Health Organization (WHO) and the Centers for Disease Control (CDC; both with "silver" status) were absent across all steps. This is despite the fact that on 2 April 2020, YouTube encouraged creators to base their coronavirus-related material on information from "reputable sources" such as WHO and CDC (YouTube Help, 2020) and that the WHO's channel was active over this time.

Turning to content diversity, a handful of videos from mainstream news channels "win" repeatedly at each step. These videos had millions of views (as of February 2021 when the analysis was undertaken), were all uploaded during our period of study, and most had charged, if not sensationalist titles. For example, in step 1, the most frequently recommended video came from Channel 4 News, with the title *Coronavirus Expert: 'War is an Appropriate Analogy'* (see Table 2). Although not featured in our "top five" list, a video uploaded by New Tang Dynasty—a problematic news channel published under the Epoch Media Group and accused of spreading misinformation (Zadrozny & Collins, 2019)—was recommended 12 times over two days at step 1, ranking sixth, with over 4 million views at the time of analysis. This video featured an interview with the discredited scientist behind the infamous "Plandemic" video, Judy Mikovits (Shepherd, 2020), in which she raised questions about the origin of Covid-19.

4.3. Feminism

For the keyword "feminism," we found that mainstream news media (Fox News; Channel 4 News), entertainment (The Late Late Show with James Corden), and educational channels (TEDx Talks and TED) falling predominantly within the "silver," "gold," and "diamond" tiers, were the clear "winners" across steps (see Table 3). As Figure 3 shows, YouTube offers more variance in how it recommends videos than it does channels for "feminism," but the difference is less pronounced than for "coronavirus." In step 1, only 11.9% of videos (n = 17) were recommended on two or more days during the period studied, whereas 23.8% of channels (n = 30) were recommended on more than one day.

Table 1. Top 20 channels recommended for "coronavirus" at each step.

Step 1	R	D	Step 2	R	D	Step 3	R	D
NBC News*	203	32	NBC News*	547	39	NBC News*	1030	39
Channel 4 News*	104	18	MSNBC*	234	26	MSNBC*	402	25
MSNBC*	89	18	Channel 4 News*	226	22	60 Minutes Australia*	299	19
BBC News*	70	20	CNN♦	149	16	Channel 4 News*	296	18
TODAY*	69	15	LastWeekTonight*	118	11	CNN♦	292	21
CNN♦	62	15	60 Minutes Australia*	112	15	LastWeekTonight*	237	11
DW News*	48	12	Global News*	107	14	Fox News*	212	17
60 Minutes Australia*	41	13	BBC News*	100	13	CNBC Television*	180	11
LastWeekTonight*	39	8	NewsNOW from FOX*	95	11	The Daily Show with Trevor Noah*	167	11
The Daily Show with Trevor Noah*	33	7	Fox News*	94	11	ABC News♦	136	10
ABC News♦	30	9	The Daily Show with Trevor Noah*	92	10	BBC News*	133	12
Global News*	29	10	TODAY*	91	12	Late Night with Seth Meyers*	132	11
NewsNOW from FOX*	28	6	DW News*	84	12	Global News*	128	10
Washington Post*	27	6	CNBC Television*	81	10	TODAY*	118	10
The Late Show with Stephen Colbert*	22	5	Fox Business*	65	8	NewsNOW from FOX*	111	9
The White House*	22	4	CBS News*	64	9	CBS News*	110	10
Fox News*	20	6	Late Night with Seth Meyers*	59	8	The Late Show with Stephen Colbert*	104	7
CBS News*	20	7	TIME*	51	7	CNBC*	83	6
Guardian News*	20	6	ABC News♦	49	8	Sky News*	80	5
CNBC Television*	19	6	Washington Post*	49	8	The White House*	73	7

Notes: Column "R" indicates the number of times the channel was recommended; column "D" indicates the number of days on which the channel was recommended more than once; symbols indicate channel subscription status—diamond symbol (♦) for "diamond" and asterisk (*) for "gold."

Regarding trends across the entire chain over time (see Table 3), TEDx Talks ("diamond" status) was the clear "winner" at each step, and PowerfulJRE (The Joe Rogan Experience podcast), which has courted controversy for being "a safe space to launder bad ideas" (Maiberg, 2018), was also prominent. Ranking 16th at step 1, *PowerfulJRE* ("diamond" status) goes on to rank second at both steps 2 and 3, outperforming mainstream news and entertainment channels. In terms of locales, US channels dominate, though to a lesser extent than they did for "coronavirus." At each step, US media sources make up between 50% (n = 10) and 60% (n = 12) of the top 20 channels. Indian channels also did well, making up between 15% (n = 3) and 30% (n = 6) of the top 20 channels at each step. This is likely related to a controversy involving Indian actress Neha Dhupia's comments about violence against women ("Neha Dhupia addresses," 2020). Among the Indian channels recommended across steps and over time (and which are distributed, roughly equally, across tiers ranging from "bronze" to "diamond"), those engaged in anti-feminist content (e.g., PeepOye Fame, Sahil Chhikara, Tanmay Bhat, and Bollywood Samachar) outperformed educational channels featuring videos on topics that include feminism (NPTEL-NOC IITM and UPSC Preparation). It is striking that none of the top 20 most recommended channels at each step self-describe as "feminist." The first self-described feminist channel to appear in our dataset is South-Korean 하말넘많 [heavytalker], which ranks 37th at step 2.

With regard to content diversity, looking at the top-recommended videos across steps, an anti-feminist trend was clear (see Table 4). Videos featuring the controversial public intellectual Jordan Peterson emerge as "winners." Following a Channel 4 News interview with Peterson which was recommended at step 1, a Joe Rogan interview with Peterson in which he criticizes the existence of "Women's Studies" departments at universities is the second most frequently recommended video at both steps 2 and 3, and a video of Peterson's 2018 book

Table 2. Top five videos recommended for "coronavirus" at each step.

Step	Video title	Channel	R	D
Step 1	Coronavirus Expert: "War is an Appropriate Analogy"	Channel 4 News	35	4
	Journalist Goes Undercover at "Wet Markets," Where the Coronavirus Started	60 Minutes Australia	31	8
	Coronavirus II: Last Week Tonight With John Oliver (HBO)	LastWeekTonight	17	3
	Coronavirus Disrupts Daily Life as Trump Declares National Emergency	TODAY	13	1
	Trump's Coronavirus Address, Blooper Reel Included	The Daily Show with Trevor Noah	12	3
Step 2	Coronavirus Expert: "War is an Appropriate Analogy"	Channel 4 News	67	4
	Watch CNBC's Full Interview With Berkshire Hathaway CEO Warren Buffett	CNBC Television	61	9
	Journalist Goes Undercover at "Wet Markets," Where the Coronavirus Started	60 Minutes Australia	61	7
	Trump's Coronavirus Address, Blooper Reel Included	The Daily Show with Trevor Noah	45	5
	Coronavirus: Last Week Tonight With John Oliver (HBO)	LastWeekTonight	44	5
Step 3	Journalist Goes Undercover at "Wet Markets," Where the Coronavirus Started	60 Minutes Australia	143	10
	Watch CNBC's Full Interview With Berkshire Hathaway CEO Warren Buffett	CNBC Television	134	10
	Coronavirus Wxpert: "War is an Appropriate Analogy"	Channel 4 News	96	5
	Coronavirus: Last Week Tonight With John Oliver (HBO)	LastWeekTonight	83	5
	Trump's Coronavirus Address, Blooper Reel Included	The Daily Show with Trevor Noah	59	5

Notes: Column "R" indicates the number of times the video was recommended; column "D" indicates the number of days on which the video was recommended more than once.

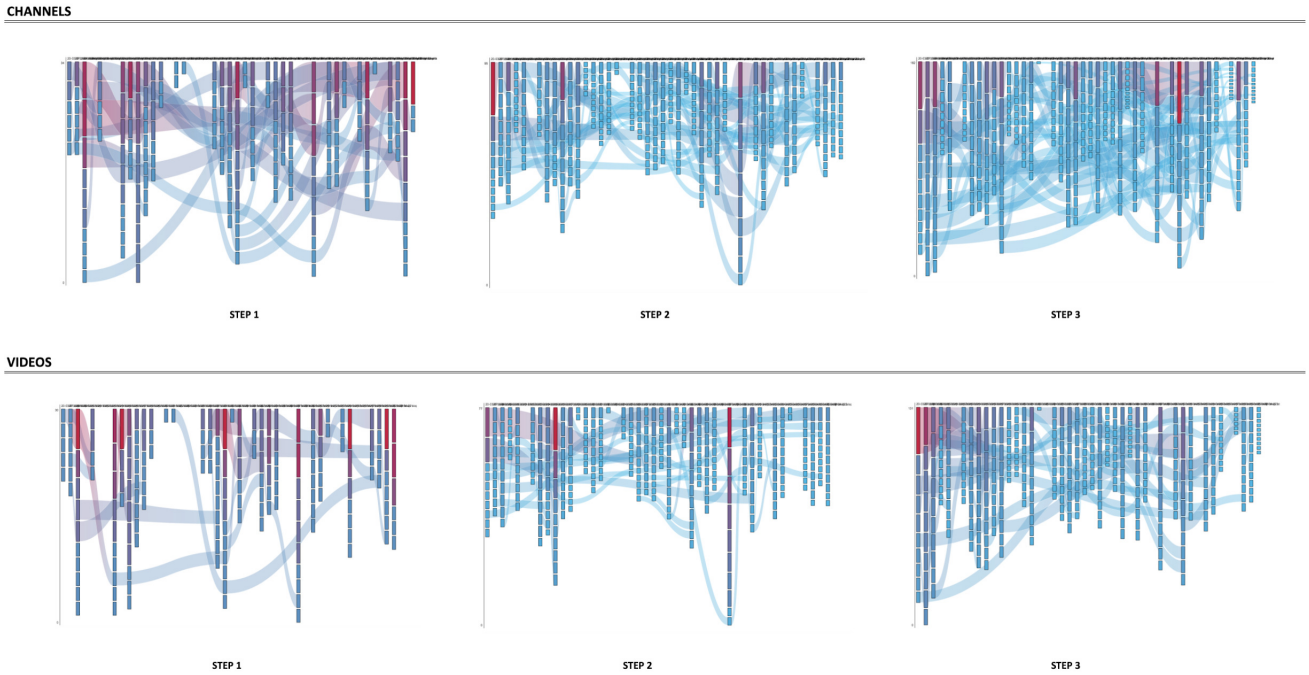


Figure 3. RankFlow visualization of "feminism" for the top 10 most recommended channels at step 1, step 2, and step 3 (top), and for the top 10 most recommended videos at step 1, step 2, and step 3 (bottom).

Table 3. Top 20 channels recommended for "feminism" at each step.

