

ROUTLEDGE FOCUS



# ALGORITHMIC GATEKEEPING FOR PROFESSIONAL COMMUNICATORS

Power, Trust, and Legitimacy

Arjen van Dalen



# ALGORITHMIC GATEKEEPING FOR PROFESSIONAL COMMUNICATORS

This book provides a critical study of the power, trust, and legitimacy of algorithmic gatekeepers.

The news and public information which citizens see and hear is no longer solely determined by journalists, but increasingly by algorithms. Van Dalen demonstrates the gatekeeping power of social media algorithms by showing how they affect exposure to diverse information and misinformation and shape the behavior of professional communicators. Trust and legitimacy are foregrounded as two crucial antecedents for the acceptance of this algorithmic power. This study reveals low trust among the general population in algorithms performing journalistic tasks and a perceived lack of legitimacy of algorithmic power among professional communicators. Drawing on case studies from YouTube and Instagram, this book challenges technological deterministic discourse around “filter bubbles” and “echo chambers” and shows how algorithmic power is situated in the interplay between platforms, audiences, and professional communicators. Ultimately, trustworthy algorithms used by news organizations and social media platforms as well as algorithm literacy training are proposed as ways forward toward democratic algorithmic gatekeeping.

Presenting a nuanced perspective which challenges the deep divide between techno-optimistic and techno-pessimistic discourse around algorithms, *Algorithmic Gatekeeping* is recommended reading for journalism and communication researchers in related fields.

**Arjen van Dalen** is Professor WSR at the Center for Journalism of the University of Southern Denmark. He wrote his PhD dissertation on Political Journalism in Comparative Perspective and is co-author of an award-winning book on this topic with Cambridge University Press. He published in journals such as *Political Communication*, *Journalism*, the *International Journal of Press/Politics*, and *Public Opinion Quarterly*. His research on algorithmic gatekeeping is funded by the Independent Research Fund Denmark and Carlsberg Foundation.

## **Disruptions: Studies in Digital Journalism**

Series editor: Bob Franklin

*Disruptions* refers to the radical changes provoked by the affordances of digital technologies that occur at a pace and on a scale that disrupts settled understandings and traditional ways of creating value, interacting, and communicating both socially and professionally. The consequences for digital journalism involve far-reaching changes to business models, professional practices, roles, ethics, products, and even challenges to the accepted definitions and understandings of journalism. For Digital Journalism Studies, the field of academic inquiry which explores and examines digital journalism, disruption results in paradigmatic and tectonic shifts in scholarly concerns. It prompts reconsideration of research methods, theoretical analyses, and responses (oppositional and consensual) to such changes, which have been described as being akin to “a moment of mind-blowing uncertainty.”

Routledge’s book series, *Disruptions: Studies in Digital Journalism*, seeks to capture, examine, and analyze these moments of exciting and explosive professional and scholarly innovation which characterize developments in the day-to-day practice of journalism in an age of digital media, which are articulated in the newly emerging academic discipline of Digital Journalism Studies.

### **Disrupting Chinese Journalism**

Changing Politics, Economics and Journalistic Practices of the Legacy Newspaper Press

*Haiyan Wang*

### **Algorithmic Gatekeeping for Professional Communicators**

Power, Trust, and Legitimacy

*Arjen van Dalen*

For more information about this series, please visit: [www.routledge.com/Disruptions/book-series/DISRUPTDIGJOUR](http://www.routledge.com/Disruptions/book-series/DISRUPTDIGJOUR)

# **Algorithmic Gatekeeping for Professional Communicators**

Power, Trust, and Legitimacy

**Arjen van Dalen**

 **Routledge**  
Taylor & Francis Group  
LONDON AND NEW YORK



First published 2023

by Routledge

4 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge

605 Third Avenue, New York, NY 10158

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2023 Arjen van Dalen

The right of Arjen van Dalen to be identified as author of this work has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

The Open Access version of this book, available at [www.taylorfrancis.com](http://www.taylorfrancis.com), has been made available under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 license.

*Trademark notice:* Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

*British Library Cataloguing-in-Publication Data*

A catalogue record for this book is available from the British Library

*Library of Congress Cataloguing-in-Publication Data*

Names: Van Dalen, Arjen, 1980- author.

Title: Algorithmic gatekeeping for professional communicators : power, trust and legitimacy / Arjen van Dalen.

Description: Abingdon, Oxon ; New York : Routledge, 2023. |

Series: Disruptions : studies in digital journalism | Includes bibliographical references and index.

Identifiers: LCCN 2023003541 (print) | LCCN 2023003542 (ebook)

| ISBN 9781032450711 (hardback) | ISBN 9781032450728

(paperback) | ISBN 9781003375258 (ebook)

Subjects: LCSH: Social media and journalism. | Algorithms. | Filter bubbles (Information filtering)

Classification: LCC PN4766.V36 2023 (print) | LCC PN4766

(ebook) | DDC 302.23/1--dc23/eng/20230227

LC record available at <https://lcn.loc.gov/2023003541>

LC ebook record available at <https://lcn.loc.gov/2023003542>

ISBN: 978-1-032-45071-1 (hbk)

ISBN: 978-1-032-45072-8 (pbk)

ISBN: 978-1-003-37525-8 (ebk)

DOI: 10.4324/9781003375258

Typeset in Times New Roman

by MPS Limited, Dehradun

# Contents

|   |           |
|---|-----------|
| <i>List of Figures</i>                              | vii       |
| <i>List of Table</i>                                | viii      |
| <i>Acknowledgments</i>                              | ix        |
| <b>1 Introduction</b>                               | <b>1</b>  |
| 1.1 <i>Algorithmic power</i>                        | 1         |
| 1.2 <i>Conceptualizing algorithmic gatekeeping</i>  | 2         |
| 1.3 <i>Technological determinism: hard and weak</i> | 4         |
| 1.4 <i>Theoretical considerations</i>               | 8         |
| 1.5 <i>Legitimacy and trust</i>                     | 11        |
| 1.6 <i>Outline and content of the book</i>          | 14        |
| <b>2 Algorithm Aversion</b>                         | <b>17</b> |
| 2.1 <i>Introduction</i>                             | 17        |
| 2.2 <i>Making sense of news algorithms</i>          | 18        |
| 2.3 <i>Research questions</i>                       | 21        |
| 2.4 <i>Method</i>                                   | 22        |
| 2.5 <i>Results</i>                                  | 23        |
| 2.6 <i>Conclusion</i>                               | 29        |
| <b>3 The Mediating Power of Algorithms</b>          | <b>31</b> |
| 3.1 <i>Introduction</i>                             | 31        |
| 3.2 <i>Misinformation and algorithms</i>            | 32        |
| 3.3 <i>YouTube and autism spectrum disorder</i>     | 34        |

|          |   |           |
|----------|---|-----------|
| 3.4      | <i>The mediating power of YouTube's algorithms</i>        | 36        |
| 3.5      | <i>Research questions</i>                                 | 38        |
| 3.6      | <i>Method</i>   | 38        |
| 3.7      | <i>Results</i>  | 40        |
| 3.8      | <i>Conclusion</i>   | 43        |
| <b>4</b> | <b>Structuring Power</b>                                  | <b>46</b> |
| 4.1      | <i>Introduction</i>                                       | 46        |
| 4.2      | <i>Influencers: politically relevant strategic actors</i> | 48        |
| 4.3      | <i>Algorithmic power</i>                                  | 49        |
| 4.4      | <i>Research questions</i>                                 | 51        |
| 4.5      | <i>Method</i>   | 52        |
| 4.6      | <i>Results</i>  | 54        |
| 4.7      | <i>Conclusion</i>   | 61        |
| <b>5</b> | <b>Algorithmic Power, Trust, and Legitimacy</b>           | <b>63</b> |
| 5.1      | <i>Algorithmic power</i>                                  | 63        |
| 5.2      | <i>Algorithms, trust, and legitimacy</i>                  | 64        |
| 5.3      | <i>Implications</i>                                       | 65        |
|          | <i>References</i>   | 70        |
|          | <i>Index</i>  | 82        |

# Figures

|  |    |
|--|----|
| 1.1 Annual number of social science publications about news algorithms, social media algorithms, filter bubbles, and echo chambers | 5  |
| 2.1 Who do you think is best at selecting the following type of content?   | 23 |
| 2.2 Should humans or algorithms make decisions in different areas of public life?  | 28 |
| 2.3 Should humans or algorithms decide which news you receive? (Generational differences)  | 29 |
| 3.1 YouTube search results after searching for videos about autism using six different search terms                                | 40 |
| 3.2 YouTube recommendations after five types of videos about autism  | 42 |
| 4.1 The relation between Influencers and Instagram's algorithms as a balance of power  | 56 |



# Table

2.1 Algorithm aversion experiment

26

# Acknowledgments

One of the central arguments in this book is that algorithmic gatekeeping can not be understood in isolation since it is the result of the interaction between the algorithm and its surroundings. The same can be said for the writing of this book, which is the result of my interactions with generous colleagues and institutions.

First of all, I would like to thank the Independent Research Fund Denmark, which generously supported my research on algorithmic gatekeepers under grant 7015-0090B and the Carlsberg Foundation which supported the writing of this book and Open Access Funding with monograph grant CF19-0544. Kristina Bøhnke, Annette Schmidt, and research support at the SDU's Faculty of Business and Social Sciences provided very professional support in the several rounds of funding application. A special thanks go to Erik Albæk, Claes de Vreese, and Michael Latzer for their help in the application process.

The Center for Journalism and the Department of Political Science and Public Management at the University of Southern Denmark (SDU) provided a good base to conduct my research. Many thanks to the following students without whose help it would not have been possible to conduct the empirical analyses in the book: Anne-Marie Fabricius Markussen, Camilla Lund Knudsen, Freja Møkkelbost, Julie Holmegaard Milland, Kasper Viholt Nielsen, Kirsten Vestergaard Prescott, Kristina Finsen Nielsen, Kristine Tesgaard, Ninna Nygård Jæger, Sasha Dilling, Selma Marthedal, Sofie Høyer Christensen, and Thor Bech Schrøder. Tina Guldbrandt Jakobsen managed the employment of these students.

Some of the results in Chapter 2 were earlier presented in the report MFI undersøger Synet på nyhedsalgoritmer. Ralf Andersson, Didde Elnif, and Morten Skovsgaard provided valuable feedback on this report which also helped shape the argument in Chapter 2.

Drafts of the manuscript were discussed in the Journalism Research group at SDU. Thanks to research group members Camilla Bjarnøe, Christina Pontoppidan, David Hopmann, Didde Elnif, Erik Albæk,

Fransiska Marquart, Jakob Ohme, Joe Bordacconi, Jonas Blom, Kim Andersen, Lene Heiselberg, Lisa Merete Kristensen, Majbritt Christine Severin, Morten Skovsgaard, Morten Thomsen, Peter Bro, Rasmus Rønlev, and Sofie Filstrup for their challenging and constructive feedback. In the last stage of the writing process, I discussed some of the central ideas with colleagues at the Digital Democracy Centre at SDU. I am sure this Centre will be a place for continuous conversation around the central themes of this book.

During the writing of this book, I dove into the emerging research literature on the topic of algorithmic gatekeeping, which was highly inspirational. Here, the work of Nick Diakopoulos, Taina Bucher, Axel Bruns, Tarleton Gillespie, Kevin Munger, and Joseph Philips deserves special mention.

Before the pandemic, it was possible for me to visit three research institutions where I worked on data collection and presented selected findings from this book. From ASCOR at the University of Amsterdam, I would like to specially thank Rens Vliegthart and Mark Boukes. Bert Bakker generously shared his knowledge on pre-registration. From Södertörn University, I would like to thank Ester Appelgren, Mikolaj Dymek, Gunnar Nygren, and Nina Springer. From my stay at the Leibniz Institute for Media Research, Hans-Bredow-Institut (HBI), and University of Hamburg, I would like to thank Cornelius Puschmann, Jonathon Hutchinson, Jessica Kunert, and Michael Brüggemann.

I would also like to thank the conference and workshop organizers and participants who allowed me to test early versions of the arguments presented in this book. Here, I would like to highlight in particular the 2018 ECREA pre-conference Information diversity and media pluralism in the age of algorithms, hosted by Edda Humprecht, Judith Möller, and Natali Hellberger; the ECREA Journalism Studies Section Conference 2019 convened by Folker Hanusch and the Journalism Studies Section at the University of Vienna; the ECREA Political Communication Section Conference 2019 hosted by Agnieszka Stępińska and the Faculty of Political Science and Journalism at the Adam Mickiewicz University in Poznań; the Knowledge-Resistance-workshop in 2020 upon invitation by Jesper Strömbäck of the University of Gothenburg; and an Augmented Journalism network-webinar in 2020, organized by Carl-Gustav Linden, Mariëlle Wijermans, and Andreas L Opdahl.

A further word of gratitude goes to Routledge. I specially want to thank Bob Franklin, Editor of Routledge's *Disruptions: Studies in Digital Journalism*-series for his confidence and encouragement, as well as Hannah McKeating and Elizabeth Cox for their guidance in the publication process. Three anonymous reviewers provided valuable input to the book.

A large part of the work on this book took place during the Covid pandemic. Personal thanks go to Anna Batog, Niels Lachmann, Thomas Paster, and my family in The Netherlands for their support during this time. Head of Department Signe Pihl-Thingvad and Vice-Head Morten Kallestrup are thanked for allowing me to work at the university during the lockdown. I want to thank Dean Jens Ringsmose for his encouraging words to the staff of the faculty during this period. A special word of thanks goes to everyone who reached out to me during the lockdown. That was very much appreciated!



# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

# 1 Introduction

## 1.1 Algorithmic power

The news and public information which citizens see and hear are increasingly influenced by algorithms. Algorithms are “encoded procedures for transforming input data into a desired output, based on specified calculations” (Gillespie, 2014, p. 167). Such automated procedures fulfill a gatekeeping role and automatically select, sort, and prioritize public information (e.g., Bucher, 2018; Diakopoulos, 2019; Kruikemeier et al., 2021; Napoli, 2014; Thurman et al., 2019; Wallace, 2018).

News and social media algorithms are powerful forces (see Bucher, 2018; Diakopoulos, 2019; Gillespie 2014; Just and Latzer, 2017). Worldwide, people get more and more of their political information through algorithm-controlled online social media, such as Facebook, Google, YouTube, or Instagram. This is especially true for younger generations (Newman et al., 2019). When algorithms act as gatekeepers, they impact on the information which the audience receives and on the way people communicate. Prioritization and recommendation algorithms affect exposure to counter-attitudinal views by either narrowing (Bakshy et al., 2015) or broadening the perspectives which the audience encounters (Helberger, 2019). Replacing human-curated content with algorithmically curated news, like in Apple News, affects exposure to soft news about celebrities (Bandy and Diakopoulos, 2020a). The information which algorithms on social media platforms show can affect voter turnout (Bond et al., 2012) or the persuasiveness of misinformation (Bode and Vraga, 2015). Recommendation algorithms on platforms like TikTok play an important role in social movements, as they influence the reach of calls to action (Bandy and Diakopoulos, 2020b). On the one hand, algorithms on social media platforms can enhance the spread of extremist views (e.g., Massanari, 2017). On the other hand, algorithms are used to moderate the online debate and censor inappropriate speech (e.g., Cobbe, 2021). Changes in social media algorithms can limit or boost traffic to legacy news outlets

## 2 Introduction

(Trielli and Diakopoulos, 2019) and influence the type of content which these outlets produce (Diakopoulos, 2019, pp. 178–180). Automation in news production transforms the news-making process (e.g., Linden, 2017; Thurman, 2011; Van Dalen, 2012; Wu et al., 2019) and affects the relation between journalists and the audience (Dörr and Hollnbuchner, 2017).

The power of news and social media algorithms is central to this book which does not consider algorithms as an autonomous force that exercises a uni-directional influence on society at large. This book also tries to steer away from treating algorithms as a scape goat for negative societal developments such as radicalization or polarization. Instead, its purpose is to understand how the power of gatekeeping algorithms emerges and is exerted in the interaction with other actors. The first aim of the book is to make the power of algorithmic gatekeepers tangible. It documents how algorithms affect the content which the audience is exposed to (Chapter 3) and how they shape the behavior of professional communicators (Chapter 4). This shows how algorithmic power is situated in the interplay between platforms, audiences, and professional communicators. To be justified, such power needs to be accepted by the general population and professional communicators. Therefore, the second aim of this book is to study two crucial antecedents for the acceptance of algorithmic power: trust and legitimacy. It analyzes trust and the acceptance of algorithmic gatekeeping among the general population (Chapter 2) and evaluates the legitimacy of algorithmic power (Chapter 4).

### 1.2 Conceptualizing algorithmic gatekeeping

Journalistic gatekeeping has been defined as “the process of selecting, writing, editing, positioning, scheduling, repeating and otherwise massaging information to become news” (Shoemaker et al., 2009, p. 73). Building on this general definition, algorithmic gatekeeping can be understood as the influence of automated procedures on the process of selecting, writing, editing, scheduling, repeating, and otherwise massaging information to become news. These automated procedures are applied by news organizations, but equally important, they are a defining feature of social media platforms. This book deals with the influence of automated procedures on the production of news, as well as on the selection of news.

Technically, algorithms can be described as programmed procedures that are followed by a computer system. Bucher (2018) distinguishes between deterministic and machine learning algorithms. Deterministic algorithms follow the same predefined steps and do not adapt. Machine learning algorithms on the other hand continuously adapt the way they function as they learn to optimize their performance. More broadly, algorithms are a symbol for the increasing automation of decision-making in

public life. The “*algorithmic turn*” (Napoli, 2014) in gatekeeping which is central in this book is part of a larger societal trend, where more and more societal tasks are automated (e.g., Kitchin and Dodge, 2011; Lessig, 1999; Steiner, 2012; see Chapter 2). While algorithms’ role in gatekeeping is the main focus, the book ties into the growing journalism research field studying news automation more broadly (e.g., Diakopoulos, 2019; Linden, 2017; Thurman et al., 2019). In line with this literature on news automation, the book does not deal narrowly with technical lines of computer codes but understands algorithms which influence gatekeeping as “heterogeneous and diffuse sociotechnical systems, rather than rigidly constrained and procedural formulas” (Seaver, 2017 in Bandy and Diakopoulos, 2020a, p. 38).

To understand the process and effects of algorithmic gatekeeping, algorithms need to be studied in their interplay with other actors, most notably professional communicators, and the general public.

Journalistic gatekeeping research goes back to White’s (1950) study of how an individual journalist acts as gatekeeper and decides which information makes it through the gate and becomes news, and which information is kept out. Building and expanding on this pioneering research, later studies have identified the influence of the organizational context and extra-media influences on the decisions made by individual journalists. Journalistic gatekeeping has come to mean more than only selecting news, as later definitions also include the writing and editing of news (Shoemaker et al., 2009). The increasingly important role of automation in the news-making process and the role of social media platforms in the distribution of news have made the gatekeeping process even more complex. Gatekeeping decisions are made by diverse sets of actors, most notably journalists and other strategic professionals, individual amateurs as well as algorithms (see Wallace, 2018). Gatekeeping has become decentralized and is in many instances the consequence of the interaction between different actors. This makes it difficult to isolate the consequences of algorithms in the gatekeeping process or to use the language of directional causality when discussing their influence.

Where gatekeeping research traditionally studied gatekeeping within journalistic news organizations, other professional communicators play an increasingly important role in gatekeeping in today’s media environment. On algorithmically controlled social media platforms, audiences encounter diverse kinds of information about public affairs side-by-side, ranging from entertainment to hard news, and from conspiracy theories to the correction of misinformation. A clear example of these new sources of information are Influencers who post about public affairs on social media (see Chapter 4). They exemplify how technological, societal, and media-internal developments have blurred the boundaries around the journalistic profession (e.g., Lewis, 2012). Due to these faded boundaries, studies of



## 4 Introduction

algorithmic gatekeeping need to consider professional communicators in a broad sense.

This book has a normative starting point as it is interested in the effects of automated procedures on news, understood as information about current affairs which allows citizens to participate in public life and which supports the functioning of the public sphere (e.g., Strömbäck, 2005). As described in the introduction to this chapter, several concerns have been raised about the influence of algorithms on the news which reaches the public. The empirical analyses presented here tie into three broader questions about the democratic impact of algorithmic gatekeeping.

*First*, do algorithms foster the spread of misinformation on social media platforms? Misinformation can negatively affect social cohesion, can undermine the legitimacy of election and democratic governance, or can be directly damaging to people's health (e.g., Lewandowsky et al., 2012). Chapter 3 analyzes the role of YouTube's recommendation algorithm in the spread of misinformation.

*Second*, do algorithms prioritize information which confirms people's worldview rather than present them with various perspectives? This would have negative democratic consequences since it could make people's beliefs more polarized and undermine a unified public sphere (Stroud, 2010). The algorithm audit presented in Chapter 3 studies whether watching a video with misinformation leads YouTube's algorithm to recommend more videos confirming this perspective.

*Third*, do algorithms undermine deliberation on social media platforms? For a well-functioning public sphere, it is important that people engage in debate and are willing to express their own point of view. Research has shown that the general population is less willing to speak out on algorithmically driven social media platforms than in face-to-face communication (Neubaum and Kramer, 2018). Chapter 4 analyzes how Influencers perceive their power vis-à-vis Instagram's algorithms to shed light on the effects of these perceptions on their public opinion expression.

### 1.3 Technological determinism: hard and weak

This book studies how the gatekeeping power of algorithms takes shape in the interaction of algorithms with other actors. This approach can be contrasted with a technological deterministic view which often dominates discourse around the democratic consequences of algorithmic gatekeeping. Technological determinism sees technological developments as an autonomous force, which creates negative societal effects (e.g., Shade, 2003). While the benefit of this view is that it raises awareness for the role of media technology and its consequences, in its purest form (hard technological determinism) the role of technology gets overemphasized (see Smith and Marx, 1994). It is a technology-centric approach with limited

attention to social forces which shape the technology, allow for the growth in its use, and mediate the societal effects of technology. Also, hard technological determinism pays limited attention to variations across societal groups or potential positive effects (e.g., Shade, 2003). Technology often becomes a scape goat in this discourse.

A review of the literature on filter bubbles and echo chambers illustrates the limits of hard technological determinism for the understanding of algorithmic gatekeeping. Debate about the power of algorithms is strongly influenced by fears around filter bubbles and echo chambers, despite empirical evidence challenging these fears (see Figure 1.1). In 2011, Eli Pariser published the book *The Filter Bubble: What the Internet is Hiding from You* coining the term filter bubble. According to the filter bubble argument, (1) people hardly get exposed on social media to information that challenges their worldview, and (2) this is due to algorithms that prioritize content that is similar to content which people have previously interacted with, over more diverse content. Concretely, when people with different interests type in the same search term in Google, Google's search algorithms would return completely different information. This should reflect them having interacted with different types of content previously. A similar effect should be present in Facebook's news feed, where people with different political identities see news and political information with a different political outlook. As a consequence of these online filter bubbles, the range of information to which people become exposed would over time get narrower and

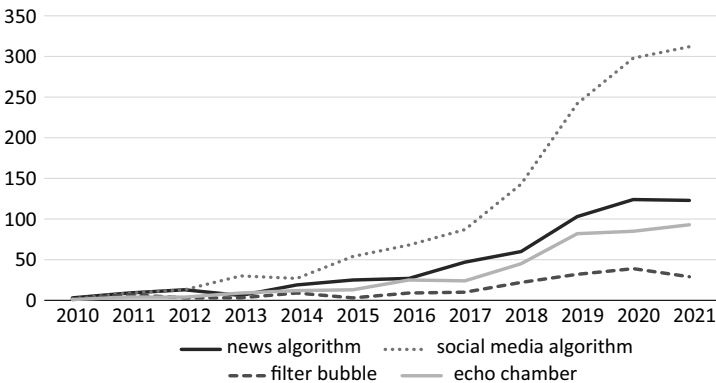


Figure 1.1 Annual number of social science publications about news algorithms, social media algorithms, filter bubbles, and echo chambers.

Note: Based Social Sciences Citation Index (SSCI); numbers of annually published articles with these topics.

narrower, which in turn creates self-reinforcing spirals. By prioritizing information that confirms people’s worldview, algorithms would amplify confirmation biases: people’s tendency to seek out information in line with pre-existing beliefs (e.g., Nickerson, 1998), thus making them less open to and curious about views that challenge their preconceptions.

Filter bubbles are closely related to echo chambers. As Nguyen (2020) argues, it is important to distinguish between filter bubbles and echo chambers, although the two terms in practice often are used interchangeably (see also Bruns, 2019). While a filter bubble refers to a situation where individuals are isolated from information that challenges their worldview, an echo chamber is a group-level phenomenon where the members of the group share the same worldview and reinforce this worldview by on the one hand continuously confirming this to one another and, on the other hand, discrediting voices with another perspective. Such groups can be found on social media such as Facebook or YouTube, for example, around incels, climate change deniers, or anti-vaxxers. An “echo chamber is a social epistemic structure from which other relevant voices have been actively excluded and discredited” (Nguyen, 2020, p. 141). The term echo chamber was popularized by Cass R. Sunstein in the first decade of the 21st minutes focusing primarily on the internet and the blogosphere. In his 2017 book *#Republic*, Sunstein expanded the discussion of echo chambers to social media. As with filter bubbles, critics have pointed to algorithms behind social media as a villain and cause of echo chambers. On Facebook, algorithms might drive people toward groups of like-minded individuals and keep them entrapped in these groups (e.g., Sunstein, 2017).

In their most extreme form, such concerns about algorithm-driven filter bubbles and echo chambers reflect the central characteristics of technological determinism: algorithms are seen as an autonomous force leading to societal consequences such as radicalization and polarization, with limited attention to variation in the way different societal groups use the technology or to the way societal trends interact with the effects of algorithms.

Empirical studies of filter bubbles and echo chambers challenge this perspective (e.g., Flaxman et al., 2016; Geiß et al., 2021; Nechushtai and Lewis, 2019; Puschmann, 2019). A thorough discussion of relevant studies led Bruns (2019, p. 96) to conclude that “mainly, the debate around these concepts and their apparent impact on society and democracy constitutes a moral panic.” Based on their review of the literature, Zuiderveen Borgesius et al. (2016, p. 1) answered the question *Should we worry about filter bubbles?* with the conclusion that “at present there is little empirical evidence that warrants any worries about filter bubbles.” They list the following reasons why worries about filter bubbles might be overstated.

- Without personalization algorithms, people tend to limit the range of views they expose themselves to. Selectively exposing oneself to information in line with one's preferences is not a new phenomenon. It has existed long before social media and algorithms.
- Social media and the internet also offer new opportunities for encountering surprising information which widens one's worldview.
- At the time of writing, personalization algorithms were not technically advanced enough to give each audience member information in line with what they want.

Several empirical studies indeed support the observation that algorithms do not have the strong personalization effects which critics suggest. Haim et al. (2018), for example, studied the personalization effect of Google News and found very limited differences in search results depending on what people had previously searched for. Möller et al. (2018) studied how different types of news recommendation systems on a newspaper website affect diversity of recommended content. They found no evidence that basing recommendations on user history leads to limited diversity in recommendations.

Two studies by the Reuters Institute for the Study of Journalism not only challenge the idea that algorithm-driven social media narrow the range of views to which the audience is exposed (Fletcher and Nielsen, 2018a, 2018b). They even showed evidence suggesting the exact opposite: on algorithm-driven social media, people get exposed to more diverse information than people who do not use such social media at all. In one study, they illustrated empirically that people who use search engines like Google to find news are exposed to more diverse and more balanced news sources than people who do not use search engines. They attribute this to what they call automated serendipity, “forms of algorithmic selection that expose people to news sources they would not otherwise have used” (Fletcher and Nielsen, 2018a, p. 977). In another study, Fletcher and Nielsen (2018b) studied whether people get exposed to news sources without explicitly looking for them (so-called incidental exposure) when using social media channels. They concluded that people who use Facebook, YouTube, and Twitter primarily for entertainment, get exposed to more online news channels than people who do not use social media at all. Boulianne et al. (2020) likewise challenged the idea that exposure to algorithm-driven news sources leads to partisan polarization because people are no longer exposed to opposing views.

Some studies illustrate that algorithms can foster selective exposure, but that these effects cannot be attributed to algorithms alone. A study into exposure to cross-cutting information on Facebook by Bakshy et al. (2015) found that Facebook's algorithms decreased content diversity. However, this effect of algorithms on content diversity was remarkably smaller than the effects of people's own tendency to self-select into

networks with people with similar political preferences (Bakshy et al., 2015). Dubois and Blank's (2018) argue that the effect of algorithms should be studied in the context of the broader media environment, not focusing on single social media channels. In a representative survey among the British population, they found that only a minority of the respondents never or rarely disagree with content they encounter on social media and never or rarely check diverse political sources online. They estimate that only 8% of the population might be in danger of being caught in an online echo chamber, and these people might still be exposed to diverse information offline through personal conversations with friends and family. Bodó et al. (2019) revealed that there are large parts of the population who prefer diversity-enhancing automated news recommendations over diversity-limiting news recommendations. These people are less likely to fall into a negative spiral where their news preferences combined with news recommendations limit the diversity of the information they are exposed to. On the other hand, there are parts of the population who are less interested in diversity and who might get caught in "diversity reducing feedback loops" (Bodó et al., 2019, p. 206). This research shows that algorithmic gatekeeping can have distinct effects across various societal groups.

In sum, empirical research challenges the hard technological deterministic view of the negative influence of algorithmic gatekeeping on filter bubbles and echo chambers. Instead, this research is more supportive of a softer version of technological determinism, acknowledging that the influence of algorithmic gatekeeping should be studied in a broader context; that the influence is not necessarily negative; that algorithmic gatekeeping affects people differently; and that it is the way social actors and social factors shape and use these algorithms which determine their power. This soft technological determinism is also supported by other algorithmic gatekeeping studies beyond research into filter bubbles and echo chambers. Bandy and Diakopoulos (2021) for example showed that Twitter features more junk news (understood as news which does not help the audience to fulfill its citizen role) when Twitter's timeline is driven by algorithms, compared to when it is sorted chronologically. Part of this can be explained by the specific working of the algorithm, but they conclude that the role of the algorithm is "a fairly minor, supporting role in shifting media exposure for end users, especially considering upstream factors that create the algorithm's input – factors such as human behavior, platform incentives, and content creation" (Bandy and Diakopoulos, 2021, p. 1).

## **1.4 Theoretical considerations**

### *Supply-and-demand framework*

One theoretical starting point to understand how algorithmic gatekeeping is situated in the interplay between platforms, professional communicators,

and audiences is offered by Munger and Phillips' (2022) supply-and-demand framework. Munger and Phillips (2022) list factors which explain the emergence of communities around right-wing content on YouTube, or more broadly, the emergence of alternative media clusters. This framework helps to see the effects of social media algorithms in context and to understand how they interact with other technical affordances and audience desires and behaviors.

On the supply side, Munger and Phillips (2022) highlight the affordances of the YouTube platform which are favorable to the growth of communities around right-wing content. The possibility to discover content through video suggestions by YouTube's algorithms is one of these supply-side features of the YouTube platform which fosters alternative media clusters. YouTube's algorithms determine which videos show up in a search, which videos are suggested on YouTube's starting page, and which videos are suggested to watch next after watching a video. There is a large degree of similarity between videos which have been watched previously and what YouTube's algorithms suggest watching next (e.g., Airoidi et al., 2016; O'Callaghan et al., 2015). This could foster the growth of communities around specific types of content. Central in the supply-and-demand model is the notion that YouTube's algorithms are far from the only factor driving the growth of right-wing communities. There are other affordances which make the growth of such communities more likely: YouTubers can actively influence whether their videos will be discovered by adding descriptive keywords as metadata (tagging) or by using clickbait-style headlines. Furthermore, video is a form of content which is relatively cheap to produce and can convey more impacting emotions than text. The specific monetization structure of YouTube favors the building of communities and the continuous stream of new content. Popular YouTubers can sell advertising around their videos, offer channel membership, or create content specifically for YouTube Premium subscribers. YouTube can demonetize videos around sensitive topics, but this still leaves YouTubers with the opportunity to use alternative crowdfunding sources to get paid.

Isolating YouTube's algorithms as the driving force behind communities around right-wing content on the platform would also overlook the importance of audience demand for such content (Munger and Phillips, 2022). On the demand side, we find factors such as a loss of identity due to societal changes, or feelings of disconnection and isolation in the offline world, which might drive people to search for a feeling of community on social media platforms. They may develop para-social relations with YouTubers, who substitute for friends and meaningful connections which the audience lacks in real life. The video format seems to be particularly attractive to audiences looking for such connections, as it is less demanding to process than for example reading long blog posts.

While Munger and Phillips (2022) use the supply-and-demand model to explain the presence of right-wing communities on YouTube, the model can likewise be used to understand that algorithms are only one factor among several factors influencing the spread of content on social media platforms. This framework is a useful antidote against a hard technologically deterministic understanding of the power of news and social media algorithms.

### *The structuring power of algorithms*

Discussions on how to perceive the power of algorithms mirror similar discussions about the influence of the media on politics (see also Klinger and Svensson, 2018). Here, two broad perspectives can be distinguished (e.g., Van Aelst et al., 2008). First, an actor perspective, where power is seen as “a successful attempt by A to get B to do something that he would otherwise not do” (Dahl, 1957, p. 203). Here, being in power is seen as having the upper hand in an asymmetric relation. In the second perspective, the power of the media is assessed from a structural perspective. Here, the focus is more on the way the media set the stage on which politics is performed and the conditions within which politicians have to maneuver. According to this perspective, the media do not determine specific political outcomes, but instead determine the unwritten rules which politicians have to adapt to when they pursue their goals.

In line with a soft technological deterministic view on algorithmic gatekeeping, the power of algorithms can be better understood from a structural rather than actor perspective. Following Dahl’s definition of power, seeing the algorithm–human relation as a power balance would require the algorithm to purposefully force users to do something against their will. To purposefully force others to do something would require agency: the ability to autonomously act out of free will. However, it is problematic to talk about an algorithm as having agency in the same way as humans would have agency. Klinger and Svensson (2018) stress that algorithms lack evaluative and reflective capacities. Even though algorithms might be self-learning, this does not mean that they make deliberative decisions on what is desirable or the right trajectory of action when faced with unexpected situations. Without such evaluative and reflective capacities, algorithms lack the intentionality that defines human agency. Bucher (2018, pp. 50–54) underlines the fact that the way algorithms operate is the result of the interaction of different actors and forces, including the people who designed and programmed the algorithm, the algorithm itself as well as the users whose behavior also affects the process and outcomes of the algorithms. The assemblage of actors which determine the working of the algorithm makes it difficult to see algorithms as possessing agency.

Instead of studying the power of social media algorithms for an actor perspective, it makes sense to see these algorithms as a structuring power. The algorithms help social media users to obtain specific resources (such as access to followers) or fulfill certain needs (e.g., feeling part of a community). Consequently, these algorithms structure how social media users interact with each other and the social media platform. In doing so, the algorithms do not determine the exact decisions which the social media users make but set the ground rules which the social media users have to adhere to if they want to be in the best position to obtain these resources or fulfill these needs. Bucher (2012, p. 1165) made this concrete by showing that “algorithmic architectures dynamically constitute certain forms of social practice around the pursuit of visibility.” Facebook’s algorithm determines which posts are shown to whom and in which order and thus whose views and output are visible. In order to be visible, Facebook users are forced to post regularly and have an incentive to like and comment on other people’s posts. As we shall see in Chapter 4, Instagram’s algorithms have a similar shaping power over Influencers.

### 1.5 Legitimacy and trust

Algorithmic gatekeeping, which positively affects the quality of the public sphere and the ability of people to fulfill their citizen role, needs to be legitimate and trusted. This is particularly relevant because algorithms are opaque: their role is often invisible. Even when algorithms are visible their working remains obscure (Bucher, 2018). Since people cannot monitor how algorithms work, they have to accept their power based on a perception that they are just and fair (in other words see them as legitimate) and they have to be willing to be vulnerable to the actions of the algorithms (in other words trust them).

Legitimacy can be defined as “a generalized perception or assumption that the actions of an entity are desirable, proper and appropriate within some socially constructed system of norms, values, beliefs, and definitions” (Suchman, 1995, p. 574). When people see the power of algorithmic gatekeepers as legitimate, they accept the role which algorithms play in the public sphere and see this role as justified and rightful. A distinction can be made between objective and subjective approaches to legitimacy (Hurd, 2012). Following the objective approach, we can discuss whether algorithmic gatekeepers support the public sphere according to some normative standards or specific views on democracy. Following the subjective approach, we can study whether people perceive algorithmic gatekeeping as acceptable. From this perspective, legitimacy “exists only in the beliefs of an individual about the rightfulness of rule” (Hurd, 2012). Ideally, there is a close link between objective and subjective legitimacy and the general public and communication professionals perceive algorithmic gatekeepers



as just when they indeed support the public sphere. If power is not perceived as legitimate, people might still adhere to it, but motivated by possible personal gains or lack of alternatives rather than because they believe the power is just and fair. Algorithmic power without legitimacy might lead to resistance against the algorithms and could ultimately be an incentive for people to leave a social media platform (Velkova and Kaun, 2021). While sources of legitimacy of media institutions have been extensively studied (e.g., Skovsgaard and Bro, 2011; Van Dalen, 2019b), less is known about which criteria people use to assess whether the power of algorithms is legitimate. Van Dalen (2019b) summarizes five sources of legitimacy for media institutions: the character of the professionals (in other words, whether journalists are perceived as good people), connections to other legitimate institutions (in other words, whether media institutions can piggyback on its relation with other institutions, which are seen as legitimate, such as the state or experts), ethical standards (such as the objectivity norm), constituents (in other words, the target group who benefit from the actions of media institutions, such as the public at large), and beneficial outcomes (such as the important democratic roles which the media fulfill by holding governments accountable or informing the electorate). Chapter 4 will describe similar legitimacy criteria for social media algorithms.

Trust can be defined as “the willingness of a trustor to be vulnerable to the actions of a trustee based on the expectations that the trustee will perform a particular action, irrespective of the ability to monitor or control that other party” (Mayer et al., 1995, p. 712). Given the enormous amount of information available online, it would be a challenging task for the public to individually go through all sources and select relevant information. Similar to the trusted news media (see Coleman, 2012), trusted algorithmic gatekeeping can thus reduce complexity and help to minimize efforts which allows the audience to fulfill their citizen role. Gatekeeping algorithms are complex and get continuously updated. Thus, social media users do not have the possibility to monitor or control the algorithms. When using social media and automated news, the users are vulnerable to the actions of the algorithms. They are dependent on the algorithms to on the one hand produced, select, and prioritize information which is relevant for them, and on the other hand, make sure that they do not miss essential information or get wrongly informed. To trust algorithms thus requires humans to believe that the algorithms do a good job in selecting, prioritizing, and creating information.

In order to invoke trust, social media platforms often underline the fact that their algorithms lack subjectivity and personal biases (Gillespie, 2014, p. 179): “The careful articulation of an algorithm as impartial (even when that characterization is more obfuscation than explanation) certifies it as a reliable sociotechnical actor, lends its results relevance

and credibility, and maintains the provider's apparent neutrality in the face of the millions of evaluations it makes." However, as we shall see in Chapter 2, trust in social media algorithms is not only determined by platforms and the working of their algorithms (the object of trust or trustee), but as much by the social media users who use them (the trustors). Trust in algorithms is influenced by people's understanding of how these automated computer systems work, which in turn is strongly shaped by comparisons with how the human mind works (Shariff et al., 2017). Such comparisons could either result in attributing human characteristics to the algorithmic systems, or to sharply distinguish between tasks which are suitable for machines on the one hand, and tasks which should be left to humans on the other (e.g., Lee, 2018). Research into algorithmic forecasting has for example shown that people are often unwilling to rely on the advice of algorithms, even when such advice outperforms human forecasting (Burton et al., 2020). Thus, the promise of algorithmic objectivity does not necessarily translate into public trust in these algorithms. This makes it relevant to study how people perceive news algorithms and how this is related to trust.