Step 1	R	D	Step 2	R	D	Step 3	R	D
TEDx Talks♦	76	22	TEDx Talks♦	198	27	TEDx Talks♦	696	30
The Late Late Show with James Corden♦	32	9	PowerfulJRE♦	56	14	PowerfulJRE♦	286	23
Fox News*	25	9	Channel 4 News*	48	13	SET India♦	116	7
TED♦	16	6	NBC News*	47	8	NBC News*	116	10
Channel 4 News*	16	6	PeepOye Fame♠	45	6	TED♦	111	12
PeepOye Fame♠	16	4	SET India♦	31	5	Fox News*	84	7
The University of Melbourne♣	13	4	TED♦	30	8	92nd Street Y♠	82	6
NewsNOW from FOX*	13	3	Tanmay Bhat*	28	4	Channel 4 News*	77	8
NPTEL-NOC IITM♠	9	3	Fox News*	26	5	Talent Replay*	76	5
NBC News*	9	3	TIME*	23	3	LastWeekTonight*	70	6
The White House*	9	2	How To Academy♠	22	5	Got Talent Global♦	64	5
Washington Post*	8	2	CBS News*	22	5	Top Viral Talent♦	62	5
Fox Business*	8	2	Fox Business*	22	4	PeepOye Fame♠	53	4
LastWeekTonight*	7	2	Bollywood Samachar*	19	4	Habertürk TV♠	52	5
SET India♦	7	2	LastWeekTonight*	16	3	TVO Docs♣	39	3
PowerfulJRE♦	7	2	Talent Replay*	15	2	How To Academy♠	38	5
UPSC Preparation♣	7	1	Sahil Chhikara♣	15	4	Top 10 Talent*	37	3
The Hill*	6	2	After Work Reactions♣	15	3	The Agenda with Steve Paikin♠	37	4
Ninja Nerd Science♠	6	3	Gauthali Entertainment♥	15	4	Tanmay Bhat*	34	3
Jordan B Peterson*	6	3	NPTEL-NOC IITM♠	14	3	After Work Reactions♣	33	3

Notes: Column "R" indicates the number of times the channel was recommended; column "D" indicates the number of days on which the channel was recommended more than once; symbols indicate channel subscription status—diamond symbol (♦) for "diamond," asterisk (*) for "gold," spades (♠) for "silver," clubs (♣) for "bronze," and hearts (♥) for "opal."

presentation is in the top five videos recommended at both steps 2 and 3. Beyond Peterson, there is a notable presence of videos from Indian channels that seem to mock or disparage Neha Dhupia (e.g., videos with titles such as *Destroying Pseudo Feminists Neha Dhupia and Nikhil Chinapa*).

Among the top five recommended videos for "feminism" at each step, only the videos related to the Neha Dhupia controversy were uploaded during our period of study. All remaining videos, some of which were unrelated to the topic of feminism, were uploaded to YouTube years beforehand. For example, a 2015 video on the health dangers of wireless radiation by Dr. Devra Davis (see Table 4), the appearance of which might be explained by activity around Covid-19 on YouTube at the time of data collection, including public discussions related to 5G conspiracy theories (Bruns et al., 2020).

4.4. Beauty

For "beauty," channels promoting DIY crafts and beauty hacks (5-Minute Craft, 5-Minute Crafts GIRLY, 123 GO!, and 5-Minute Crafts VS), falling within the "gold" and to a lesser extent "diamond" tiers, dominate the ranking of most recommended channels across steps (see Table 5). As observed for "coronavirus" and "feminism," as Figure 4 shows, we observed substantial content diversity, especially at step 1. While only 20% of videos (n = 17 videos) were recommended on two or more days in step 1 during the period studied, 31.4% of channels (n = 22) were recommended on more than one day.

Regarding trends across the entire chain over time (see Table 5), "5-minute" channels are the clear "winners" across steps, together accounting for between 40% (n = 8) and 45% (n = 9) of the top 20 channels at each step. These US-based channels pertain to TheSoul Publishing, an online publisher subject to claims of "gaming" YouTube's algorithm, including by uploading videos frequently and using clickbait strategies (Jennings, 2018). Similarly, Troom Troom, a channel of "mysterious international origin," which posts DIY/hack videos and takes an approach similar to that of the "5-minute" channels (Jennings, 2018), was recommended several times across different days in both step 2 and step 3.

In contrast to our findings for "coronavirus" and "feminism," mainstream news media channels are completely absent from our top 20 recommendations across steps for "beauty." Native-YouTube channels clearly dominate.

Table 4. Top five videos recommended for "feminism" at each step.

Step	Video title	Channel	R	D
Step 1	Pitch Perfect Riff-Off With Anna Kendrick & The Filharmonics	The Late Late Show with James Corden	22	7
	Meeting the Enemy: A Feminist Comes to Terms With the Men's Rights Movement | Cassie Jaye	TEDx Talks	19	8
	"The Truth About Mobile Phone and Wireless Radiation" —Dr. Devra Davis	The University of Melbourne	17	6
	Destroying Pseudo Feminists Neha Dhupia and Nikhil Chinapa (MTV Roadies Revolution) #AkasshReacts	Peepoye	16	4
	Jordan Peterson Debate on the Gender Pay Gap, Campus Protests and Postmodernism	Channel 4 News	13	5
Step 2	Destroying Pseudo Feminists Neha Dhupia and Nikhil Chinapa (MTV Roadies Revolution) #AkasshReacts	Peepoye	37	5
	Joe Rogan Experience #877 With Jordan Peterson	PowerfulJRE	25	9
	This Title is Her Choice—Roadies Cringe Mahotsav	Tanmay Bhat	24	4
	Jordan B. Peterson on 12 Rules for Life	How To Academy	24	6
	Bollywood Angry Reaction on Neha Dhupia Roadies Controversy@Bollywood Samachar	Bollywood Samachar	17	4
	Top 10 Funny Performances Got Talent	Talent Replay	98	8
Step 3	Joe Rogan Experience #877 With Jordan Peterson	PowerfulJRE	86	12
	Jordan B. Peterson on 12 Rules for Life	How To Academy	58	9
	How to Learn Any Language in Six Months | Chris Lonsdale	TEDx Talks	46	5
	The Mathematics of Weight Loss | Ruben Meerman | TEDxQUT (Edited Version)	TEDx Talks	45	4

Notes: Column "R" indicates the number of times the video was recommended; column "D" indicates the number of days on which the video was recommended more than once.

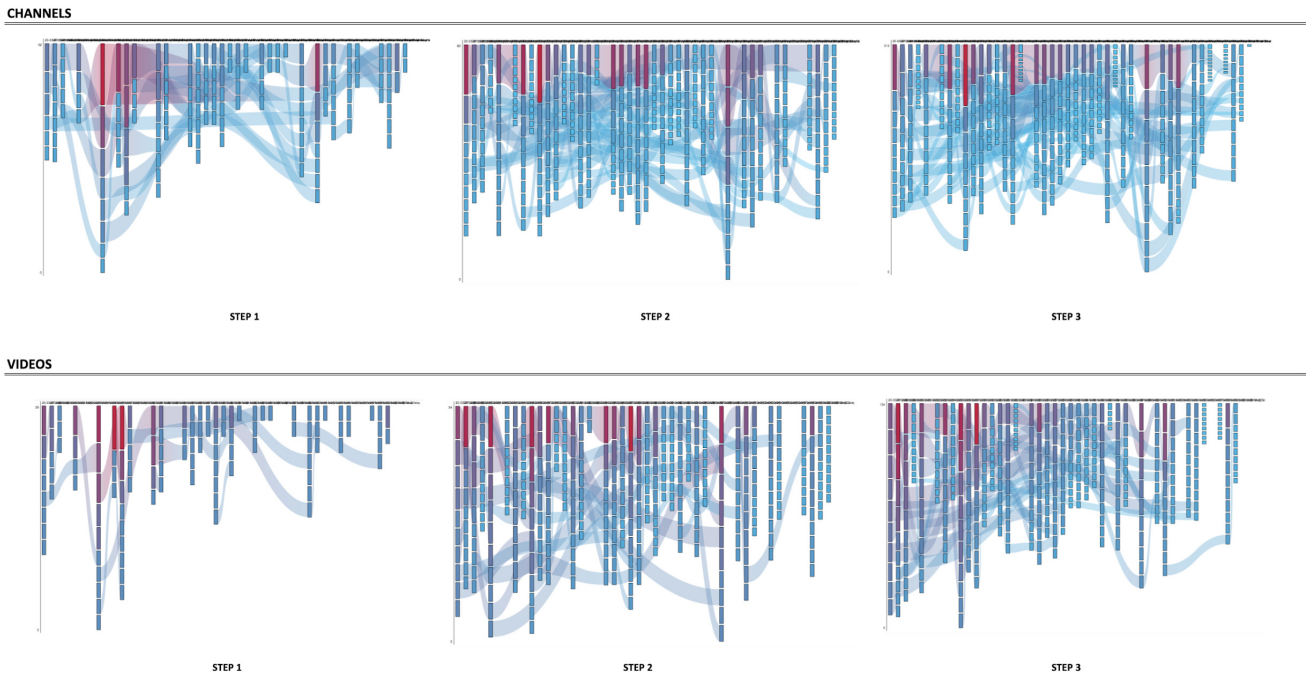


Figure 4. RankFlow visualization of "beauty" for the top 10 most recommended channels at step 1, step 2, and step 3 (top), and for the top 10 most recommended videos at step 1, step 2, and step 3 (bottom).

Table 5. Top 20 channels recommended for "beauty" at each step.

Step 1	R	D	Step 2	R	D	Step 3	R	D
5-Minute Crafts♦	79	23	5-Minute Crafts♦	316	33	5-Minute Crafts♦	923	33
5-Minute Crafts♦ GIRLY	39	15	5-Minute Crafts♦ GIRLY	87	17	5-Minute Crafts♦ GIRLY	188	15
Vogue*	21	6	5-Minute Crafts VS*	66	14	Vogue*	181	15
123 GO! *	19	5	Vogue*	48	8	5-MINUTEN-TRICKS*	130	8
5-Minute Crafts VS*	17	7	123 GO! *	43	7	5-Minute Crafts VS*	123	12
5-Minute Crafts FAMILY♦	17	8	5-Minute Crafts FAMILY♦	41	10	Dawn Gallagher♣	106	7
123 GO Like! *	12	5	123 GO! SCHOOL*	40	7	123 GO! SCHOOL*	93	8
Beauty Lady	11	5	5-MINUTEN-TRICKS*	35	7	Troom Troom♦	90	6
TIK TOK N SANAM♣	8	4	Dawn Gallagher♣	29	4	5-Minute Crafts FAMILY♦	75	7
Crafty Panda♦	7	3	123 GO Like! *	29	6	Jen Phelps♣	58	4
Dawn Gallagher♣	6	2	Troom Troom♦	26	3	123 GO Like! *	56	5
5-Minute Crafts PLAY♦	6	2	Crafty Panda♦	16	3	123 GO! *	50	5
5-Minute Crafts Tech*	6	3	DIY Unique Ideas	14	3	DIY Unique Ideas	43	4
123 GO! SCHOOL*	6	3	Kim Lianne*	13	3	Kim Lianne*	41	3
Crazy Shayna	6	3	Crazy Shayna	13	3	Jen Luvs Reviews♠	41	3
DIY Unique Ideas	5	2	Beauty's Big Sister♣	12	3	Allie Glines♠	39	3
Dominique Sachse*	4	2	Ali Andreea♠	11	3	MarionCameleon♠	37	3
Kim Lianne*	4	2	Jessica Braun♠	11	2	TEDx Talks♦	32	3
RosyMcMichael*	4	2	Zachary Michael♠	11	2	Kelly Strack♠	32	3
Cassandra Bankson*	4	2	Yasmina Filali♣	10	2	Julia Mazzucato♣	32	2

Notes: Column "R" indicates the number of times the channel was recommended; column "D" indicates the number of days on which the channel was recommended more than once; symbols indicate channel subscription status—diamond symbol (♦) for "diamond," asterisk (*) for "gold," spades (♠) for "silver," and clubs (♣) for "bronze"; for channels without specified subscription figures, rows were left with no symbol.