For a deeper understanding of why legitimacy and trust are relevant for algorithmic gatekeeping, it is helpful to look at why legitimacy and trust matter for journalistic gatekeeping. The press is often described as the fourth estate since it performs the important democratic role of informing citizens and holding government accountable. Still, the basis of authority of journalism is different from the basis of authority of the first three powers: the workings of the legislative, executive, and judiciary are guided by clear rules and regulations which are laid down in constitutions and democratic procedures (e.g., Cook, 1998, p. 109). Journalists' work is rather based on unwritten social patterns of behavior and routines which are followed across organizations and remain rather stable over time (Cook, 1998). In the Western world, journalists' work is protected by freedom of expression and their editorial independence is guaranteed in press laws. Other professions like medics or lawyers have stricter guidelines and inclusion and exclusion mechanisms into the profession than journalism which is in many Western countries not a protected title (Skovsgaard and Bro, 2011). On the one hand, this gives journalists and news organizations freedom to organize their work according to their own standards rather than standards and guidelines set by other institutions. On the other hand, this also comes with the downside that journalists cannot claim to adhere to written rules and procedures when they want to justify their power and claim legitimacy. Trust in the mainstream media is a necessary condition for the legitimacy of the press. If the press is not trusted by the general public, it is easier for politicians to ignore criticism from journalists or obstruct efforts by the media to hold them accountable (e.g., Van Dalen, 2019a).

Like journalistic gatekeeping, algorithmic gatekeeping has long been shielded from government regulation. European regulation stipulates that courts or regulatory authorities may not impose on journalists and news organizations certain ways of conducting their work (Helberger et al., 2020). This gives news organizations freedom to use algorithms in their reporting “as long as doing so does not conflict with the enjoyment of fundamental rights of others” (Helberger et al., 2020). An argument can be made that automated prioritization and selection of information on social media platforms are also forms of (quasi-)editorial judgment (Leerssen, 2020). Still, the rights and freedoms also come with responsibilities, in particular to recognize and deal with the negative consequences of algorithmic gatekeeping on fundamental rights and the functioning of the public sphere. Examples of such negative consequences include the circulation of hate speech, interference with election results, or the spread of health-related disinformation.

Concerns about the rapid implementation of algorithms in society and debate about the negative ethical implications have led to the formulation of several ethical guidelines. Hagendorff’s (2020) analysis of 22 ethical guidelines for AI however showed that such guidelines are generally non-binding and lack mechanisms of enforcement. This limits the role of ethical guidelines in making the general public and professional communicators accept algorithmic power. Concerns about negative consequences of algorithm-driven social media platforms combined with the limited effects of ethical guidelines have led to legislative initiatives to counter democratic risks posed by these platforms. Examples of this are the European Union’s AI Act, Digital Services Act, and Digital Market Acts (e.g., Helberger and Diakopoulos, 2022). These Acts demand platforms to be transparent about the algorithms used for recommendations; oblige large platforms to allow independent audits of their measures to prevent abuse of their systems; and force them to provide researchers access to data (European Commission, 2022a). The implications of such regulation for trust in algorithmic gatekeeping will be addressed in the final chapter.

## 1.6 Outline and content of the book

Three empirical chapters further our understanding about the power of news algorithms by focusing on, respectively, the general public, platforms, and professional communicators.

Chapter 2 analyzes how the *general public* perceives and approves of news algorithms. Following an overview of the antecedents of algorithm approval found in the literature, the chapter reports the results of three surveys among the general Danish population. The results show that trust in news algorithms and approval of news algorithms is low. Behind

this seems to lie an implicit comparison with human journalists, who are seen as more objective and neutral, and thus more trustworthy. Still, the general population is more approving of algorithms playing a role in news selection than in other areas in public life. This is especially true for younger generations and when humans are involved in the process.

Chapter 3 analyzes the mediating power of algorithms on *social media platforms*. Mediation refers to “the relaying of second-hand (or third-party) versions of events and conditions which we cannot directly observe for ourselves” (McQuail, 2010, p. 83). Media organizations have always had a mediating role between the audience and broader society. Social media algorithms play a similar role. When they filter, prioritize, or censor certain information, they influence which views and perspectives are most likely to reach the audience. This mediating power is analyzed in a case study of the presence of different types of misinformation in YouTube search results and recommendations around autism. The chapter shows that YouTube’s search and recommendation algorithms do not recommend misinformation which is directly harmful. Still, some evidence was found that the algorithms suggest and recommend less harmful misinformation. These results can not only be attributed to the algorithms since they are also shaped by other supply as well as demand factors on YouTube. Still, they raise important questions about how social media platforms should handle their responsibility and what duty and opportunity they have to use algorithms to downplay or correct misinformation.

Chapter 4 analyzes how *professional communicators* like Influencers perceive the power balance between themselves and the algorithms of the platforms on which they communicate. A study of how professional communicators perceive and react to the power of Instagram’s algorithms shows how dependent they are on these non-transparent algorithms for access to user engagement. The algorithms limit their autonomy. This can have serious democratic consequences such as a tendency toward mainstream perspectives and limited room for doubts or nuance. Against this background, the chapter lists criteria which professional communicators use to assess the legitimacy of algorithmic power.

The final chapter reflects on the findings of the preceding chapters in the light of the aims of the book. The chapter highlights lessons learned about the power, trust, and legitimacy of news algorithms and the challenges they pose for social media platforms and news organizations. Ultimately, concrete initiatives are proposed to strengthen the trustworthiness and legitimacy of news algorithms.

Empirically, these chapters build on the results of three representative surveys, two survey-embedded experiments, a quantitative content analysis of videos recommended by YouTube’s algorithms and a qualitative analysis of the way professional communicators relate to social media

algorithms. The surveys and experiments are conducted in Denmark, while the other analyses study English-language content. Thus, the empirical results should be understood in the context of Western media systems. Instead of analyzing much studied social media platforms like Twitter or Facebook, the manuscript focuses on YouTube and Instagram. As described in Chapters 3 and 4, these two platforms are used by large parts of the population, especially younger segments. They distinguish themselves from other social media platforms by their visual nature. On YouTube and Instagram, audiences encounter public information of all kinds, ranging from entertainment to hard news, and from conspiracy theories to the correction of misinformation. These are also the platforms where new actors like celebrities and Influencers compete for attention with mainstream outlets and expert sources. Thus, these platforms exemplify what the new media environment looks like, making them exemplary cases to study the role of algorithms in disseminating information and in affecting communication behavior. Furthermore, YouTube and Instagram actively change and update their algorithms to counter potential negative consequences. This provides a basis for discussing how social media platforms should responsibly deal with the power of their algorithms. Before narrowing the focus to these two platforms, Chapter 2 first maps generalized trust and approval of news algorithms among the general population.

## 2 Algorithm Aversion

### 2.1 Introduction

The previous chapter discussed the distinctive ways in which algorithms play an increasingly important role in selecting news and writing news (see Diakopoulos, 2019). How the public perceives such news algorithms affects how they interact with the information written and selected by these automated computer systems (Bucher, 2018; Rader and Gray, 2015). People often lack awareness of the role of algorithms in news production and selection (Eslami et al., 2015). Those members of the general public who are aware that news algorithms exist make sense of them based on simplified mental models of how the technologies function rather than a deep understanding of their working. These mental models are strongly influenced by two mental shortcuts. The first shortcut is the tendency to compare algorithms to humans (Hofstadter, 1995). The second shortcut is the machine heuristic (Sundar, 2008), the idea that decisions made by algorithms are neutral, since these automated computer programs follow prescribed rules. As we will see in this chapter, these heuristics and comparisons affect the way people perceive, trust, and approve of news algorithms. Based on representative surveys and two survey-embedded experiments, the chapter addresses the following questions: *How do people perceive the strengths and weaknesses of news algorithms compared to human journalists?* and *Do people trust and approve of news algorithms?*

To address these questions, this chapter first presents a review of the literature on news algorithm perceptions and approval. This review focuses on the comparisons between humans and algorithms and on the machine heuristic. Previous research on approval of news algorithms is placed in context by relating the findings to the broader literature of algorithm approval. Next, new empirical evidence shows the impact of machine heuristics on perceptions of strengths and weaknesses of news algorithms. The data further reveal an alarming lack of trust in news selected and written by algorithms compared to news selected and written

by humans. People however remain more positive toward algorithms playing a role in the selection of news than in other areas of public life. This is especially the case when algorithms and humans work together in selecting the news, and among younger generations. The final chapter of this book will return to these findings and point to human-algorithmic collaboration as a way to build trust toward news algorithms.

## 2.2 Making sense of news algorithms

A central feature of news algorithms which affects how they are perceived is their opacity; the lack of transparency and clarity around how they operate (Eslami et al., 2019). Many people lack awareness of the existence of algorithms which prioritize and select information, such as the Facebook News Feed curation algorithm (Eslami et al., 2015), or Yelp's review filtering algorithm (Eslami et al., 2019). Even when people are aware of the existence of news algorithms, there is a lack of understanding about the way such algorithms make decisions and a lack of awareness of their effects. Many college students who use Google to access news for example were not aware that search results are personalized (Powers, 2017).

People who are confronted with new technologies under uncertain conditions like the ones created by the opacity of the news algorithms, develop their own mental models to make sense of how these technologies function (Bucher, 2018, pp. 96–7). Such mental models are based on different types of input, ranging from representations in popular culture, public debate, and media coverage, to so-called folk theories grounded on personal everyday experiences with news algorithms (Bucher, 2018). Following DeVito et al. (2017, p. 3165), folk theories can be defined as “intuitive, informal theories that individuals develop to explain the outcomes, effects, or consequences of technological systems, which guide reactions to and behavior towards said systems.” DeVito et al. (2017) distinguish between operational theories and abstract theories. Operational theories relate to the specific working of algorithms and concrete rules which the algorithms follow (e.g., popular content will be prioritized on Twitter). Abstract theories on the other hand represent a generic idea of the function and broad role of algorithms on the platform, without specification of concrete mechanisms or specific outcomes (e.g., Twitter's algorithm affects a user's timeline).

These mental models are often based on (implicit) comparisons with humans. Research in robotics has for example documented that people often attribute human characteristics to automated systems, such as moods or emotions. This anthropomorphism helps people to make sense of the working of these abstract systems (Hofstadter, 1995; Zarouali, Makhortykh et al., 2021). These implicit comparisons can be more

favorable toward either the automated systems or humans performing similar tasks. When assessing the credibility of news and information online, people are faced with more uncertainty than when consuming news through more traditional channels like newspapers or television where the sender can clearly be distinguished. Under such uncertain circumstances, the technical affordances, such as whether there is the possibility for interactivity, are used as cues to assess the credibility of information (Sundar, 2008). The cues trigger heuristics or rules of thumb, which affect credibility assessments. One mental shortcut that affects how people think about news algorithms is the so-called machine heuristic (Sundar, 2008). The machine heuristic assumes that because a machine follows pre-defined rules it is less susceptible to subjectivity or unfair favoritism than humans are: “if a machine chooses the story, then it must be objective in its selection and free of ideological bias” (Sundar, 2008, p. 83). Sundar and Nass (2001) found that people rated the same stories as of higher quality when they were allegedly selected by computer systems rather than human journalists. The machine heuristic and presumed lack of bias might account for this result (see also Nielsen, 2016).

Contrary to what the machine heuristic suggests, news is perceived as less credible when it is allegedly written by a “robot reporter” compared to when a human journalist is portrayed as the author (Franklin Waddell, 2018). This can be explained by expectancy violation theory. Using the label robot could lead to high expectations about the objectivity of the content. When the actual content then does not live up to these expectations, people will be disappointed, and the credibility will be perceived as lower (Franklin Waddell, 2018).

This research by Sundar (2008), Sundar and Nass (2001), and Franklin Waddell (2018) suggests that the machine heuristic has different effects on the credibility of automatically selected news than on automatically written news (Franklin Waddell, 2018, p. 249). This could reflect that people understand news selection as a less creative and thus less human task than news writing. Lee (2018, p. 4) makes a distinction between “tasks that require more ‘human’ skills (e.g., subjective judgment and emotional capability) and those that require more ‘mechanical’ skills (e.g., processing quantitative data for objective measures).” People had equal trust in either algorithms or humans completing mechanical tasks, such as making a work schedule. When it comes to tasks which are perceived as requiring a human skill (such as hiring or work evaluations), algorithms were perceived as less fair than humans, since algorithms do not have intuition and cannot make subjective judgments (Lee, 2018). Other studies confirm this distinction. Bigman and Gray (2018) showed that for moral decisions, people are algorithm averse: when it comes to decisions in moral areas, such as life and death decisions in the military, medicine, or law, humans are preferred as decision-makers



compared to algorithms. Logg et al. (2019), on the other hand, found algorithm appreciation when it comes to numerical estimates. Here, algorithmic advice is preferred over human advice. These different preferences of algorithms versus humans reflect how people perceive the relative strengths and weaknesses of algorithms and how this differs from the way they think the human mind operates (Logg et al., 2019).

Dietvorst et al. (2015) showed that apart from assumptions about the general working of algorithms, experience with algorithms matters for algorithm trust and approval. When people have seen an algorithm make a mistake, they quickly lose faith and subsequently will avoid algorithms. When a human makes a mistake, this is more easily forgiven and does not affect trust in other decisions to the same degree, even when the mistake is more severe than the one made by the algorithm.

Representations of and discussions about algorithms in the mainstream media might likewise influence the approval of news algorithms. From research on robotics, it is known that representations of robots in popular culture can either make people more accepting toward such forms of automation, or more skeptical. Following the “Hollywood Robot Syndrome” media depictions of robots affect people’s general perceptions of robots (Sundar et al., 2016). When they recall friendly movie robots, people are generally more supportive of robots’ roles in society more generally. This also makes them more positive toward automatically generated news stories (Franklin Waddell, 2018). Media depictions may also affect robot support negatively. Following the Frankenstein complex, depictions of evil robots might trigger negative connotations of robots, focusing on their shortcomings and the danger they pose to humans (Kaplan, 2004; Kim and Kim, 2018). Similar to the Frankenstein complex, it could be expected that the simple use of the term “algorithm” might make people skeptical, given the negative attention to algorithm-driven filter bubbles and echo chambers in public and media discourse.