There is a strong commercial element to many of the top-recommended channels for "beauty," for example, many YouTubers test and review products. US-based channels continue to feature prominently: between 60% (n = 12) and 75% (n = 15) of the top 20 channels at each step are US-based. Regarding popularity, between 75% (n = 15) and 80% (n = 16) of the top 20 channels at each step fall within the "silver," "gold," or "diamond" tiers.

Turning to videos, a video from 5-Minute Crafts, offering up "cooking tricks," is the clear recommendation winner across steps and, in general, TheSoul Publishing's "5-Minute' channels" videos occurred most frequently (see Table 6). The only other videos in the top five were uploaded by Vogue and American beauty expert Dawn Gallagher. Notably, all the recurring videos from TheSoul Publishing's "5-minute" channels used fully capitalized letters (e.g., "33 GIRLY HACKS YOU DIDN'T KNOW BEFORE"), exemplifying the use of clickbait tactics to "game" the YouTube algorithm (Jennings, 2018). In terms of a diversity of viewpoints and representation, the top-recommended videos offer a commercialized and gendered representation of beauty and a limited representation of people of color.

5. Discussion

Our investigation shows significant variation in recommended videos (content diversity) over time and across queries, especially at step 1. This finding aligns with the company's longstanding commitment to "diversification" in the "up next" section (Davidson et al., 2010; The YouTube Team, 2019a). Yet, YouTube's operationalization of media diversity deserves further attention. First, we found that the platform clearly prioritizes certain channels (source diversity) over time and across steps, which provided important insights into how YouTube operationalizes "authoritativeness" in practice. US channels dominated across queries, down the chains, and over time, which highlights the cultural dominance of the US on YouTube (Rieder et al., 2020). From our data, it also seems clear that YouTube, following political and social pressure, makes decisions to categorize certain topics societally significant and truth-oriented enough for heavy-headed platform intervention (e.g., vaccination, climate change, elections), while others (e.g., gender, politics, and beauty) are considered less so. For "coronavirus," for example, YouTube amplified US news partners in the chain. Users' location

Table 6. Top five videos recommended for "beauty" at each step.

Step	Video title	Channel	R	D
Step 1	100 Cooking Tricks That Will Help You to Cut Costs || Live	5-Minute Crafts	16	6
	Hilary Duff's Busy Mom Makeup Routine | Beauty Secrets | Vogue	Vogue	14	4
	33 Girly Hacks You Didn't Know Before	5-Minute Crafts VS	10	4
	17 Genius Ideas for Girls || Hair and Makeup Transformations	5-Minute Crafts	9	2
	31 Colorful Hair Hacks for a Flawless Look	5-Minute Crafts GIRLY	9	4
Step 2	100 Cooking Tricks That Will Help You to Cut Costs || Live	5-Minute Crafts	90	19
	101 Easy Yet Genius Kitchen Hacks You've Never Seen Before	5-Minute Crafts	49	10
	Makeup Techniques for Women Over 40! Dawn and Joseph	Dawn Gallagher	29	5
	100 Best Cooking Hacks Live	5-Minute Crafts	23	8
	33 Girly Hacks You Didn't Know Before	5-Minute Crafts VS	21	5
Step 3	100 Cooking Tricks That Will Help You to Cut Costs || Live	5-Minute Crafts	234	22
	100 Best Cooking Hacks Live	5-Minute Crafts	164	18
	Makeup Techniques for Women Over 40! Dawn and Joseph	Dawn Gallagher	105	8
	All-Time Best Life Hacks Everyone Should Know	5-Minute Crafts	93	12
	100 Best Kitchen Tips || Cooking Hacks, Easy Recipes and Yummy Ideas	5-Minute Crafts	79	8

Notes: Column "R" indicates the number of times the video was recommended; column "D" indicates the number of days on which the video was recommended more than once.

information, though, influences the news channels surfaced by YouTube, as the appearance of Australian news channels (e.g., 60 Minutes Australia) in our data demonstrates. This is in line with YouTube's announcement that it was surfacing local trusted news outlets for newsworthy events (Mohan & Kyncl, 2018). For "feminism" and "beauty," in contrast, YouTube-native anti-feminist content creators (e.g., PeepOye Fame; Sahil Chhikara), and "5-Minute" channels, respectively, dominated the "up next" section over time and across steps, raising the question of how easily channels operated by "entrepreneurs" and powerful publishing companies can become "authoritative voices" on topics with clear consumer and niche markets.

When moderating at the level of the channel, YouTube has had some success: YouTube-native niche entrepreneurs promoting conspiracy theories have reportedly experienced a drop in views since the platform updated its systems to demote borderline content (Thomson, 2020). However, our findings reveal the problems associated with prioritizing content from news partners (and from users that self-certify as verified accounts) as an approach to operationalizing the promotion of authoritative content (Caplan, 2020), especially when channels such as Fox News have been known to circulate misinformation, channels including PowerfulJRE are known for laundering "bad ideas" (Maiberg, 2018) and 5-minutes Crafts channels engage in clickbait practices (Jennings, 2018).

Second, while YouTube might be committed to offering video diversity in the "up next" section, we found that the videos most recommended for each of our queries did not feature a breadth of genres, viewpoints, or framings. For "beauty," YouTube's "up next" section favored channels that upload highly stereotyped, commercialized, and gendered content, and for "feminism" it prioritized channels run by male YouTubers with strong anti-feminist views. We consider these findings to indicate YouTube has not effectively addressed content diversity from a social perspective (failing to attend to factors such as race, gender, nationality, sexuality, and ability).

Our findings also indicate a clear correlation between frequently recommended videos and channels, and popularity and "freshness" (proxies for "quality"). All the channels that were "winners" in the recommendation chain across queries fell predominantly within the "silver,'' "gold," and "diamond" tiers, which means that these media sources are part of an "elite" group representing less than 1% of all YouTube channels (Rieder et al., 2020). For videos, almost all of the most frequently recommended videos had accrued millions of views. The recency signal was also evident in our data: The most frequently recommended channels were uploading videos regularly during the period of our data collection, and most frequently recommended videos were often recently uploaded. However, we also found older "viral" videos repeatedly recommended, some of which were potentially problematic in terms of misinformation, especially for "coronavirus" and "feminism."

Recency and popularity alone, though, are insufficient to explain why certain problematic videos and less popular channels appear in the "up next" recommendation chain. Platform and issue vernaculars also play a part. Content creators are increasingly aware of the importance of gaming social media algorithms to boost visibility (Bishop, 2019), and they implement and test various optimization tactics—e.g., use of relevant keywords in headlines—to increase their chances of being amplified by YouTube's recommendations systems, which was visible in both "feminism" and "beauty." Optimization tactics might explain the appearance of some channels with "opal" and "bronze"

status—so-called micro-influencers (Boerman, 2020)—within the top 20 most recommended channels for "beauty," such as former model Dawn Gallagher and beauty and fashion YouTuber Julia Mazzucato.

For "feminism," audience viewing patterns and the "data void" problem (Golebiewski & boyd, 2019) might explain the overrepresentation of anti-feminist YouTube content creators around discussions of "feminism" (Döring & Mohseni, 2019). Arguably, YouTube has a much richer repository of content in the case of "coronavirus" and "beauty" than it does for "feminism," which might result in recommendations of less popular and/or less relevant content for that search term. Data voids are especially concerning when they have been successfully exploited by actors pushing problematic agendas, such as those that are anti-feminist or misogynistic. Although YouTube was alerted to this issue in 2015 (Golebiewski & boyd, 2019, p. 29), and despite highly popular feminist YouTubers being active on the platform (e.g., Jouelzy, Feminist Frequency), our study indicates that five years later, it is still a problem.

Last, our analysis shows that the algorithms underpinning the "up next" feature, as with ranking, are sensitive to newsworthy events and controversies (Rieder et al., 2018, p. 63). This was visible in the "feminism" data where India-based channels that had uploaded new content to YouTube were recommended at high rates after a gender-based controversy relating to actress Neha Dhuphia. Sensitivity to current events shows the importance of studying YouTube's "related videos" algorithm over time and as part of the broader media system in which YouTube exists, where different internal and external agencies converge to influence how the platform recommends content to users.

6. Conclusion

This article has provided new evidence about what the algorithms underpinning YouTube's "up next" feature "do" over time, down the recommendation chain, and in relation to specific issue spaces. We paid attention to YouTube's "moments of choice" that find their form in algorithmic processes (e.g., a commitment to content diversity and to the promotion of authoritative voices) and the impact of platform and issue vernaculars on what content gets surfaced in the "up next" section, albeit without fully accounting for the effects of personalization. Critically, we have also shown how corporate understandings of "diversity," "quality," and "authoritativeness," and their operationalization in practice, can have significant limitations in terms of improving the types of content that are amplified by automated recommendations systems and, potentially, the type of information users are exposed to in relation to certain issue domains. For future research, our approach to "platform observability" (Rieder & Hofmann, 2020, p. 21) might be usefully combined with studies that build on issue comparisons while also accounting for personalization.

Acknowledgments

The authors would like to thank Guangnan Zhu for his work in preparing the RankFlow files for our visualizations, and Bernhard Rieder for his useful comments on earlier versions of this article. Funding for this project was received from the QUT Digital Media Research Centre and the Australian Research Council Centre of Excellence for Automated Decision-Making and Society.

Conflict of Interests

The authors declare no conflict of interests.

Supplementary Material

Supplementary material for this article is available online in the format provided by the author (unedited).

References

Airoldi, M., Beraldo, D., & Gandini, A. (2016). Follow the algorithm: An exploratory investigation of music on YouTube. *Poetics*, *57*, 1–13. https://doi.org/10.1016/j.poetic.2016.05.001

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Bergen, M. (2019, April 2). YouTube executives ignored warnings, letting toxic videos run rampant. *Bloomberg*. https://www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant

Bishop, S. (2018). Anxiety, panic, and self-optimization: Inequalities and the YouTube algorithm. *Convergence*, *24*(1), 69–84. https://doi.org/10.1177/1354856517736978

Bishop, S. (2019). Managing visibility on YouTube through algorithmic gossip. *New Media & Society*, *21*(11/12), 2589–2606. https://doi.org/10.1177/1461444819854731

Boerman, S. C. (2020). The effects of the standardized Instagram disclosure for micro- and meso-influencers. *Computers in Human Behavior*, *103*, 199–207. https://doi.org/10.1016/j.chb.2019.09.015

Bruns, A., Harrington, S., & Hurcombe, E. (2020). "Corona? 5G? or both?": The dynamics of Covid-19/5G conspiracy theories on Facebook. *Media International Australia*, *177*(1), 12–29. https://doi.org/10.1177/1329878X20946113

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, *14*(7), 1164–1180. https://doi.org/10.1177/1461444812440159

Burgess, J., & Green, J. (2018). *YouTube: Online video and*

participatory culture (2nd ed.). Polity Press.

Burgess, J., & Matamoros-Fernández, A. (2016). Mapping sociocultural controversies across digital media platforms: One week of #gamergate on Twitter, YouTube, and Tumblr. *Communication Research and Practice*, *2*(1), 79–96. https://doi.org/10.1080/22041451.2016.1155338

Caplan, R. (2020, December 18). Pornhub is just the latest example of the move toward a verified internet. *Slate Magazine*. https://slate.com/technology/2020/12/pornhub-verified-users-twitter.html

Chadwick, A. (2017). *The hybrid media system: Politics and power*. Oxford University Press.

Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. In P. Covington, J. Adams, & E. Sargin (Eds.), *Proceedings of the 10th ACM conference on recommender systems* (pp. 191–198). ACM. https://doi.org/10.1145/2959100.2959190

Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., & Sampath, D. (2010). The YouTube video recommendation system. In J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingstone, & D. Sampath (Eds.), *Proceedings of the fourth ACM Conference on recommender systems* (pp. 293–296). ACM. https://doi.org/10.1145/1864708.1864770

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. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, *59*(2), 56–62. https://doi.org/10.1145/2844110

Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media, *Digital Journalism*, *5*(7), 809–828. https://doi.org/10.1080/21670811.2016.1208053

Döring, N., & Mohseni, M. R. (2019). Male dominance and sexism on YouTube: Results of three content analyses. *Feminist Media Studies*, *19*(4), 512–524. https://doi.org/10.1080/14680777.2018.1467945

Faddoul, M., Chaslot, G., & Farid, H. (2020). *A longitudinal analysis of YouTube's promotion of conspiracy videos*. ArXiv. http://arxiv.org/abs/2003.03318

Gibbs, M., Meese, J., Arnold, M., Nansen, B., & Carter, M. (2015). #Funeral and Instagram: Death, social media, and platform vernacular. *Information, Communication & Society*, *18*(3), 255–268. https://doi.org/10.1080/1369118X.2014.987152

Glasser, T. L. (1984). Competition and diversity among radio formats: Legal and structural issues. *Journal of Broadcasting*, *28*, 127–142.

Golebiewski, M., & boyd, D. (2019). *Data voids: Where missing data can easily be exploited*. Data & Society. https://datasociety.net/wp-content/uploads/2019/11/Data-Voids-2.0-Final.pdf

Haim, M., Graefe, A., & Brosius, H. (2018). Burst of the filter bubble? *Digital Journalism*, *6*(3), 330–343. https://doi.org/10.1080/21670811.2017.1338145

Helberger, N. (2011). Diversity by design. *Journal of Information Policy*, *1*, 441–469. https://doi.org/10.5325/jinfopoli.1.2011.0441

Helberger, N. (2012). Exposure diversity as a policy goal. *Journal of Media Law*, *4*(1), 65–92. https://doi.org/10.5235/175776312802483880

Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, *7*(8), 993–1012. https://doi.org/10.1080/21670811.2019.1623700

Horwitz, R. B. (2005). On media concentration and the diversity question. *Information Society*, *21*(3), 181–204.

Jennings, R. (2018, November 12). YouTube is full of cringey, clickbait DIY channels: They're even weirder than you think. *Vox*. https://www.vox.com/the-goods/2018/11/12/18065662/troom-troom-5-minute-crafts-youtube-diy-prank

Jugovac, M., & Jannach, D. (2017). Interacting with recommenders: Overview and research directions. *ACM Transactions on Interactive Intelligent Systems*, *7*(3). https://doi.org/10.1145/3001837

Ledwich, M., & Zaitsev, A. (2020). Algorithmic extremism: Examining YouTube's rabbit hole of radicalization. *First Monday*, *25*(3). https://doi.org/10.5210/fm.v25i3.10419

Lewis, R. (2018). *Alternative influence: Broadcasting the reactionary right on YouTube*. Data & Society. https://datasociety.net/wp-content/uploads/2018/09/DS_Alternative_Influence.pdf

Li, H. O.-Y., Bailey, A., Huynh, D., & Chan, J. (2020). YouTube as a source of information on Covid-19: A pandemic of misinformation? *BMJ Global Health*, *5*(5). https://doi.org/10.1136/bmjgh-2020-002604

Loecherbach, F., Moeller, J., Trilling, D., & van Atteveldt, W. (2020). The unified framework of media diversity: A systematic literature review. *Digital Journalism*, *8*(5), 605–642. https://doi.org/10.1080/21670811.2020.1764374

Maiberg, E. (2018, September 8). The Joe Rogan Experience is a safe space to launder bad ideas. *VICE*. https://www.vice.com/en/article/9kv9qd/the-joe-rogan-experience-is-a-safe-space-to-launder-bad-ideas

McKelvey, F., & Hunt, R. (2019). Discoverability: Toward a definition of content discovery through platforms. *Social Media + Society*, *2019*. https://doi.org/10.1177/2056305118819188

Mohan, N., & Kyncl, R. (2018, July 9). Building a better news experience on YouTube, together. *Youtube Official Blog*. https://blog.youtube/news-and-events/building-better-news-experience-on

Möller, J., Trilling, D., Helberger, N., & van Es, B. (2018). Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society*, *21*(7), 959–977. https://doi.org/

10.1080/1369118X.2018.1444076
Mozilla Foundation. (2020). *YouTube regrets*. https://foundation.mozilla.org/en/campaigns/youtube-regrets
Napoli, P. M. (1999). Deconstructing the diversity principle. *Journal of Communication*, *49*(4), 7–34. https://doi.org/10.1111/j.1460-2466.1999.tb02815.x
Nechushtai, E., & Lewis, S. C. (2019). What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations, *Computers in Human Behavior*, *90*, 298–307. https://doi.org/10.1016/j.chb.2018.07.043
Neha Dhupia addresses Roadies controversy after "weeks of vitriol": "My dad's WhatsApp is flooded with abuses." (2020, March 15). *Hindustan Times*. https://www.hindustantimes.com/tv/neha-dhupia-addresses-roadies-controversy-after-weeks-of-vitriol-my-dad-s-whatsapp-is-flooded-with-abuses/story-e3yg0d540ZEo8oCFKsQCDL.html
O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2015). Down the (white) rabbit hole: The extreme right and online recommender systems. *Social Science Computer Review*, *33*(4), 459–478. https://doi.org/10.1177/0894439314555329
Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin Books.
Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira, W. (2020). Auditing radicalization pathways on YouTube. In M. H. Ribeiro, R. Ottoni, R. West, V. A. F. Almeida, W. Meira (Eds.), *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 131–141). ACM.
Rieder, B. (2016). RankFlow [Computer software]. PolSys. http://labs.polsys.net/tools/rankflow
Rieder, B. (2017). Scrutinizing an algorithmic technique: The Bayes classifier as interested reading of reality. *Information, Communication & Society*, *20*(1), 100–117. https://doi.org/10.1080/1369118X.2016.1181195
Rieder, B. (2020). *Engines of order: A mechanology of algorithmic techniques*. Amsterdam University Press.
Rieder, B., Coromina, Ò., & Matamoros-Fernández, A. (2020). Mapping YouTube: A quantitative exploration of a platformed media system. *First Monday*, *25*(8). https://doi.org/10.5210/fm.v25i8.10667
Rieder, B., & Hofmann, J. (2020). Towards platform observability. *Internet Policy Review*, *9*(4). https://doi.org/10.14763/2020.4.1535
Rieder, B., Matamoros-Fernández, A., & Coromina, Ò. (2018). From ranking algorithms to "ranking cultures": Investigating the modulation of visibility in YouTube search results. *Convergence*, *24*(1), 50–68. https://doi.org/10.1177/1354856517736982
Roose, K. (2020, October 24). How the epoch times created a giant influence machine. *The New York Times*. https://www.nytimes.com/2020/10/24/technology/epoch-times-influence-falun-gong.html
Sandvig, C., Hamilton, K., & Karahalios, K. (2014). *Auditing algorithms: Research methods for detecting discrimination on internet platforms* [Conference paper]. Data and Discrimination: Converting Critical Concerns into Productive Inquiry. Seattle, WA, US.
Savage, M. (2009). Contemporary sociology and the challenge of descriptive assemblage. *European Journal of Social Theory*, *12*(1), 155–174. https://doi.org/10.1177/1368431008099650
Schmitt, J. B., Rieger, D., Rutkowski, O., & Ernst, J. (2018). Counter-messages as prevention or promotion of extremism?! The potential role of YouTube recommendation algorithms. *Journal of Communication*, *68*(4), 780–808. https://doi.org/10.1093/joc/jqy029
Seyfert, R., & Roberge, J. (2016). *Algorithmic cultures: Essays on meaning, performance and new technologies*. Routledge.
Shen, H., DeVos, A., Eslami, M., & Holstein, K. (2021). Everyday algorithm auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), Article 433. https://doi.org/10.1145/3479577
Shepherd, K. (2020, May 9). Who is Judy Mikovits in "Plandemic," the coronavirus conspiracy video just banned from social media? *The Washington Post*. https://www.washingtonpost.com/gdpr-consent/?next_url=https%3a%2f%2fwww.washingtonpost.com%2fnation%2f2020%2f05%2f08%2fplandemic-judy-mikovits-coronavirus%2f
Siddiqui, S. (2008). YouTube and feminism: A class action project. *Feminist Collections: A Quarterly of Women's Studies Resources*, *29*(1), 24–25.
Solsman, J. E. (2018, January 10). YouTube's AI is the puppet master over most of what you watch. *CNET*. https://www.cnet.com/news/youtube-ces-2018-neal-mohan
Sunstein, C. R. (2001). *Republic.com*. Princeton University Press. vThe YouTube Team. (2019a, January 25). Continuing our work to improve recommendations on YouTube. *YouTube Official Blog*. https://blog.youtube/news-and-events/continuing-our-work-to-improve
The YouTube Team. (2019b, June 5). Our ongoing work to tackle hate. *YouTube Official Blog*. https://blog.youtube/news-and-events/our-ongoing-work-to-tackle-hate
The YouTube Team. (2019c, December 3). The four Rs of responsibility, part 2: Raising authoritative content and reducing borderline content and harmful misinformation. *YouTube Official Blog*. https://blog.youtube/inside-youtube/the-four-rs-of-responsibility-raise-and-reduce
Thomson, C. (2020, September 18). YouTube's plot to

silence conspiracy theories. *Wired*. https://www.wired.com/story/youtube-algorithm-silence-conspiracy-theories

Vrijenhoek, S., Kaya, M., Metoui, N., Möller, J., Odijk, D., & Helberger, N. (2021). Recommenders with a mission: Assessing diversity in news recommendations. In F. Scholer & P. Thomas (Eds.), *CHIIR '21: Proceedings of the 2021 conference on human information interaction and retrieval* (pp. 173–183). ACM. https://doi.org/10.1145/3406522.3446019

YouTube. (n.d.). *How does YouTube provide more quality information to users?* https://www.youtube.com/intl/ALL_au/howyoutubeworks/our-commitments/fighting-misinformation/#raising-quality-info

YouTube Creators. (n.d.). *Awards*. https://www.youtube.com/creators/awards

YouTube Help. (2020). *Coronavirus disease 2019 (Covid-19) updates*. https://support.google.com/youtube/answer/9777243?hl=en

Zadrozny, B., & Collins, B. (2019, August 20). Trump, QAnon and an impending judgment day: Behind the Facebook-fueled rise of The Epoch Times. *NBC News*. https://www.nbcnews.com/tech/tech-news/trump-qanon-impending-judgment-day-behind-facebook-fueled-rise-epoch-n1044121

Zhao, Z., Hong, L., Wei, L., Chen, J., Nath, A., Andrews, S., & Chi, E. (2019). Recommending what video to watch next: A multitask ranking system. In Z. Zhoo, L. Hong, L. Wei, J. Chen, A. Nath, S. Andrews, A. Kumthekar, M. Sathiamoorthy, X. Yi, & E. H. Chi (Eds.), *Proceedings of the 13th ACM conference on recommender systems* (pp. 43–51). ACM. https://doi.org/10.1145/3298689.3346997

About the Authors

Ariadna Matamoros-Fernández is a lecturer in digital media in the School of Communication at the Queensland University of Technology, chief investigator at the Digital Media Research Centre, and associate investigator at the national ARC Centre of Excellence for Automated Decision-Making and Society.

Joanne E. Gray is a lecturer in digital media in the School of Communication at the Queensland University of Technology and chief investigator at the Digital Media Research Centre.

Louisa Bartolo is a PhD student in the Digital Media Research Centre at the Queensland University of Technology. She is also a student member of the ARC Centre of Excellence for Automated Decision-Making and Society.

Jean Burgess is professor of digital media in the Digital Media Research Centre and School of Communication at the Queensland University of Technology, and associate director of the national ARC Centre of Excellence for Automated Decision-Making and Society.

Nicolas Suzor is professor of law in the Digital Media Research Centre and School of Law at the Queensland University of Technology, and chief investigator in the national ARC Centre of Excellence for Automated Decision-Making and Society.

COGITATIO

Media and Communication (ISSN: 2183–2439)
2021, Volume 9, Issue 4, Pages 250–259
https://doi.org/10.17645/mac.v9i4.4086

Article

Mediated by Code: Unpacking Algorithmic Curation of Urban Experiences

Annelien Smets *, Pieter Ballon and Nils Walravens

imec-SMIT, Vrije Universiteit Brussel, Belgium; E-Mails: annelien.smets@vub.be (A.S.), pieter.ballon@vub.be (P.B.), nils.walravens@imec.be (N.W.)

* Corresponding author

Submitted: 29 January 2021 | Accepted: 24 April 2021 | Published: 18 November 2021

Abstract
Amid the widespread diffusion of digital communication technologies, our cities are at a critical juncture as these technologies are entering all aspects of urban life. Data-driven technologies help citizens to navigate the city, find friends, or discover new places. While these technology-mediated activities come in scope of scholarly research, we lack an understanding of the underlying curation mechanisms that select and present the particular information citizens are exposed to. Nevertheless, such an understanding is crucial to deal with the risk of the socio-cultural polarization assumedly reinforced by this kind of algorithmic curation. Drawing upon the vast amount of work on algorithmic curation in online platforms, we construct an analytical lens that is applied to the urban environment to establish an understanding of algorithmic curation of urban experiences. In this way, this article demonstrates that cities could be considered as a new materiality of curational platforms. Our framework outlines the various urban information flows, curation logics, and stakeholders involved. This work contributes to the current state of the art by bridging the gap between online and offline algorithmic curation and by providing a novel conceptual framework to study this timely topic.

Keywords
algorithmic curation; algorithmic mediators; context media; smart cities; spatiality; urban algorithms

Issue
This article is part of the issue “Algorithmic Systems in the Digital Society” edited by Sanne Kruikemeier (University of Amsterdam, The Netherlands), Sophie Boerman (University of Amsterdam, The Netherlands), and Nadine Bol (Tilburg University, The Netherlands).

1. Introduction

The pervasiveness of digital communication technologies in our urban environments results in large amounts of data. These high volumes of data not only provide the means to monitor particular aspects of urban life, such as air quality or traffic; they also increasingly subject it to mediation by code (Amin & Thrift, 2002). Technology-mediated activities, in so-called smart cities, are indeed (re)producing urban spatiality (Ballon & Smets, 2021; Ridell & Zeller, 2013): citizens use digital applications to discover new destinations (e.g., Tripadvisor), places to eat or rest (e.g., Airbnb), or navigate the urban environment (e.g., Waze). Mediation, however, inherently comes with curation. While mediation refers to the organizational structure of an intermediary in between multiple actors, curation deals with the specific activity of selecting and presenting the information through the intermediary (Rader & Gray, 2015). Currently, this type of curational activities of digital intermediaries is a contentious topic among media and communication scholars. This line of research focuses on the curational role of algorithms in online platforms, such as Facebook or Twitter, and their ability to shape how we see, experience, and understand the world (Bucher, 2018; Kitchin & Dodge, 2011; Willson, 2017).

The intensified mediation through data-driven urban technologies, however, results in the curational practices of algorithms no longer being limited to the mere online world (Foth, 2017). The ubiquity of these technologies is reconfiguring time-space relationships and consequently citizens’ relation with the city and each

other, both online and offline. This raises fundamental questions as to what extent the urban space can still fulfill its role as public space and facilitate social encounters (McQuire, 2016). Therefore, the question of how we experience the urban space is more than a trivial question; it relates to the degree in which we interact with others and experience unpredictable encounters, which has been considered essential to foster necessary skills for cosmopolitan civility (Jacobs, 1961; Sennett, 1978). It is therefore crucial to understand how the use of algorithms and their curational practices could change our urban experience, both from an individual and societal point of view. What are the consequences of citizens' exposure to these algorithms? Can they exclude people from particular public places? Are citizens prone to end up in so-called "urban filter bubbles" (Smets et al., 2019)? These are some of the many questions that illustrate that urban algorithms should be subject of critical research as well (Foth, 2017; Graham, 2005). While the intersection of urbanism and digital technologies (e.g., algorithms) is typically in scope of the smart city discourse, to the best of our knowledge, little attention has been paid to the curational role of algorithms in this field. At the same time, the current study of algorithmic curation itself is mainly grounded in traditional communication studies and lacks a foundation that takes into account the specific nature of the urban context (Foth, 2017). This work sets out to address this shortcoming and develop a framework to organize the research of algorithmic curation of urban experiences. Drawing upon the vast amount of work of algorithmic curation in online platforms, we construct an analytical lens that is subsequently translated to the urban environment, in order to establish our framework on algorithmic curation of urban experiences.

2. Methodological Approach

The aim of this work is to unpack algorithmic curation of urban experiences. In other words, to provide an analytical framework that can be used to guide the study of algorithmic curation of urban experiences. In the last decade, media and communication scholars increasingly started to discuss algorithmic curation in online platforms, such as Facebook or Google Search (e.g., Rader, 2017; Trielli & Diakopoulos, 2019). While it is clear that we cannot simply apply their understanding of algorithmic curation to the physical urban environment (Foth, 2017), we can draw some important insights from these works and use it to develop an analytical perspective that can be applied to the urban context. After all, this intensified mediation through data-driven urban technologies can simply be considered as a new materiality of curational platforms: *context media*. Thanks to the affordances provided by new technologies (e.g., location services, sensing capabilities, etc.), platforms such as Waze or Uber are able to tailor their services towards an individual context. These services are not only taking into account one's individual preferences, but also increasingly rely on additional information such as what you are doing, where you are, and what that particular context looks like. An example is Uber's controversial technique of "surge pricing" where cab fares depend on contextual factors such as weather or traffic conditions (Guda & Subramanian, 2019). This notion of context media has been described in literature using different terminologies, such as "locative" or "urban" media (de Waal, 2013). However, this relationship and (dis)similarities with algorithmic curation in merely online platforms has remained largely unexplored.

This work mainly draws upon a literature review that has been conducted using Google Scholar as research database. Google Scholar is known to have a broad reach in terms of disciplines; however, this broadness is also a challenge in terms of volume (Martín-Martín et al., 2018). We therefore only included English peer-reviewed articles, from both journals and conferences, and academic book sections. To select the search terms for our literature study, we started from the key concepts of our research scope (cities, algorithmic curation, and experiences), and, for each of these concepts, we constructed a list of related terms (e.g., urban space[s], smart city, algorithmic shaping, citizen experience[s]). These search terms were then combined in such a way that our search results would represent the intersection of the three key concepts. This was a challenging exercise: Some combinations did not result in any (or very few) results, whereas others resulted in an explosion in terms of retrieved documents (over 10 thousand results). We learned that the particular combination of "algorithmic curation" and "urban experiences" was a constraining factor, which strengthens the hypothesis that algorithmic curation of urban experiences has rarely been discussed. Consequently, we had to find the balance between sensitivity and specificity. We decided to relax our search conditions in two ways. First, we only combined search terms related to "city" and "algorithmic curation" and collected those results. Second, we combined the three key concepts but no longer used the particular wording of "algorithmic curation" and relaxed the search term to "algorithm" instead. In what follows we will refer to these searches as Search 1 (city + algorithmic curation) and Search 2 (city + algorithm + citizen experiences). This literature review was performed in December 2019. Search 1 resulted in a final selection of 65 articles, based on the above-mentioned criteria, accessibility, and relevance for our work. Interestingly, although search terms related to "city" were explicitly mentioned, none of these articles actually dealt with algorithmic curation in urban environments. The majority of these studies discuss algorithmic curation in online (social) media environments, such as platforms like Facebook or Twitter. The articles were nonetheless highly relevant as they were the basis to develop our understanding of algorithmic curation in general (cf. Section 3). Search 2, on the other hand,

resulted in articles that were very much centered on cities and urban contexts. However, since we searched for "algorithms" instead of "algorithmic curation" most of the articles discussed algorithms from a rather technical point of view. More specifically, from the 71 articles that matched our above-mentioned criteria, another 60 items were excluded because their content did not match our research scope at all. The remaining 11 articles were used to construct the analytical framework for the urban context (cf. Section 4).

In the remainder of this work, we will address our research question about how algorithmic curation of urban experiences can be studied. We will first elaborate on the notion of algorithmic curation (based on results from Search 1 and develop an understanding of the building blocks that we consider to be core of its study. Subsequently, we will translate these building blocks to the urban scene and elaborate them based on the literature in our review (results from Search 2) and complement them with examples known by the authors. Following this, we present an integrative approach to study algorithmic curation of urban experiences and discuss how this can be put into practice.

3. Studying Algorithmic Curation

By analyzing the literature that deals with algorithmic curation we identify four main constructs to guide its study: information flows, feedback loops, curation logics, and multi-stakeholder configurations.

3.1. Information Flows and Feedback Loops

Algorithmic curation is most commonly defined as "organizing, selecting, and presenting subsets of a corpus of information for consumption" (Rader & Gray, 2015, p. 1). Many authors indeed refer to particular mechanisms that influence particular flows of information, such as selecting, organizing, filtering, prioritizing, classifying, and associating (Eslami et al., 2015; Liu, 2010, 2012; Prado, 2014; Rader & Gray, 2015; Shapiro & Hall, 2018; Thorson & Wells, 2015b). In this stream of thought, the algorithm is thus considered to be (part of) a digital information intermediary. Algorithmic curation is therefore often associated with gatekeeping: "The process of culling and crafting countless bits of information into the limited number of messages that reach people each day" (Shoemaker & Vos, 2009, p. 1). However, whereas traditional gatekeeping theories emphasize the negating role of such processes (Thorson & Wells, 2015a), the notion of curation rather stresses the idea of promoting content (Swords, 2017). In this sense, curation is more appropriate for our contemporary media environments characterized by information overload and attention scarcity (Thorson & Wells, 2015a). An intermediary having this gatekeeping (or curational) power has "a major lever in the control of society" (Bagdikian, 1983, p. 226) and therefore scholars plea for a systematic way to analyze the gatekeeping functions of algorithmic curation, so-called "algorithmic audits" (Bandy & Diakopoulos, 2019; Sandvig et al., 2014).

Despite the opaqueness of algorithms ("black boxes"), people seem to develop strategies to game the algorithm. Even though they do not know how the algorithm works, by experimenting with it and adapting their behavior, they develop strategies to make use of the algorithm in their own favor (Bucher, 2017; Eslami et al., 2016). As put by Rader and Gray (2015): "They adapt their behavior to correspond with how they believe the system works, in order to accomplish their goals for using the system" (p. 8). Some users, for example, claim to use another computer to prevent ending up in a "filter bubble" (Bilandzic et al., 2018). Napoli (2018) even talks about an entire industry that "has arisen around optimizing content for social media curation algorithms" (p. 8). This means that there is an (unconscious) response to the feedback loop characteristics of these systems. These feedback loops are hence an important factor in shaping the overall system behavior and should therefore be considered in the analysis of algorithmic curation (Rader & Gray, 2015).