Surveys have given insight into how the general public approves of news algorithms. Zarouali et al. (2021) documented that misconceptions about media algorithms are common among the general population. Araujo et al. (2020) showed that news recommendations attributed to AI were seen as equally fair and useful as news recommendations attributed to human editors. Fletcher and Nielsen (2019) showed that people across different media systems even prefer automatically selected news stories over stories selected by journalists. At the same time, there was broad concern that such automatically selected news might give a narrow perspective (Nielsen, 2016). This reflected a generalized skepticism where respondents are critical toward both journalists and algorithms. Thurman, Moeller et al. (2019) found that people prefer automatically selected news when it is based on their own prior news consumption over

news which is automatically selected based on what peers have seen. This could be related to on the one hand overconfidence in one's own ability to select relevant news, and on the other hand to performance problems of automatically generated peer recommendations. Younger respondents are generally more aware of the role of algorithms in their social media feeds and also take a more active role in trying to influence which content algorithms show in their social media feeds (e.g., Bucher, 2018). This might explain why younger generations are more positive toward news algorithms and algorithm-driven news personalization than older generations (Bodó et al., 2019; Fletcher and Nielsen, 2019).

### 2.3 Research questions

Building on the literature described above, two research questions are posed to contribute to our understanding of news algorithm perceptions and approval among the general population.

Research question 2.1:

*How do people perceive the strengths and weaknesses of news algorithms compared to human journalists?*

As previous research has shown, perceptions of news algorithms are heavily shaped by comparisons to human journalists. This chapter makes these comparisons explicit, by asking whether people think that algorithms or human journalists are better at selecting a specific type of content. In line with the machine heuristic, it could be expected that algorithms are seen as better than humans at selecting information, which is neutral, objective, balanced, and trustworthy. On the other hand, selecting such information could be seen as a “human,” rather than “mechanical” task and should therefore be left to journalists. Following concerns about filter bubbles and echo chambers, the chapter explores whether algorithms or humans are perceived as better at selecting information, that is surprising, relevant, and offers different perspectives, will be explored.

Research question 2.2:

*Do people trust and approve of news algorithms?*

Trust and approval of news algorithms are closely connected to how their strengths and weaknesses are perceived. Generalized trust in algorithms selecting and writing news will firstly be compared to generalized trust in human journalists and news organizations. This analysis of generalized trust will be supplemented by a more narrowly focused analysis of trust in news algorithms, studying how people react to scenarios where either human journalists or an algorithm makes a journalistic mistake. Following

Dietvorst et al. (2015), it could be expected that people are less forgiving toward algorithms making mistakes than toward human journalists making mistakes. Next, the approval of news algorithms will be compared to approval of news algorithms making decisions in other areas of public life, such as court cases, hiring decisions, or hospital patient prioritization. This will give insight into whether news selection is seen as a more human or more mechanical task than other important decisions in public life (Lee, 2018). In the analysis of trust in algorithms and approval, special attention will be given to age differences. Following Bodó et al. (2019) and Fletcher and Nielsen (2019), it can be expected that younger generations are more positive toward news algorithms than older generations.

## 2.4 Method

To answer these research questions, this chapter reports the results of three representative surveys among the Danish population. These surveys included both questions about perceptions of news algorithms as well as two survey-embedded experiments. As described above, previous surveys about perceptions of news algorithms have given insight into approval and trust in news algorithms around the world (e.g., Fletcher and Nielsen, 2019; Thurman, Moeller et al., 2019). This chapter expands on these surveys by directly comparing perceptions of news algorithms with perceptions of human journalists, and by placing the approval of news algorithms in context by comparing this to approval of algorithms in other areas of public life. The first survey-embedded experiment tries to establish whether the simple use of the value-loaded term *algorithms* affects approval of news algorithms. The second survey-embedded experiment follows the approach by Bigman and Gray (2018) and presents the respondents with different scenarios in order to compare approval of news algorithms with approval of human journalists and how mistakes affect this approval.

Concretely, the analysis is based on the following three surveys:

- The question whether journalists or algorithms are better at selecting different types of content (Figure 2.1) was examined in a survey among 1225 Danish social media users. These data were collected between 25 March and 7 April 2020.
- The questions about trust in news algorithms compared to trust in other actors as well as the questions comparing approval of news algorithms with approval of algorithms in other areas of public life were studied in a survey among 1210 social media users in the period between 15 and 30 April 2020.

Both these surveys were conducted by the polling company Epinion, and participants were recruited from Norstats Online panel. All respondents were between 18 and 65 years old and use Facebook at

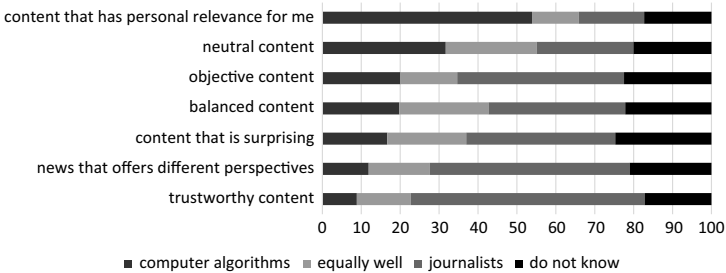


Figure 2.1 Who do you think is best at selecting the following type of content?

Notes: N = 1225 social media users between 18 and 65. Question wording: Every news website, mobile app or social network makes decisions about what news content to show to you. Increasingly, these decisions are made by computer algorithms, which use different criteria to select which news stories to show. Who do you think is better able to select content with the following qualities? 1. Human editors and journalists; 2. Computer algorithms; 3. Equally well, 4. I do not know (% of population choosing each category).

least once a month. Both studies were weighted to be representative of the Danish population on the parameters of gender, age, and region.

- The survey-embedded experiment comparing approval of news algorithms with approval of human journalists was part of a survey completed by 1200 Danes. These data were collected between 17 and 25 November 2021. This survey was conducted by polling company DMA research, and participants were recruited from Bilendi's online panel. Respondents were between 18 and 88 years old and 81.2% of the respondents were regular Facebook users. The data were weighted to be representative for the Danish population on the parameters of gender, age, region, and education.

The final part of the analysis reports differences in approval of news algorithms between four generations. The operationalization of the four generations is based on a classification by Pew Research Centre (Dimock, 2019): Generation Z: born in 1997 or later; Millennials: 1981–1996; Generation X: 1965–1980; Boomers: 1946–1964.

The exact question wordings and experimental conditions in the three surveys are described in the following section together with the results.

## 2.5 Results

### *Man vs machine*

Respondents were asked whether they thought that algorithms or human journalists and editors are best at selecting different types of

content (see Figure 2.1). When it comes to content with personal relevance for the user, the majority of respondents think that algorithms outperform human journalists. Only 16% of respondents have more faith in human journalists in this respect. Personally, relevant content seems to come with a downside. Less than 20% of respondents think that algorithms are better than humans at selecting content that is surprising or news that offers different perspectives. Here human journalists are clearly preferred. This seems to indicate that fears of filter bubbles influence people's assessment of news algorithms.

In total, 30% of social media users think that algorithms are better at selecting neutral content than humans, which is in line with the machine heuristic described above. This is more than the share of respondents who think that human journalists are best at this. Even though news algorithms are seen as more neutral than human journalists, they are at the same time seen as less able to select objective or balanced content. Here the share of respondents who prefer human journalists is almost twice as large. Thus, the perceived technical neutrality of algorithms is not seen as the same as absence of bias.

When it comes to selecting trustworthy information, respondents overwhelmingly prefer human journalists (60%) over computer algorithms (9%). Low trust in news algorithms was confirmed in further survey questions. When asked directly how much social media users trust news selected by computer algorithms, 72% of respondents answer that they have no or low trust, 25% has neither high nor low trust, and only 3% has trust or high trust ( $n = 595$ ). To test whether this low trust might be due to negative connotations with the term algorithm rather than an expression of true concerns with their role in news selection, we randomly divided respondents into two groups. One group was asked how much trust they have in news selected by algorithms, and one group was asked how much trust they have in news selected by automated computer systems. Also, among the group who was asked about automated computer systems, trust was very low: only 2% of respondents said they (highly) trust such news ( $n=615$ ).

This low trust in news selected by algorithms stands out even more clearly when it is compared to trust in the news media and to trust in automated computer systems. Around half of the respondents have (a great deal of) trust in the news media (52%) or journalists (44%). Less than one in six respondents say they trust artificial intelligence (12%), robots (14%), or computer algorithms (15%), but such trust levels are still significantly higher than trust in news selected by algorithms. Thus, when it comes to trust in news selected by computer algorithms, the whole is smaller than the sum of the parts.

*Algorithm aversion*

The lack of trust in news algorithms was confirmed in an experiment studying the approval of news written by news algorithms. Following the literature on algorithm aversion discussed above, an experiment was set up to test two hypotheses. Hypothesis 1 predicts that people approve less of automated computer systems writing articles than of journalists writing such articles. Hypothesis 2 predicts that approval decreases more when an automated computer system makes a journalistic error than when a journalist makes a journalistic error. To test these hypotheses, 1200 respondents were randomly divided into eight groups of 150 respondents. Each of the respondents was asked to think of the following scenario:

*A Danish news website publishes articles about each question to a minister in parliament and the subsequent answer.*

*These articles are written by a journalist*

OR

*These articles are written by an automated computer system.*

The experiment furthermore varied whether the scenario indicates that a human journalist checks the automated articles before publication and whether a mistake was made in one of the published articles by incorrectly referring to former PM Poul Schlüter instead of current PM Mette Frederiksen as Prime Minister in a question about Corona (see Table 2.1). The different scenarios are based on an actual mistake made in an automated story on the website of the Danish news medium Altinget (see Andreassen, 2020). After reading one of the scenarios, the respondents indicated on a scale from 1 (strongly disagree) to 5 (strongly agree) whether they agree that (a) it is appropriate for a journalist/automated computer system to write these articles, (b) a journalist/automated computer system should be forbidden from writing these articles, and (c) I trust the journalist/automated computer system writing these articles. The first two statements were adapted from Bigman and Gray (2018). After reversing the answers to statement b, the three questions were combined into one scale measuring approval (Cronbach's alpha .69,  $M = 2.44$ ,  $SD = .97$ ).

The results of the experiment revealed general disapproval of the idea of automated computer systems writing news stories. Approval across the six conditions where the automated computer system writes the news stories is well below the midpoint of the scale ( $M = 2.21$ ,  $SD = .90$ ), and around one point on the five-point scale lower than approval of journalists writing news stories about ministerial questions

Table 2.1 Algorithm aversion experiment

| <i>Condition</i> | <i>Articles are written by</i> | <i>Mentions whether articles are checked?</i>                             | <i>Mentions mistake?</i> | <i>Additional information</i>  | <i>Approval<sup>2</sup><br/>M (S.D.)</i> |
|------------------|--------------------------------|---|--------------------------|--|--|
| 1                | A journalist                   | No  | No                       |  | 3.33 (.79)                               |
| 2                | A journalist                   | No  | Yes                      |  | 2.96 (.82)                               |
| 3                | An automated computer system   | No  | No                       |  | 2.20 (.88)                               |
| 4                | An automated computer system   | Yes: The articles are not checked by human journalist before publication. | No                       |  | 2.18 (.95)                               |
| 5                | An automated computer system   | Yes: The articles are checked by human journalist before publication.     | No                       |  | 2.33 (.95)                               |
| 6                | An automated computer system   | No  | Yes                      |  | 2.18 (.90)                               |
| 7                | An automated computer system   | Yes: The articles are checked by human journalist before publication.     | Yes                      |  | 2.19 (.83)                               |
| 8                | An automated computer system   | Yes: The articles are not checked by human journalist before publication. | Yes                      | “The editor of the news website said they take the mistake just as serious as if it was made by a human.” <sup>3</sup> | 2.15 (.90)                               |

*Notes:* Danish news website publishes articles about each question to a minister in parliament and subsequent answer. These articles are written by .... <sup>1</sup>Recently one of these published articles mistakenly referred to Poul Schlüter instead of Mette Frederiksen as Prime Minister in a question about Corona. <sup>2</sup>Scale from 1 (fully disapprove) to 5 (fully approve).

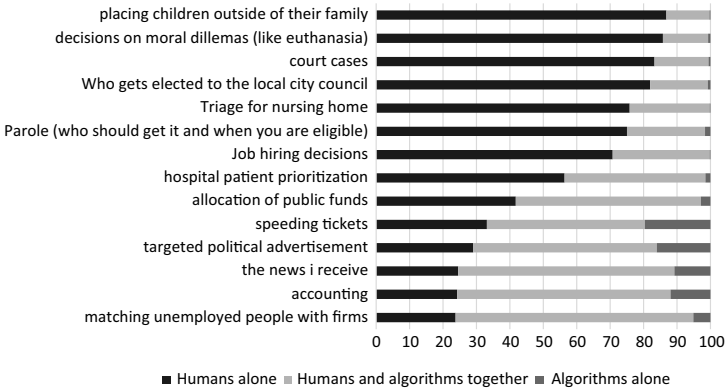
( $M = 3.15$ ,  $SD = .82$ ). As expected, there is a significant difference in the approval of news written by journalists (scenario 1;  $M = 3.33$ ,  $SD = .79$ ) and news written by automated computer systems (scenario 3;  $M = 2.20$ ,  $SD = .88$ ,  $t(295) = 11.72$ ,  $p < .001$ ). Thus, there is support for Hypothesis 1. What is more, people are more approving of news written by a journalist who makes a mistake (scenario 2;  $M = 2.96$ ,  $SD = .82$ ) than of an automated computer system when no mistake is mentioned (scenario 3,  $t(293) = 7.69$ ,  $p < .001$ ).