3.2. Curation Logics and Multi-Stakeholder Configurations

Such analysis should not be limited to the algorithm itself. Scholars like Kitchin (2017b) and Seaver (2019) call for the study of algorithms in their "full socio-technical assemblage" which requires an assessment including all actors. Indeed, algorithms do not originate from a void and they should be examined as sociotechnical constructs influenced by their context of creation and use (Seaver, 2019). Grounding in the literature concerned with information flows in media environments, Thorson and Wells (2015a) argue that different curating actors do not exist next to each other and rather show a significant degree of overlap or intersection. They call for an empirical investigation of the extent to which these interactions occur, with particular attention to the "curation logics" of these actors. These logics refer to the "particular interests, norms, incentives and network positions" that play an essential role in the decision about which content to present (Thorson & Wells, 2015a, p. 5). The identification of these curation logics is argued to be "useful to structure theory and guide empirical research" (Landerer, 2013, p. 248).

The importance of the curation logics of multiple actors refers to the multi-sidedness of algorithmic curation. In the literature, we observe three stands of research each discussing a different side: (1) End-users—examining how algorithmic curation influences users' content exposure on social media (Bandy & Diakopoulos, 2019; Diakopoulos, 2015; Nelimarkka et al., 2018; Rader, 2017; Yatid, 2019), their exposure to diverse news (Ku et al., 2019; Wohn & Bowe, 2016) or fake news (Cohen, 2018), or users' feelings and beliefs about the

algorithmic curation (Bucher, 2017; Eslami et al., 2015, 2016; Rader & Gray, 2015); (2) providers—studying the actor who provides the information that is curated, such as journalists who create online news content (Usher, 2017); (3) operators—examining the algorithm and its operator, for example Bernal's (2018) discussion on how Facebook's business model makes the fake news problem inevitable.

Summarizing, the study of algorithmic curation should be guided by the following questions: What information flows are being curated? Which feedback loops can be identified? Who are the different stakeholders involved? What are their curation logics? However, in order to apply this to the urban context we need to be able to identify the algorithmic mediators that curate the information and eventually understand how these relate to one another, as well as the urban experience itself.

4. Setting the Urban Scene

So far, we mainly discussed curation in online platforms such as Facebook. However, the premise of this work is that a new materiality of curational platforms arises: the urban scene. There is indeed a growing line of work that focusses on the spatiality of media such as Foursquare, Pokémon Go, or Waze. However, the personal applications that people use in their everyday life are not the sole algorithmic mediators in the urban sphere. Indeed, the last decade has been characterized by cities increasingly implementing digital technologies, both in the public environment and public services (Ballon & Smets, 2021). The main difference between purely online platforms such as Facebook and those that operate at the intersection of offline and online, is their ubiquity. As a consequence, the identification of these mediators and the information flows they curate is particularly challenging. Before addressing this, we start with a brief discussion of the essential building blocks of information—data.

4.1. Urban Data Landscape

When discussing data in an urban setting, most people think of "urban operational data" generated by sensors monitoring air quality or cameras counting road traffic. However, there are many other data sources contributing to the urban data landscape (Kitchin, 2017a). Mobile phone operators generate data about people's location; and social media and websites such as TripAdvisor or Airbnb generate data that to a large extent reflect personal experiences of citizens (photos, reviews, likes, etc.; Cervantes et al., 2016; Foth et al., 2011). Another category of "data providers" are organizations such as financial institutions or retail chains that generate data on financial transactions and purchases. All of these companies and platforms are increasingly making their data publicly available through API's or sell them through data brokers (Kitchin, 2017a). Another emergent type of data source is crowdsourcing or citizen science. Here, citizens actively contribute to data collections (e.g., Open Street Map) or even install their own data collection infrastructure to collect data about their neighborhood. The distinctive feature of most of this data (and related technologies) is their velocity, or real time character (Kitchin, 2014). This enables a wide range of curational activities that require algorithmic mediators that can analyze and act upon this data simultaneously. Perhaps the most common example is traffic, where real time data from various sources is used to fuel personal navigation applications such as Waze. However, other more innovative scenarios exist where such real time data is used to physically alter the urban infrastructure. For example, by means of changing the function of a particular zone (e.g., vehicles vs. pedestrians). The remainder of this section is set out to identify those different types of algorithmic mediators and discuss their multi-stakeholder configurations.

4.2. Urban Information Flows

Following our understanding of algorithmic curation developed in the previous section, we first need to identify digital intermediaries in cities in order to analyze the algorithmically curated urban information flows. Drawing from case descriptions discussing data-driven urban technologies, we identify two main categories of urban mediators that will be core in our conceptual framework: networked people and urban infrastructure (Figure 1).

4.2.1. Networked People

This category represents digital services that citizens directly interact with. Indeed, in contemporary cities, citizens are participating members of a greater collective and technology aims to strengthen this connectedness among citizens as well as their environment (Foth et al., 2011). In this sense, citizens are increasingly networked people. This category consists of personal applications such as smartphone applications or websites. For example, a smartphone application providing personalized recommendations to tourists about points of interest in a particular area (Cervantes et al., 2016). The navigation application Waze is another exemplary case of how people use digital interfaces to consult information and generate knowledge about their urban environment. In this case, the application is often no longer a mere information intermediary, but also acts as an actual guide through the roads with the least traffic. Personal applications can also be used to connect with other people in the environment. Citysocializer is such an application that aims to facilitate meeting new people and making friends in your own city by attending social events. Apart from applications and services to meet new people, popular applications like Facebook, Yelp, or Find My (Friends) also allow users to share their location with their friends.

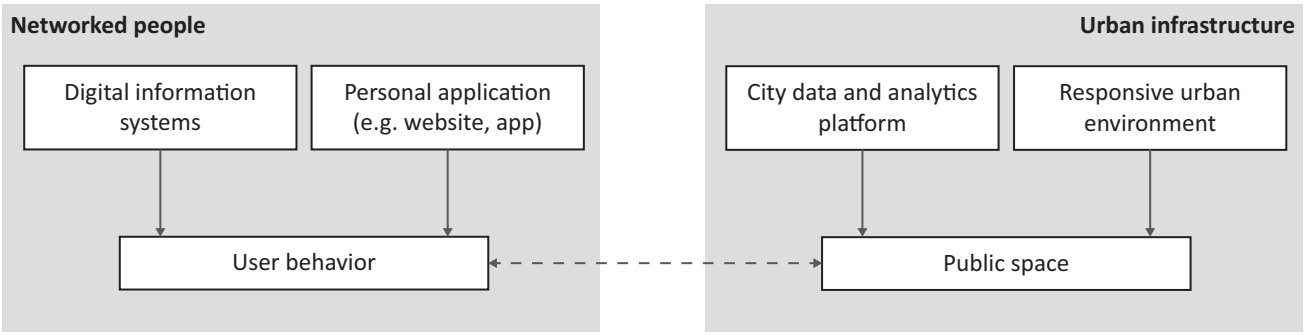


Figure 1. Identified algorithmic mediators in the urban context, and their impact on user behavior and the public space.

Another type of application in this category are community platforms such as Nextdoor or Hoplr. These applications are oriented towards community building and bringing together people from the same neighborhood by facilitating the exchange of information or goods. By establishing this sense of community, these platforms also aim to enhance the overall experience of a particular neighborhood (De Meulenaere et al., 2020). Apart from personal applications, we also distinguish more generally available digital information systems, such as digital signage that can be found on city squares. Citizens can consult these systems to gather information about the environment, such as information on upcoming events or an overview of pleasant walking routes in the area (Koto & Bandung, 2016).

Each of these services thus mediates information and, in this way, also inevitably engages in curational practices by deciding which information to filter, select, or present. Similar to the Facebook algorithm deciding which friends most frequently pop up on your newsfeed, the Citysocializer algorithm could greatly influence your new group of friends by recommending particular types of events. The information that is curated by these intermediaries could greatly affect citizens behavior, for example by suggesting where to go. Therefore, we assemble the impact of these curated information flows in the notion of user behavior, the collection of actions undertaken by citizens. The latter represents a crucial connection to the physical public space, as will become evident in the following sections.

4.2.2. Urban Infrastructure

The urban infrastructure represents a second category of technologies that mediate the urban experience. The main difference with the previous category is the level of human agency. In the category of networked people, citizens have to take the initiative to consult information, i.e., they "pull" information. Contrary, the information that is mediated by the urban infrastructure is rather "pushed" upon the citizen. This category of mediators thus acts independently from any user interaction. Although an in-depth discussion goes beyond the scope of this article, this distinction has far-reaching consequences, not in the least for the degree to which citizens can control the algorithmic curation they are subjected to or to what extent they can develop the skills to cope with it.

We identify two types of intermediaries in this category: (1) city data and analytics platforms, and (2) responsive urban environments. The former refers to algorithm-based decision-support systems, such as dashboards or knowledge discovery tools that can inform city administrations about changes to the urban infrastructure (Al Nuaimi et al., 2015; Foth et al., 2011; Hanzl et al., 2012; Lim et al., 2018; Mora et al., 2017). This could be, for example, a citizen experience dashboard that "allows public administrations to understand the real expectations of the citizens to optimize investments or even predict the potential impact on citizens when redesigning the services" (Abella et al., 2017, p. 51). These dashboards are driven by data on citizens' experiences that could be provided directly (e.g., through surveys) or collected indirectly (e.g., posts on social media). Changes to the urban infrastructure could relate to public services such as public transport but also to urban planning and city infrastructure. An example are so-called "urban digital twins" (Mohammadi & Taylor, 2017): (real-time) spatiotemporal data-driven models of the city that allow policy makers to rely on predictive modelling techniques to study the impact of a particular change to the urban infrastructure. Such a digital twin can be used to predict what happens to the air quality when a particular area would be turned into a pedestrian zone. This kind of city data and analytics platforms thus enable city administrations to make informed decisions about possible changes to the urban infrastructure, both public services and the public space.

A responsive urban environment, on the other hand, refers to an element of the physical urban space that is capable of acting upon data and adapting itself to it. A digital advertisement that adapts to the emotions of passers-by, and thereby eventually influences their experience, for example. This kind of technology is able to capture people's emotions and react to it, so-called "emphatic media" (McStay, 2016). The latter differs from the digital signage described in the previous section, as this one adapts itself regardless of any user interaction.

Another example of urban responsiveness is illustrated by adaptive street environments: "the morphological transformation of the urban space enabled by smart sensors and responsive materials for adaptive options" (Andreani et al., 2019, p. 18). An example could be dynamic adjustments of traffic lanes with respect to vehicle flows or increasing presence of bikers and pedestrians, or adaptive traffic lights based on the actual traffic demand.

4.2.3. Interaction Effect

Finally, we want to address the missing link in the discussion so far: the relationship between networked people and the urban infrastructure. As said before, we consider public space to be constantly (re)produced through human activities (cf. the dotted line in Figure 1). Consequently, it is important to take into account how algorithmic curation could play a role in this relationship. For example, in our exemplary case of Waze, evidence shows that the use of Waze increases the traffic flow in smaller streets, which are often only designed for local traffic (Fisher, 2020; Macfarlane, 2019). As a consequence, these small streets suffer from the increasing volume of through traffic, e.g., resulting in increased congestion or accelerated deterioration of the road surface. This means that even when citizens do not use Waze themselves, their urban experience could be influenced by others who do. At this particular intersection of individual user behavior and the public space, the identification of conflicting interests among various actors is most emergent.

4.3. Urban Curation Logics and Multi-Stakeholder Configurations

Our understanding of algorithmic curation of urban experiences also needs to encompass an analysis of the curation logics, i.e., the logics behind the mechanisms that curate the information flows. Here, one should identify the different stakeholders involved and their corresponding interests and motivations. However, the specific group of stakeholders and their interaction typically depends on the specific case, and therefore we omit a detailed discussion here. On a general level, we identify the following stakeholder categories: End-users, the ones who receive the information, most often citizens in the urban context; providers, those who (indirectly) provide the data, e.g., retailers, points of interest or city administrations; service operators, the entity who operates the service in which the algorithms are embedded; and society, those who represent the stakes of society, i.e., the common good.

It is clear that these categories are non-exclusive and thus serve as a means to guide the identification of the stakeholders rather than a strict categorization. While the first three categories are in line with the categories previously identified in our discussion in Section 3.2, the fourth one has not yet been formulated explicitly. The societal impact of online algorithmic curation is most often considered as an aspect of the service operator itself (cf. discussions on accountability). However, in this case of the urban environment, we value the explicit modelling of this category as a separate one, because in public space there is the inherent involvement of city administrations and/or governments who have to safeguard the public interest, which might conflict with others' interests (Smets et al., 2020). The increased traffic in residential areas due to the use of Waze (Fisher, 2020) is again an example of such conflicting curational logics. While the end-user might indeed want to take the fastest route from point A to B, from a societal point of view this might not be desired due to the resulting congestion or deterioration in small streets. At first sight, this might seem a mere practical issue in the sense that people lose time or roads should be maintained more frequently. However, this increased traffic also changes the safety and hence the social function of these streets. Whereas normally children can play on the streets, this will no longer be the case when the traffic increases. This not only affects the children, but also the parents who often get to know each other by getting out on the streets while their children play. This example clearly illustrates the complex interplay between the curation logics of multiple stakeholders and will be a crucial building block of the study of algorithmic curation.

5. Towards an Integrative Approach

We conclude this discussion by presenting an integrative framework (Figure 2) that illustrates the identified urban algorithmic mediators and information flows that might eventually alter the urban experience, thereby addressing our main research question. Our framework illustrates that there are multiple ways for algorithms to curate urban experiences, however, the actual relationships are much more complex compared to how they are depicted in Figure 2. As highlighted before, we should also take into account feedback loops: For example, citizens sharing their experiences on social media could provide data for urban digital twins (Mohammadi & Taylor, 2020).

By breaking down the different steps from data to the urban experience, this framework allows to analyze the curation mechanisms (selecting, organizing, etc.) that act upon the information flows and thus structure the study of algorithmic curation of urban experiences. This framework not only allows to structure empirical research to investigate a particular phenomenon, but also to facilitate comparative research or study the normative aspects of algorithmic curation. The latter could for example relate to the formulation of particular normative principles for algorithmic systems, where the framework allows to investigate if and to what extent this principle should apply to different information flows or how it can be operationalized.

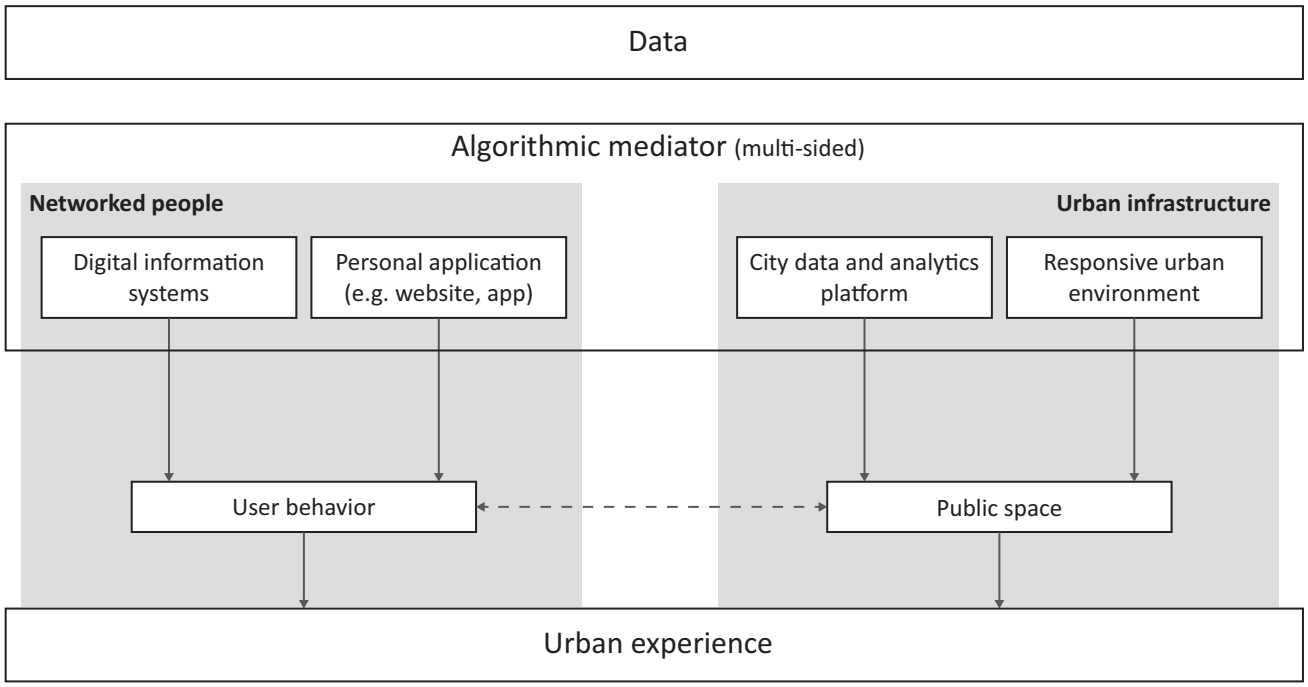


Figure 2. Framework on algorithmic curation of urban experiences.

The operationalization of the research would then require studying algorithms, which is known to be a difficult task. Different research methods have been suggested such as interviewing designers, reverse engineering, or examining pseudo-code (Kitchin, 2017b). We suggest that the study of these algorithms is accompanied by a stakeholder mapping in order to take into account the broader socio-technical setting. The stakeholder categories mentioned in the previous section can be used as a starting point to identify the specific stakeholders involved and how their interests define or influence the curation logics. Here, the most important question is to identify the overall strategic objectives of each actor and how these potentially result in incompatible or competing curation logics. There is also the possibility to conduct user studies looking closely into the citizens' experiences of these curational activities. To do so, there is a broad range of user research methodologies that already has been extensively used in urban contexts, such as surveys, focus groups or large-scale experimental designs often referred to as "living labs" (Ballon & Schuurman, 2015).

6. Conclusions

This work originated from our interest in algorithmic curation in urban environments, and the observations that on the one hand, the current study of algorithmic curation lacks a foundation that takes into account the specific nature of the urban context, whereas, on the other hand, the smart city discourse that is particularly concerned with the interplay of digital technology and urbanism, fails to capture this curational aspect of algorithms. To address this shortcoming, we present an analytical approach to study algorithmic curation of urban experiences, building upon prior work in media and communication studies, and elaborated upon by means of case examples in the urban realm. More specifically, we identified urban algorithmic mediators, consisting of networked people and urban infrastructure, discussed the notion of urban curation logics, and categorized potential stakeholders. The latter can be used as a starting point to identify the specific stakeholders involved, and how their interests define or influence the curation practices.

The main limitation of this study is the degree to which each of the components in the framework is elaborated. Although we acknowledge this shortcoming, we believe that our work could serve as a starting point for extensions or adjustments based on further research. We acknowledge that the current framework is indeed just a conceptual one and would thus benefit from an empirical verification. Future work could address this by applying the framework to actual use cases. This will not only demonstrate its empirical value but also help to refine it. As such, the study at hand attempts to be a first valuable contribution to the critical study of algorithmic curation in urban contexts and remains open to empirical verification, extensions, and adjustments as more research in this field emerges.

In this way, we present a first approach towards the study of algorithmic curation in urban environments and more specifically the algorithmic curation of urban experiences. Our analysis indicates some similarities with algorithmic curation in online platforms. However, the physical characteristics of the urban environment require an adjusted approach to study algorithmic

curation in the urban context. After all, algorithms can influence urban experiences without a direct technological interface towards the citizen, for example through the urban infrastructure. Moreover, a significant interaction effect exists between individual user behavior and the public space that can have complex (in)direct consequences. These findings illustrate that there is definitely a continuation of algorithmic curation from the online to the offline world, and that its study requires a full sociotechnical approach both in terms of actors and physical places. We believe that this framework can set the scene for further research in this field that not only considers the curational practices themselves, but also investigates related concepts and phenomena in more depth. For example, if and how citizens might develop strategies to "game the algorithm." Inspired by the German artist Simon Weckert (2020), who generated a virtual traffic jam in Google Maps by walking through a street with 99 mobile phones in a handcart, we could imagine that citizens adopt a similar hack to avoid Waze-users disturbing their quiet neighborhood. By extending the scope of algorithmic curation to the urban environment, we hope this work inspires other scholars to study phenomena that we know online, in the offline world as well.

Acknowledgments

We are particularly grateful to Rob Heyman for his contribution to an earlier version of this work. We also thank the anonymous reviewers whose suggestions helped to improve and clarify our manuscript.

Conflict of Interests

The authors declare no conflict of interests.

References

Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2017). A model for the analysis of data-driven innovation and value generation in smart cities' ecosystems. *Cities*, *64*, 47–53. https://doi.org/10.1016/j.cities.2017.01.011

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, *6*(1), 25.

Amin, A., & Thrift, N. (2002). *Cities: Reimagining the urban*. Polity Press.

Andreani, S., Kalchschmidt, M., Pinto, R., & Sayegh, A. (2019). Reframing technologically enhanced urban scenarios: A design research model towards human centered smart cities. *Technological Forecasting and Social Change*, *142*, 15–25.

Bagdikian, B. H. (1983). *The media monopoly*. Beacon Press.

Ballon, P., & Schuurman, D. (2015). Living labs: Concepts, tools, and cases. *Info*, *17*(4). https://doi.org/10.1108/info-04-2015-0024

Ballon, P., & Smets, A. (2021). De slimme stad: Stedelijke dataficatie in theorie en praktijk [The smart city: Urban datafication in theory and practice]. In G-J. Hospers & P. Renooy (Eds.), *De Wereld van De Stad* [The world of the city]. Berghauser Pont.

Bandy, J., & Diakopoulos, N. (2019). *Auditing news curation systems: A case study examining algorithmic and editorial logic in apple news*. arXiv. https://arxiv.org/abs/1908.00456

Bernal, P. (2018). Fakebook: Why Facebook makes the fake news problem inevitable. *Northern Ireland Legal Quarterly*, *69*(4), 513–530.

Bilandzic, A., Casadevall, D., Foth, M., & Hearn, G. (2018). Social and spatial precursors to innovation: The diversity advantage of the creative fringe. *The Journal of Community Informatics*, *14*(1), 160–182. https://doi.org/10.15353/joci.v14i1.3408

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, *20*(1), 30–44. https://doi.org/10.1080/1369118X.2016.1154086

Bucher, T. (2018). *If…Then: Algorithmic power and politics*. Oxford University Press.