There are no significant differences between the six scenarios in which the articles are written by automated computer systems. Thus, contrary to Hypothesis 2, whether the algorithm made a mistake or not did not affect approval. Nor did approval improve when the scenario mentions that human journalists check the articles written by the automated system before publication (scenarios 5 and 7), or when the scenario mentions that the mistake is taken just as seriously as if it was made by a human.

### *Algorithms everywhere*

Automated news selection is only one example of the growing role of algorithms in public life. Figure 2.2 shows that social media users are more positive about the use of algorithms in news selection and dissemination than in other areas of public life. News selection is one of the areas in public life where least people think only humans should make decisions (24%). In line with low trust in news algorithms described above, only a small minority (11%) thinks that algorithms alone should select the news one receives. On the other hand, the large majority (65%) thinks that algorithms should play a role in news selection, as long as there are also human journalists involved. Also, for tasks like matching unemployed people with firms, accounting, speeding tickets, and the allocation of public funds, the majority of the respondents think that algorithms should play a role, but primarily with human oversight. For other areas of public life, the public is much more skeptical about the added values of algorithms, compared to humans making decisions. This is the case for decisions that have a life-changing impact on people's life, such as hospital patient prioritization, parole, triage for nursing homes, or removing children from their families. Also, when it comes to democratic decision-making, such as elections or court cases, only a handful of respondents see a role for algorithms and hardly anyone thinks that algorithms alone should make decisions.





*Figure 2.2* Should humans or algorithms make decisions in different areas of public life?

*Notes:* N = 1210 social media users between 18 and 65. Question wording: Decisions in society are increasingly made by computer algorithms or with the support of computer algorithms. In your view, what is the best way to make decisions in the following areas? 1. Humans alone should make decisions, 2. Humans should make the decisions together with algorithms, 3. Algorithms alone should make decisions (% of respondents choosing each category).

### *Gen Z*

Figure 2.3 shows how different generations think about whether algorithms should play a role in news selection. For all generations, the majority of respondents think that it is best that algorithms and human together select news, but apart from that there are strong inter-generational differences. Among the Boomer generation (born between 1946 and 1964), 40% believe that humans alone should decide which news they receive, and only a small minority thinks that this should be decided by algorithms alone. The younger the respondents, the more positive they are toward news algorithms. Among Generation Z (born after 1997), only 6% believe that humans alone should decide which news they receive. No less than 27% of this generation believes that algorithms should select news without the interference of human journalists. A secondary analysis of the other results presented confirms that Generation Z is the most positive toward news algorithms. They are more positive than older generations about how trustworthy, balanced, and objective news selected by algorithms is (results not shown). Still, when asked directly, only 8% of Generation Z say that they trust news selected by news algorithms.

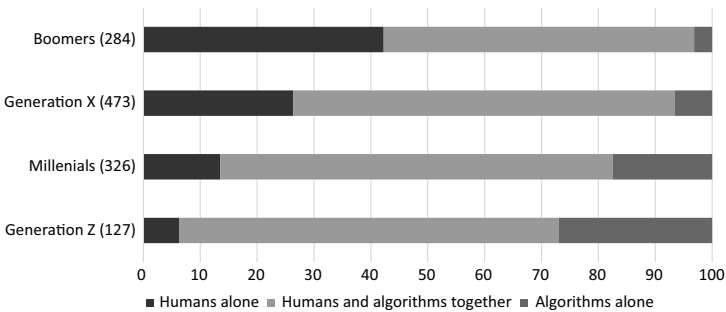


Figure 2.3 Should humans or algorithms decide which news you receive? (Generational differences).

*Notes:* Number of respondents per generation in brackets. Question wording: Decisions in society are increasingly made by computer algorithms or with the support of computer algorithms. In your view, what is the best way to make decisions in the following areas? The news I receive 1. Humans alone should make decisions, 2. Humans should make the decisions together with algorithms, 3. Algorithms alone should make decisions (% of each generation choosing each category).

## 2.6 Conclusion

This chapter presented the results of surveys studying news algorithm perceptions and trust among the general population. The surveys first and foremost showed a lack of trust in news algorithms. Only a very small minority said that they trust algorithms selecting news, which stood in marked contrast with trust in journalists selecting news. When it comes to trust in algorithms writing news, the results were similar. People even prefer journalists who make mistakes to write news rather than algorithms, which do not make mistakes. These results reveal low generalized trust in algorithms, which is not dependent on the actual performance of news algorithms. The low levels of trust in automatically selected and created news were not triggered by the term “algorithm” alone. When the more neutral term “automated computer systems” was used people also express low levels of trust. This lack of trust is accompanied by skepticism about the ability of news algorithms to select content which is objective and presents diverse perspectives.

On the more positive side, the majority of the respondents think that algorithms are better than journalists at selecting news with personal relevance. People are more approving of news algorithms than of algorithms playing a role in other areas of public life, such as ethical and moral decision-making. Across generations, there is general acceptance of algorithms playing a role in news selection as long as there are also humans involved in the process. This suggests that the mental model of

what it means to select and write information is still very much based on the traditional image of a human journalist. This seems to change from generation to generation. The Boomer generation which grew up with newspapers and traditional news outlets like radio and television was less supportive of algorithmic news selection. There was more support for algorithmic news among younger generations, in particularly Generation Z who from a young age have used social media for news consumption. This could indicate that over time the general population will become more used to and thus more accepting of news algorithms.

The next chapter will shift the focus from broad perceptions of algorithmic gatekeepers to a more narrow analysis of the mediating power of algorithms on the social media platform YouTube.

# 3 The Mediating Power of Algorithms

## 3.1 Introduction

From the 2016 US Presidential elections to debates about climate change or COVID-19 vaccines, concerns have been raised worldwide about the spread of misinformation on social media. Central in discussions about misinformation on social media is the mediating power of algorithms. As this chapter addresses, algorithms are blamed for stimulating the spread of misinformation, but at the same time seen as potential remedies against the spread of such information. Thus, algorithms play an important role in maintaining the boundaries between legitimate and illegitimate information.

This chapter presents an analysis of the role of social media algorithms in the spread and correction of misinformation. This is analyzed through a case study of YouTube's algorithms and (mis)information concerning the treatment of autism. The treatment of autism spectrum disorder (ASD) is chosen as a case to analyze the role of algorithms in spreading and countering misinformation for two reasons. (1) Social media have become an important platform where people share and search for information about autism (e.g., Saha and Agarwal, 2015). (2) Apart from the importance of social media as a channel, the treatment of autism is an interesting case, since it allows distinctions to be made between truthful information, misinformation, and correcting misinformation (see below). Thus, the research question which this chapter addresses is: *Do YouTube's algorithms recommend truthful information, misinformation, or corrections to misinformation concerning the treatment of autism?*

Before addressing this question, this chapter conceptualizes misinformation and discusses the double role of social media algorithms in spreading and limiting misinformation. This is followed by a presentation of the case and a review of previous research on the mediating power of YouTube's algorithms. The empirical analysis of the presence of misinformation and corrections to misinformation among YouTube's

search results and recommendations reveals that the algorithms suggest some misinformation, but at the same time seem to limit the spread of directly harmful misinformation. The final chapter of this book will return to these results to discuss how social media platforms should best handle the mediating power of their algorithms.

### **3.2 Misinformation and algorithms**

From the perspective of the sender, misinformation can be defined as “false or inaccurate information, especially that which is deliberately intended to deceive” (Lexico, 2021). From the perspective of the receiver, misinformation can be defined as “information that is initially believed to be valid but is subsequently retracted or corrected” (Ecker et al., 2014, p. 293). Misinformation can have severe negative consequences, both at the societal level and at the individual level. On the societal level, the general population, electorate, or political elite may make decisions which go against the best interests of society. On the individual level, misinformation can lead people to make decisions which are directly harmful for themselves and others. Misinformation is worse than ignorance, since research has consistently shown that once people have encountered misinformation it is difficult to correct and even when people know that information is false, it remains influential (e.g., Ecker et al., 2014 for an overview).

Misinformation is not a new phenomenon, but the social media environment has made it a highly pressing topic for at least three reasons. First, it is easy to publish misinformation on social media which can potentially reach a large audience when it spreads virally. Second, it is more difficult for social users to distinguish misinformation from reliable information than it would be on more traditional channels. In newspapers and on television, it is easier for the audience to identify the sender of a message than on social media. Third, once misinformation has been published and shared on social media, it is difficult to control and correct this information. Vosoughi et al. (2018) have shown that on Twitter, false information reaches more people and is spread faster than truthful information. They point to the newness of false information and its emotional appeal as possible explanations for why this type of content spreads so rapidly.

The presence and spread of misinformation on social media have led to broad concerns among the general population and public authorities. A poll in the United States in 2020 showed that two-thirds of Americans thought that social media have a negative impact on the country and that the main reason for this was concerns about the spread of misinformation (Auxier, 2020). Public authorities like the FBI (2020) and the European Commission (2022b) have warned against the spread of misinformation on social media.

Algorithms play a double role in this spread of misinformation. On the one hand, algorithms are blamed for fostering the spread of misinformation, for example by leading people from more moderate content toward more extreme content (e.g., Ribeiro et al., 2020) or by prioritizing misinformation because the content is emotional, novel, and quickly gains a good deal of traction and interaction (e.g., Avaaz, 2020).

On the other hand, social media companies use algorithms and machine learning to counter the spread of misinformation. Social media companies themselves become increasingly aware of misinformation on their platforms and take action against it. For example, Facebook states on its website that it works actively toward “stopping false news from spreading, removing content that violates our policies, and giving people more information so they can decide what to read, trust and share” (Facebook, 2020). Algorithms often play a central role in the detection of misinforming content and subsequent actions taken to correct it (e.g., Gillespie et al., 2020). YouTube for example makes an active effort to reduce recommendations for videos containing misinformation. It announced in 2019 that it takes action aimed at “reducing recommendations of borderline content and content that could misinform users in harmful ways – such as videos promoting a phony miracle cure for a serious illness, claiming the earth is flat, or making blatantly false claims about historic events like 9/11” (YouTube, 2019). This misinformation is identified using a “combination of machine learning and real people,” involving “human evaluators and experts from all over the United States to help train the machine learning systems that generate recommendations” (YouTube, 2019). With such efforts, algorithms take on a powerful role in content moderation, which can be defined as “the detection of, assessment of, and interventions taken on content or behavior deemed unacceptable by platforms or other information intermediaries, including the rules they impose, the human labor and technologies required, and the institutional mechanisms of adjudication, enforcement, and appeal that support it” (Gillespie et al., 2020, p. 2).

The intention to correct misinformation by using algorithms is inherently praiseworthy. Given the enormous amount of content uploaded onto social media platforms, automation can be a way to address problems of scale. At the same time, scholars warn of the potential risks that this new approach to content moderation poses (e.g., Cobbe, 2021; Gillespie et al., 2020). Some of the risks mentioned by Aram Sinnreich (in Gillespie et al., 2020) are the following. First the danger of false positives and negatives; machine learning might wrongfully identify some correct information as misinformation, while overlooking other cases of misinformation. Without human oversights, this will go uncorrected. Second, subtle and nuanced distinctions and assessments require human judgment. If these decisions are left to algorithms, the more rigid algorithmic logic will determine between

right and wrong, leaving little space for gray areas. Third, it is challenging to take cultural values and national regulation into account when developing algorithms that moderate speech on social media platforms where content is shared across national borders. Overtly rigid algorithms might ban content worldwide even though such content is seen as unproblematic in certain regions. Less strict algorithms might allow the circulation of messages which are not protected by freedom of speech in other parts of the world. This chapter will further discuss the role of algorithms in spreading and limiting misinformation based on an analysis of YouTube recommendations of videos about autism.

### **3.3 YouTube and autism spectrum disorder**

ASD, or briefly “autism,” is a medical diagnosis based on symptoms in two domains: “(i) deficits in social communication and social interaction; (ii) restricted repetitive behaviors, interests, and activities” (Fuentes et al., 2021, p. 962). Autism begins in the pre-school years, although it might not be diagnosed until much later in life (Centers for disease control and prevention, 2022a). There are several interventions and treatments which can reduce ASD symptoms and make everyday life easier for children with ASD. There is no scientific evidence that ASD can be cured (Centers for Disease Control and Prevention, 2022b). A number of services and products are available that allegedly should cure or heal autism, such as spiritual or religious practices or dietary changes. These methods are not scientifically proven to cure autism, although they do not directly cause harm (e.g., Lofthouse et al., 2012, p. 10; Fuentes et al., 2021). More dangerous is the recommendation to cure autism by drinking a substance called Miracle Mineral Solution that, when used according to instructions, turns into a type of bleach, which is used for industrial cleaning and should not be consumed by humans (U.S. Food & Drug Administration, 2019). This product is offered online and without scientific evidence presented as a remedy for a wide variety of conditions, ranging from AIDS to COVID-19 and autism as well.

A large range of information on the treatment of autism is available on social media platforms. Such information could take the form of mommy or daddy blogs, where parents of children with autism share their experiences, dedicated pages or communities on Facebook, or videos on YouTube, which have been seen by millions of viewers. Sharing information about autism and encountering people in similar situations can be important for parents or siblings of people with autism, since this “helps them cope with their situation in a better way and helps them feel a part of a community” (Saha and Agarwal, 2015, p. 1053). On social media platforms, the type of information spread and shared around autism is much broader than only information on which there is

scientific consensus. This can be exemplified by the role of former model, actress, and comedian Jenny McCarthy, who has a prominent voice in the debates about autism (Chivers Yochim and Silva, 2013). Jenny McCarthy talks publicly about raising her autistic son Evan and wrote a book in 2008 called *Mother Warriors; A nation of parents healing autism against all odds*. In her public conversations about autism, Jenny McCarthy positions herself against the medical establishment, as she “distrusts traditional scientific and medical expertise and exalts what have long been seen as feminine modes of knowing such as instinct and expertise” (Chivers Yochim and Silva, 2013, p. 406). McCarthy has a large following on Instagram and often appears in popular television programs to talk about autism. Clips of such visits are widely shared online and broadly watched on YouTube.

Started in 2005, YouTube is a platform where people share and watch videos. YouTube is both home to user-generated video clips and professionally produced content, either created specifically for YouTube’s platform or posted on the platform after being originally shown elsewhere (e.g., Burgess and Green, 2018). YouTube includes features of social media platforms, such as giving users the opportunity to comment on videos, allowing users to follow specific channels, and the possibility to express one’s opinion by upvoting or downvoting videos. YouTube is first and foremost an entertainment platform, featuring a wide variety of videos such as music videos, figure skating clips, or make-up tutorials. At the same time, YouTube offers a platform for informative content, such as videos from political parties, news clips, or clips from pundits who comment on current events. Many such videos present perspectives and opinions which are excluded from mainstream debate, such as right-wing content, conspiracy theories, or populist voices (e.g., Faddoul et al., 2020; Lewis, 2020).

As noted and described in the first chapter, the rise of communities around specific types of content on social media platforms should not be attributed to algorithms alone. Algorithms should be seen as one element in a broader set of factors at the supply side (incentives and opportunities offered by the platform) as well as demand side (audience characteristics) (Munger and Phillips, 2022). While Munger and Phillips’ supply-and-demand model was developed to explain the popularity of the radical right on YouTube, some of the same dynamics might also contribute to explaining the popularity of content about autism and treatments. People diagnosed with autism, or parents of children with autism, might look for connection and a feeling of being understood, which they might not always encounter in their personal environment. YouTubers who talk about autism such as Jenny McCarthy or TED talk presenters might cater to this need for connection and recognition. By referring people who watched videos about autism on to more



similar content, YouTube's algorithm might strengthen the feeling that YouTube offers a place where one can feel understood.

YouTube videos on autism can include trustworthy information, misinformation as well as messages correcting misinformation. If it is stated that autism can be treated, without claiming that it can be cured, that would be an example of truthful information. The message that autism can completely be cured goes against scientific consensus and should therefore be considered misinformation. Here, a distinction can be made between on the one hand harmless misinformation, which is not supported by scientific evidence, but which does not directly damage people's health, and on the other hand, harmful misinformation, which could damage people's health. Claiming that autism can be healed spiritually would be an example of the former and claiming that it can be cured by drinking Miracle Mineral Solution would be an example of the latter. Information that directly counters claims that autism can be cured can be considered correcting misinformation. An example of this could be warnings against drinking Miracle Mineral Solution.

### **3.4 The mediating power of YouTube's algorithms**

The specific working of YouTube's algorithms is considered proprietary information. In addition, such algorithms are self-learning and are continuously tweaked. This makes it difficult to describe the way they function. Still some general principles can be found in Covington et al. (2016) and Davidson et al. (2010). These papers describe that co-viewing patterns play an important role in the recommendations made by YouTube's algorithms: Which videos other people have watched after watching a particular video will affect which videos are recommended. In this process, the algorithms balance between homogeneity and diversity (see Van Dalen, 2021). On the one hand, recommendations after watching a particular video should be similar enough to appeal to the viewers of this video. On the other hand, it should be avoided that the exact same content is recommended or that viewers only see videos from a small number of channels. Also new content needs to have a chance to be recommended to avoid viewers from only seeing videos which were uploaded a long time ago.

Previous research on YouTube recommendations and misinformation has studied whether misinforming videos are followed by more misinformation (e.g., Hussein et al., 2020) and how recommendations link misinforming videos and videos correcting misinformation (e.g., Schmitt et al., 2018).

An audit of YouTube showed that misinforming videos concerning 9/11, conspiracies, chemtrails, flat earth, and the moon landing were often followed by more misinforming videos (Hussein et al., 2020). Such a filter

bubble effect was not found for misinforming videos about vaccines, however. This filter bubble effect for misinformation is in line with O'Callaghan et al.'s (2015) study of recommendations after watching right-wing extremist channels. Their research showed that many recommended channels after watching videos from right-wing extremist channels also feature right-wing content, such as anti-Islam or neo-nazi videos. During the Danish parliamentary elections in 2019, videos from a populist politician were mainly followed by more videos about this populist politician (Van Dalen, 2021). This research also showed that this effect was not only present after videos from a populist politician but for videos from mainstream politicians as well.

A study of vaccine misinformation on Amazon showed that Amazon's search algorithms were more likely to recommend misinforming rather than corrective information (Juneja and Mitra, 2021). Whether YouTube's search and recommendation algorithms also steer viewers toward misinformation is an open question. Beyond misinformation, research has looked into the question whether YouTube's algorithms steer people toward extremist content. An analysis into pathways of radicalization on YouTube showed that recommended channels after mainstream content may feature alt-right content (Ribeiro et al., 2020). However, these results were based on a small number of recommendations. In Denmark, little evidence was found that algorithmic recommendations on YouTube lead viewers from mainstream content toward extremist politicians (Van Dalen, 2021, see also Chen et al., 2021). Kaiser and Rauchfleisch's (2020) research showed that YouTube's algorithms were more likely to recommend right-wing content after mainstream content than mainstream content after right-wing videos. Whether YouTube also recommends misinformation after watching factually correct videos about Autism remains an open question.

It is important to study how YouTube search and recommendation results connect videos with correct information and with misinformation, since videos correcting misinformation could counter the negative effects of misinformation (e.g., Schmitt et al., 2018). For Facebook, Bode and Vraga (2015) showed that the effect of misinformation is reduced when Facebook suggests a related story that corrects such misinformation. However, research has shown that YouTube's algorithms do not only refer viewers from misinformation toward correcting information. Schmitt et al. (2018) showed that videos that are intended to prevent extremism are often connected by recommendations to extremist content.

YouTube continuously updates its algorithms and also takes responsibility by paying attention to the potential spread of misinformation on its platform. In January 2019, YouTube announced that it would reduce recommendations for videos that could misinform (Carrie Wong and Levin, 2019). Following this change, the share of recommended conspiracy videos after watching a conspiracy video has almost halved (Faddoul et al., 2020).

### 3.5 Research questions

Based on previous research on the role of algorithms in the spread of extremist content, the following questions are asked about how algorithms affect the visibility of misinformation about autism on YouTube. The research questions deal both with videos that are recommended after watching videos about autism, as well as videos that are found when searching for information about autism.

#### Research question 3.1

*Does YouTube recommend videos misinforming that autism can be cured, after searching for or watching such videos?*

Previous research reviewed above has shown that YouTube's algorithms recommend videos with a similar perspective as the ones previously watched. This would suggest that after watching videos misinforming that autism can be cured, more of this type of videos would be suggested. On the other hand, given YouTube's commitment to countering the spread of misinformation, such videos might no longer be recommended, and the algorithms could instead suggest videos that correct misinformation. This could be expected to be the case for videos about autism, since YouTube has explicitly mentioned the Miracle Cure, which is suggested to cure autism, as an example of the misinformation that it tries to prevent spreading on its platform (Faddoul et al., 2020. p. 1).

#### Research question 3.2

*Does YouTube recommend videos misinforming that autism can be cured, after searching for or watching other videos about autism?*

Following the studies by Kaiser and Rauchfleisch's (2020) as well as Ribeiro et al. (2020), it could be expected that videos which misinform viewers that autism can be cured might also be found after watching videos about the treatment of autism or about autism in general. Research by Schmitt et al. (2018) suggests that due to thematic overlap, watching videos correcting misinformation about autism might even steer viewers toward videos promoting misinformation.

### 3.6 Method

The empirical analysis consisted of scraping YouTube's search results and video recommendations and a manual content analysis to analyze the perspective on autism treatment presented in these videos. The analysis was conducted in February and March 2020, before the Coronavirus sent many countries around the world into lockdown.

First, YouTube search results were collected using the program YouTube Data Tools (Rieder, 2015; Rieder et al., 2018). Using this program, the top 50 search results were collected, sorted by relevance, for the following terms: (1) treating autism; (2) curing autism; (3) healing autism; (4) Miracle Mineral Solution; (5) autism; and (6) Jenny McCarthy autism. This resulted in 300 recommended videos, which were manually coded by a trained sociology student who had previously undertaken an extensive literature review on the treatment of autism. The coder coded for each video the perspective on autism and allocated it into one of five categories:

The video states that autism can be cured. See for example the video “Autism Cure: How to Heal Autism in 30 Days – Autism Cure with Dr. Raji”: <https://www.youtube.com/watch?v=5K3IE7670Is&t=290s>

The video states that autism can be treated but does not state that it can be cured. See for example the video “The SMART Program: Treating Autism and Autism-Related Disorders”: <https://www.youtube.com/watch?v=nAyAEoKcWZE>.

The video corrects information about autism which in the video is presented as misinformation. See for example the video “Autism Mythbusters – Dr. Shafali Jeste – UclamdChat Webinars”: <https://www.youtube.com/watch?v=swtKS-Ik-4U>

The video is about autism, but does not talk about treatment or curation. The video could for example explain symptoms or describe how autism is diagnosed. See for example the video “Things not to say to an autistic person”: <https://www.youtube.com/watch?v=d69tTXOvRq4>

The video is not about autism. Videos in this category do not contain any talk or references to autism, treatment hereof or other aspects related to autism.

An intra-coder reliability test of 59 videos showed Krippendorff’s alpha score of .70.

Next, video recommendations were collected and analyzed. First, 50 start videos were selected out of the videos found in the search described above: 10 videos arguing that autism can be cured; 10 videos arguing that autism can be treated; 10 videos correcting misinformation; 10 videos about autism without mentioning whether it can be treated or cured; and 10 videos about autism featuring Jenny McCarthy. In the second step, computer programming language Python was used to identify 10 recommended videos for each of these 50 start videos. The recommendations resemble the videos which would feature next to a YouTube video when it is watched on a computer, or below a video when it is watched on a smartphone. The perspective on the treatment of

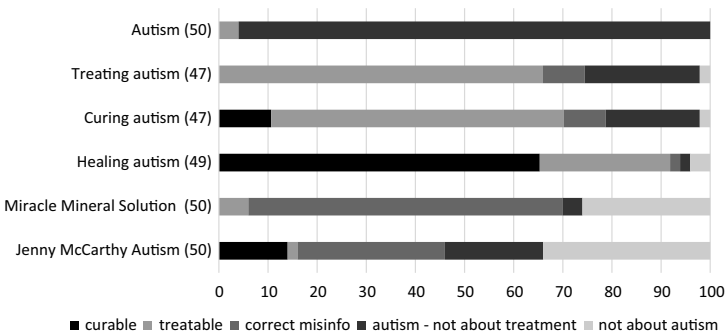
autism in the recommended videos was then classified using the same coding scheme as described above.

The focus of the analysis is on whether videos misinform viewers that autism can be cured, a claim for which no scientific evidence was found. Videos stating that autism can be cured are considered misinformation, while videos stating that autism can be treated are not considered misinformation. As is common practice in the analysis of YouTube search results and recommendations (e.g., Faddoul et al., 2020; O’Callaghan et al., 2015), these results do not take prior user history or location into account. In practice, individual search results, as well as recommendations, will be influenced by which videos the user has watched previously. Therefore, the analysis should be seen as giving a baseline understanding of the role of YouTube’s algorithms in the spread of (mis)information on autism.

### 3.7 Results

#### *Searching for autism videos on YouTube*

Figure 3.1 shows what kind of videos YouTube suggests when searching for six autism-related terms. The results show that a simple change in search terms makes a large difference in the perspective presented concerning whether autism can be cured. Simply searching for videos about autism primarily leads to videos about autism, without discussing treatment or cure. When using the search term “treating autism,” the



*Figure 3.1* YouTube search results after searching for videos about autism using six different search terms.

*Notes:* % of suggested videos in the following categories: 1. Autism presented as curable, 2. Autism presented as treatable, 3. Correcting misinformation about autism, 4. About autism without discussing treatment, 5. Not about autism. Number of analyzed results per search term in brackets.

majority of videos present autism as treatable, while none of the videos mention that autism can be cured. One in ten videos corrects misinformation. The results are largely similar when the term “curing autism” is used, but now also videos that suggest that autism can be cured are among the recommendations. However, this still makes up only 11% of the search results. The picture changes completely when searching for “healing autism.” When this term is used, two-thirds of the videos in the search results suggest that people who have correctly been diagnosed with autism can later be diagnosed as not having autism. One in four videos suggests that autism can be treated rather than cured, and, compared to the results when searching for “curing autism” or “treating autism” fewer videos either correct misinformation or are simply about autism. Misinforming videos are completely absent from the search results when users search for “Miracle Mineral Solution.” When this term is searched for, YouTube primarily returns videos that correct the misinformation that Miracle Mineral Solution is a cure to autism and point out that it is directly damaging. This suggests that YouTube’s efforts to adjust its algorithms to limit the spread of misinformation about Miracle Mineral Solution are indeed successful. Finally, the search terms “Jenny McCarthy autism” primarily return videos that are not about autism or videos correcting misinformation (both one out of every three videos). The share of videos suggesting that autism can be cured is comparable to the share of suggested videos when searching for “curing autism,” but far fewer videos deal with the treatment of autism.

Based on this analysis of the search results, the two research questions can be answered as follows. When searching for videos about cures to autism, YouTube is more likely to recommend videos presenting autism as treatable or videos correcting misinformation. The search for ‘healing autism’ is an exception, as this primarily leads to videos misinforming that autism can be cured. When searching for videos about autism in general or the treatment of autism, no misinforming videos were found.