Cervantes, O., Gutiérrez, E., Gutiérrez, F., & Sánchez, J. A. (2016). Social metrics applied to smart tourism. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*(1), 117–124.

Cohen, J. N. (2018). Exploring echo-systems: How algorithms shape immersive media environments. *Journal of Media Literacy Education*, *10*(2), 139–151.

De Meulenaere, J., Baccarne, B., Courtois, C., & Ponnet, K. (2020). Disentangling social support mobilization via online neighborhood networks. *Journal of Community Psychology*, *49*(2), 481–498. https://doi.org/10.1002/jcop.22474

de Waal, M. (2013). *De stad als interface: Hoe nieuwe media de stad veranderen* [The city as interface: How new media are changing the city]. Rotterdam.

Diakopoulos, N. (2015). Accountability in algorithmic decision-making. *Queue*, *13*(9), 126–149.

Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., Rickman, A., Hamilton, K., & Kirlik, A. (2016). First I like it, then I hide it: Folk theories of social feeds. In J. Kaye & A. Druin (Eds.), *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 2371–2382). ACM.

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015). I always assumed that I wasn't really that close to [her]: Reasoning about invisible algorithms in news feeds. In B. Begole & J. Kim (Eds.), *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 153–162). ACM.

Fisher, E. (2020). Do algorithms have a right to the city? Waze and algorithmic spatiality. *Cultural Studies*. Advance online publication. https://doi.org/10.1080/09502386.2020.1755711

Foth, M. (2017, June 26). *The software-sorted city: Big*

data & algorithms [Workshop paper]. Digital Cities 10 Workshop, Troyes, France.

Foth, M., Forlano, L., Satchell, C., & Gibbs, M. (2011). Crowdsensing in the web: Analyzing the citizen experience in the urban space. In M. Foth, L. Forlano, C. Satchell, & M. Gibbs (Eds.), *From social butterfly to engaged citizen* (pp. 353–373). MIT Press. https://doi.org/10.7551/mitpress/8744.003.0029

Graham, S. D. N. (2005). Software-sorted geographies. *Progress in Human Geography*, *29*(5), 562–580. https://doi.org/10.1191/0309132505ph568oa

Guda, H., & Subramanian, U. (2019). Your Uber is arriving: Managing on-demand workers through surge pricing, forecast communication, and worker incentives. *Management Science*, *65*(5), 1995–2014. https://doi.org/10.1287/mnsc.2018.3050

Hanzl, M., Dzik, K., Kowalczyk, P., Kwieciński, K., Stankiewicz, E., & Wierzbicka, A. L. (2012). Human geomatics in urban design—Two case studies. *Future Internet*, *4*(1), 347–361.

Jacobs, J. (1961). *The death and life of great American cities*. Vintage Books.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1–14. https://doi.org/10.1007/s10708-013-9516-8

Kitchin, R. (2017a). Data-driven urbanism. In R. Kitchin, T. P. Lauriault, & G. McArdle (Eds.), *Data and the city* (1st ed., pp. 44–56). Routledge. https://doi.org/10.4324/9781315407388-4

Kitchin, R. (2017b). Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), 14–29. https://doi.org/10.1080/1369118X.2016.1154087

Kitchin, R., & Dodge, M. (2011). *Code/space: Software and everyday life*. MIT Press.

Koto, Z., & Bandung, Y. (2016). Interactive digital signage architecture to improve user interaction on tourism information services. In M. Amin Sulthoni (Ed.), *2016 international symposium on electronics and smart devices (ISESD)* (pp. 380–385). IEEE. https://doi.org/10.1109/ISESD.2016.7886752

Ku, K. Y., Kong, Q., Song, Y., Deng, L., Kang, Y., & Hu, A. (2019). What predicts adolescents' critical thinking about real-life news? The roles of social media news consumption and news media literacy. *Thinking Skills and Creativity*, *33*, Article 100570. https://doi.org/10.1016/j.tsc.2019.05.004

Landerer, N. (2013). Rethinking the logics: A conceptual framework for the mediatization of politics. *Communication Theory*, *23*(3), 239–258. https://doi.org/10.1111/comt.12013

Lim, C., Kim, K.-J., & Maglio, P. P. (2018). Smart cities with big data: Reference models, challenges, and considerations. *Cities*, *82*, 86–99. https://doi.org/10.1016/j.cities.2018.04.011

Liu, S. B. (2010). Trends in distributed curatorial technology to manage data deluge in a networked world. *Upgrade: The European Journal for the Informatics Professional*, *11*(4), 18–24.

Liu, S. B. (2012). The living heritage of historic crises: Curating the Bhopal disaster in the social media landscape. *Interactions*, *19*(3), 20–24. https://doi.org/10.1145/2168931.2168938

Macfarlane, J. (2019). When apps rule the road: The proliferation of navigation apps is causing traffic chaos. It's time to restore order. *IEEE Spectrum*, *56*(10), 22–27.

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, *12*(4), 1160–1177. https://doi.org/10.1016/j.joi.2018.09.002

McQuire, S. (2016). *Geomedia: Networked cities and the future of public space*. Wiley.

McStay, A. (2016). Empathic media and advertising: Industry, policy, legal, and citizen perspectives (the case for intimacy). *Big Data & Society*, *3*(2). https://doi.org/10.1177/2053951716666868

Mohammadi, N., & Taylor, J. (2020). Knowledge discovery in smart city digital twins. In T. X. Bui (Ed.), *Proceedings of the 53rd Hawaii international conference on system sciences* (pp. 1656–1664). HICSS. http://hdl.handle.net/10125/639439

Mohammadi, N., & Taylor, J. E. (2017). Smart city digital twins. In P. Bonissone & D. Fogel (Eds.), *2017 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1–5). IEEE. https://doi.org/10.1109/SSCI.2017.8285439

Mora, H., Gilart-Iglesias, V., Pérez-del Hoyo, R., & Andújar-Montoya, M. (2017). A comprehensive system for monitoring urban accessibility in smart cities. *Sensors*, *17*(8), 1834. https://doi.org/10.3390/s17081834

Napoli, P. M. (2018). *What social media platforms can learn from audience measurement: Lessons in the self-regulation of "black boxes"*. SSRN. http://dx.doi.org/10.2139/ssrn.3115916

Nelimarkka, M., Laaksonen, S.-M., & Semaan, B. (2018). Social media is polarized, social media is polarized: Towards a new design agenda for mitigating polarization. In I. Koskinen & Y.-K. Lim (Eds.), *Proceedings of the 2018 on designing interactive systems conference 2018—DIS '18* (pp. 957–970). ACM. https://doi.org/10.1145/3196709.3196764

Prado, L. (2014). A protected life: Speculations on object-mediated relationships. In F. Paiva & C. Moura (Eds.), *DESIGNA 2013: Interface proceedings* (p. 361). Ubi. https://labcom.ubi.pt/ficheiros/20140608-designa2013_proceedings_flat.pdf

Rader, E. (2017). Examining user surprise as a symptom of algorithmic filtering. *International Journal of Human-Computer Studies*, *98*, 72–88.

Rader, E., & Gray, R. (2015). Understanding user beliefs about algorithmic curation in the Facebook news feed. In B. Begole & J. Kim (Eds.), *Proceedings of the 33rd annual ACM conference on human factors*

in computing systems—CHI '15 (pp. 173–182). ACM. https://doi.org/10.1145/2702123.2702174

Ridell, S., & Zeller, F. (2013). *Mediated urbanism: Navigating an interdisciplinary terrain*. SAGE.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: Converting Critical Concerns Into Productive Inquiry*, *22*, 4349–4357.

Seaver, N. (2019). Knowing algorithms. In J. Vertesi & D. Ribes (Eds.), *DigitalSTS: A field guide for science & technology studies* (pp. 412–422). Princeton University Press.

Sennett, R. (1978). *The fall of public man: On the social psychology of capitalism*. Vintage Books.

Shapiro, B. R., & Hall, R. (2018). Personal curation in a museum. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), pp. 1–22. https://doi.org/10.1145/3274427

Shoemaker, P. J., & Vos, T. P. (2009). *Gatekeeping theory*. Routledge.

Smets, A., Montero, E., & Ballon, P. (2019). Does the bubble go beyond? An exploration of the Urban filter bubble. In O. S. Shalom, D. Jannach, & I. Guy (Eds.), *Proceedings of the 1st workshop on the impact of recommender systems* (pp. 1–6). CEUR. http://ceur-ws.org/Vol-2462/paper3.pdf

Smets, A., Walravens, N., & Ballon, P. (2020). Designing recommender systems for the common good. In T. Kuflik & I. Torre (Eds.), *Adjunct publication of the 28th ACM conference on user modeling, adaptation and personalization* (pp. 276–278). ACM. https://doi.org/10.1145/3386392.3399570

Swords, J. (2017). Crowd-patronage—Intermediaries, geographies, and relationships in patronage networks. *Poetics*, *64*, 63–73.

Thorson, K., & Wells, C. (2015a). Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory*, *26*(3), 309–328.

Thorson, K., & Wells, C. (2015b). How gatekeeping still matters: Understanding media effects in an era of curated flows. In T. P. Vos & F. Heinderyckx (Eds.), *Gatekeeping in transition* (pp. 39–58). Routledge.

Trielli, D., & Diakopoulos, N. (2019). Search as news curator: The role of Google in shaping attention to news information. In S. Brewster & G. Fitzpatrick (Eds.), *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–15). ACM. https://doi.org/10.1145/3290605.3300683

Usher, N. (2017). Venture-backed news startups and the field of journalism: Challenges, changes, and consistencies. *Digital Journalism*, *5*(9), 1116–1133.

Weckert, S. (2020). *Google maps hacks: Performance & installation, 2020*. http://www.simonweckert.com/googlemapshacks.html

Willson, M. (2017). Algorithms (and the) everyday. *Information, Communication & Society*, *20*(1), 137–150. https://doi.org/10.1080/1369118X.2016.1200645

Wohn, D. Y., & Bowe, B. J. (2016). Micro agenda setters: The effect of social media on young adults' exposure to and attitude toward news. *Social Media + Society*, *2*(1). https://doi.org/10.1177/2056305115626750

Yatid, M. M. (2019). Truth tampering through social media: Malaysia's approach in fighting disinformation & misinformation. *IKAT: The Indonesian Journal of Southeast Asian Studies*, *2*(2), 203–230.

About the Authors

Annelien Smets is a doctoral researcher at imec-SMIT at the Vrije Universiteit Brussel. Her research interests include datafication and personalization in digital media and smart cities. Her doctoral research deals with the use of recommender systems in cities, with a particular focus on the value of serendipity.

Pieter Ballon is full professor at the Department of Communication Studies at Vrije Universiteit Brussel where he teaches on media and internet economics. He is director of research group imec-SMIT. His research centers on the political economy of digital platforms in media, mobile, and urban contexts. He has published widely on these topics in peer-reviewed international journals.

Nils Walravens graduated cum laude as master in communication sciences at the Vrije Universiteit Brussel in July 2007. He started working for imec-SMIT in August of 2007 as a researcher, developing expertise in the field of business modelling and policy research. In 2016, he defended his PhD on the topic of smart cities and public value, and has since worked as a senior researcher focusing on the topics of smart cities and open data at imec-SMIT.

Media and Communication (ISSN: 2183-2439)

Media and Communication is an international open access journal dedicated to a wide variety of basic and applied research in communication and its related fields. It aims at providing a research forum on the social and cultural relevance of media and communication processes.

www.cogitatiopress.com/mediaandcommunication

COGITATIO