### ***Recommendations***

Figure 3.2 shows the perspective on the treatment of autism presented in recommended videos after watching five different types of videos. After watching videos about autism, which are not about treatment, viewers do not get recommended videos that misinform that autism can be cured. After watching videos which present autism as treatable, one in five recommended videos also presents autism as treatable. Only a small number of recommendations (3%) present autism as curable. At the same time, some recommended videos correct such misinformation. Thus, there is very limited evidence that YouTube’s algorithms steer

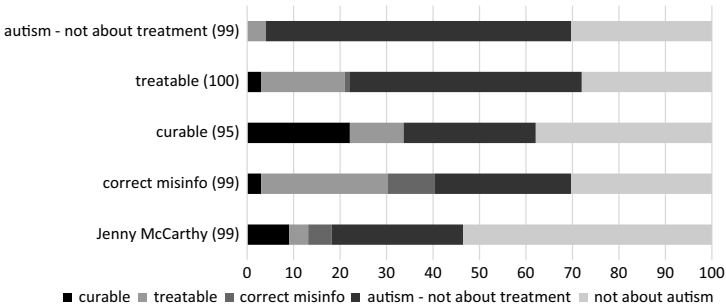


Figure 3.2 YouTube recommendations after five types of videos about autism.

Notes: % recommended videos in the following categories: 1. Autism presented as curable, 2. Autism presented as treatable, 3. Correcting misinformation about autism, 4. About autism without discussing treatment, 5. Not about autism. Number of analyzed recommendations per type of seed video in brackets.

viewers from correct information to misinformation. After watching a video presenting autism as curable, 22% of recommendations misinform that autism as curable. This is almost twice as many as the share of recommendations discussing how autism can be treated. None of the recommended videos set the record straight by correcting the claim that autism can be cured. Still, the majority of recommended videos do not discuss autism as either treatable or curable. The same goes for recommended videos after watching a video that corrects misinformation about autism. In total, 27% of recommended videos present autism as treatable and 9% likewise correct misinformation. Only three of the recommended videos after a correction of misinformation, misinform the audience. Less than half of the video recommendations are about autism, after watching a video in which Jenny McCarthy talks about autism. Here it is twice as likely that a recommended video misinforms that autism can be cured, compared to correctly informing that autism is treatable or correcting misinformation.

Based on this analysis of the recommendations, the two research questions can be answered as follows. After watching a video misinforming the viewer that autism can be cured, the viewer is more likely to be recommended a video confirming this misinformation than a video correcting this misinformation, or correctly informing that autism can be treated. Watching another video about autism seldomly leads YouTube viewers to misinformation. Still, it is more likely to be recommended misinformation after watching a video correcting misinformation than it is to receive a recommendation for a video correcting misinformation after watching a misinforming video.

### **3.8 Conclusion**

Contrary to popular concern, this chapter found no evidence that YouTube's algorithms promote misinformation about life-threatening products like Miracle Mineral Solution. When audiences search for Miracle Mineral Solution, the algorithms primarily recommend videos warning people against the use of this product. Secondary analysis showed that none of the recommended videos misinforming about the treatment of autism promoted the use of Miracle Mineral Solution. These findings are in line with work by Faddoul et al. (2020, p. 5). Another positive finding was that there was very limited evidence that YouTube's algorithms promote misinforming videos after watching or searching for videos about autism, the treatment of autism, or the correction of misinformation about autism. Still, YouTube's algorithms recommended plenty of videos that present autism as curable, despite consensus that there is no scientific support for this. It was striking to see what a difference small changes in search terms (treating vs curing vs healing) made for the suggested videos. After searching for videos healing autism, two-thirds of recommended videos suggested that either spiritual healing or dietary changes can cure autism. After watching such videos, one in five recommended videos confirms this perspective. One in five is a weaker filter bubble effect than was previously found for conspiratorial videos (32%, Faddoul et al., 2020, p. 5). Still, when such videos are recommended, this plays into people's tendency to select information that confirms their beliefs (Nickerson, 1998). Searching for videos about autism leads to plenty of suggested videos which correct misinformation. However, such videos were only seldomly suggested as videos to watch next after a video about autism. Confirming research by Schmitt et al. (2018), there was even some evidence of recommending misinforming videos after watching a video correcting such information. Given previous research, which highlighted that recommended posts on Facebook next to misinformation limit the effect of such misinformation (Bode and Vraga, 2015), there is potential here for YouTube's algorithms to further limit the spread of misinformation.

When interpreting these results, two limitations need to be considered. First, the analyzed search results and recommendations are not personalized. As is common practice in the analysis of YouTube search results and recommendations (e.g., Faddoul et al., 2020; O'Callaghan et al., 2015), these results do not take prior user history or location into account. In practice, individual search results, as well as recommendations, will be influenced by a user's viewing history. Therefore, the analysis should be seen as giving a baseline understanding of the role of YouTube's algorithms in the spread of (mis)information on autism. In real-life, YouTube will adapt the videos it proposes to what other



content users have previously interacted with. For some users, this will probably mean that less video recommendations are about autism since people will generally also have lots of other interests. For some who are highly invested in the topic, on the other hand, even more videos about the treatment of autism could be proposed, possibly with a highly specific perspective on how it can be cured. This should be topic for future research. Such research could for example monitor and compare recommendations to the same videos for various people with diverse relations to autism. Another option could be to create specific personas by watching a large number of diverse autism-related YouTube videos and analyzing the recommendations. The second limitation is related to the classification of misinformation in this study. The study narrowly defined misinformation about the treatability of autism as videos which suggest that autism can be cured. However, some of the videos about the treatment of autism might propose methods that are not based on scientific research (see Fuentes et al., 2021, p. 978 for an overview). Thus, if the study had explored misinformation about how autism could be treated, the study would likely find more misinformation on YouTube.

As described above, the results are largely in line with previous research on the role of YouTube's recommendation algorithms in the spread and limitation of misinformation in the form of conspiracy theories and hate speech. Still, there are some characteristics about the case of autism treatment that have to be considered when discussing generalizability to other topics where misinformation is debated, such as immigration, climate change, as well as the anti-vaccine movement. The treatment of autism is a highly obtrusive issue for people with autism and their relatives. This might make it more likely for viewing clusters to form around this issue than around less obtrusive issues like climate change. Compared to less obtrusive issues, the impact of autism on people's personal life might also make audiences more susceptible to non-scientific forms of knowledge and information, such as spiritual information or personal narratives like the ones provided by celebrities like Jenny McCarthy. This might explain why these videos are widely available on YouTube and why the algorithms recommend such content. Products like Miracle Mineral Solution can be directly damaging to people's health. The same goes for other medical advice, which is not supported by scientific evidence. Social media platforms have a responsibility to limit the spread of such information, which is directly harmful to individuals. In many cases, the distinction between factually correct and misinformation may however be more difficult to make, and the damage of misinformation takes longer to come to the surface. Here, one could for example think about misinformation about crime rates among immigrants. What the role of algorithms is in spreading or

limiting misinformation related to such topics should be subject to future research.

After studying the *mediating power* of algorithms in the current chapter, the next chapter will focus on the *structuring power* analyzing how communication professionals perceive and relate to the power of Instagram's algorithms.

## 4 Structuring Power

### 4.1 Introduction

This chapter addresses the power relations between Influencers and social media algorithms. This is studied by looking at the way professional communicators are affected by and react to Instagram's algorithms and the criteria they use to assess the legitimacy of these algorithms. Abidin (2015, p. 1) defines Influencers as "everyday, ordinary Internet users who accumulate a relatively large following on blogs and social media through the textual and visual narration of their personal lives and lifestyles." These Influencers make money by advertising, promoting, and endorsing products and services in their social media posts. Contrary to more traditional forms of branding or promotional communication, the way Influencers communicate with their followers is based on "perceived interconnectedness" (Abidin, 2015). Influencers give their followers a feeling of intimacy, by revealing (selected) aspects of their personal life and by asking their followers for input. This perceived interconnectedness makes Influencers very interesting for advertising brands. An endorsement of an influencer creates identification with a product (Schouten et al., 2020).

To create this interconnectedness and to nurture feelings of intimacy, Influencers rely heavily on social media platforms such as YouTube, TikTok, and most importantly Instagram. By using such platforms, they can by-pass traditional media and interact directly with their audiences. Instagram is a social media platform for sharing photos and videos, which started in 2010 and was bought by Facebook (later called Meta) in 2012. Instagram users follow friends, family members, or celebrities, comment and like their photos and videos or communicate privately by sending Direct Messages. The use of images helps Influencers to build a strong emotional connection with their followers and makes it a useful platform for self-branding (Khamis et al., 2017). Instagram is a highly popular platform which since June 2018 has more than 1 billion users (Constine, 2018). For Generations Z and Y in the United Kingdom and

the United States, Instagram was the app that most people had installed on their phone and spent most time on per day on average (Galan et al., 2019). In Denmark, 30% of the adult population used Instagram daily in 2020 (DR Medieforskning, 2021), while 93% of high school students in Denmark use the app (Mehlsen, 2020).

Instagram's algorithms determine which posts will be shown to individual users and how they are prioritized. The basic functions of algorithms on Instagram are described in a blog post on Instagram's homepage (Mosseri, 2021). Here, it is stressed that the images, which are presented by Instagram to the users, are not determined by one single algorithm, but rather by "a variety of algorithms, classifiers, and processes." These determine the order of posts shown in Feed and Stories, and the selection of new content recommended in Explore and Reels. When choosing and ranking such posts, Instagram takes characteristics of the post itself into account, such as how popular it is, one's relation with the person posting the images and one's own activities on Instagram. Based on this, Instagram predicts "how likely you are to spend a few seconds on a post, comment on it, like it, save it, and tap on the profile photo" (Mosseri, 2021).

This chapter analyzes how professional communicators perceive their relationship with Instagram's algorithms. The first research question, which this chapter explores, is *What is the perceived power of Influencers vis-à-vis Instagram's algorithms?*

As described below, Influencers have become actors with political relevance, which makes it important to study how Influencers behave in response to Instagram's algorithms and thereby give insight into how Instagram's algorithms might affect how they express themselves.

After describing how Influencers perceive their own power vis-à-vis the algorithms, the analysis turns the focus to the other side of the power dynamic, studying the question *Which criteria are used to assess the legitimacy of the power of Instagram's algorithms?* As described in Chapter 1, legitimacy plays an important role in making people adhere to power structures and trust these structures (e.g., Boulding, 1968; Boulding, 1971). Media institutions can derive their legitimacy from various sources, such as ethical standards or the character of the journalists working in these institutions (e.g., Van Dalen, 2019b). An analysis of explicit and implicit criteria used to assess the legitimacy of the power of Instagram's algorithms gives insight into how social media algorithms can build legitimacy.

To address these research questions, this chapter first conceptualizes Influencers as politically relevant strategic actors. This is followed by a review of the literature on how professional communicators act in relation to powerful social media algorithms. The research questions are then answered based on a thematic analysis of online professional discussions about changes in Instagram's algorithms during 2019. While not all of the

analyzed material is written by Influencers, the lessons and advice described in the analyzed material are aimed at communicators who use Instagram professionally, including Influencers.

#### **4.2 Influencers: politically relevant strategic actors**

The political relevance of Influencers on Instagram should not be underestimated. Firstly, Influencers have become an important source of political information. In Denmark, for example, over half of high school students say they use Instagram as a source for news (Mehlsen, 2020). Especially during the Covid crisis, Influencers were significant by relaying information from the authorities to their followers, or occasionally questioning and challenging such information. The perceived interconnectedness and authenticity make the Influencers credible sources of information for their followers, especially among niche audiences who do not feel that their voices and worldview are represented by the mainstream media (Lewis, 2020). The political importance of Influencers becomes even clearer when the definition of politics is widened beyond institutional politics and also includes power struggles around topics such as body politics, social justice, political consumerism, or the environment. Lewis (2020) describes how some Influencers have found a niche audience and built a brand around criticizing mainstream media and social justice attitudes. On the other end of the political spectrum, Influencers who “promote feminism, sex positivity, LGBTQ awareness, and racial awareness have used the same micro-celebrity practices towards different ends” (Lewis, 2020, p. 215). Such Influencers can have a strong impact on regulation or put pressure on an industry to change their practices. Khamis et al. (2017) for example describe how Vani Hari, a food blogger, lobbied to demand more openness from the food industry and mobilized her followers to pressure industries to make changes in the use of chemicals in their products. Also, Influencers who are not primarily focused on branding themselves politically, engage in political acts, for example by sharing their personal #MeToo stories, or by supporting the Black Lives Matter movement. While such acts have at times been criticized as mere lip service or joining a band wagon, these messages create awareness among their followers (e.g., Paul, 2020). According to research by Maares and Hanusch (2020), lifestyle Influencers are aware of their societal relevance and are considerate of their ethical responsibilities in particular toward younger followers. Although they do not see themselves as part of the journalistic profession, they acknowledge that they fulfill journalistic roles, such as educating their followers, and providing service and advice. These are roles that in practice often involve addressing the audience simultaneously as consumer and citizen (e.g., Eide and Knight, 1999; Mellado and Van Dalen, 2017).

The interactions of Influencers with algorithms have been described as “conscious” and “instrumental” (e.g., Cotter, 2019, p. 895), highlighting that Influencers are strategic actors. This stands in contrast with the way many ordinary users interact with social media algorithms. As described in Chapter 2, many people are unaware of the role of algorithms on the social media platforms (Eslami et al., 2015). Those who are aware that algorithms play a role often have only very generic ideas about how they function (Eslami et al., 2016). Avid social media users, especially younger and tech-savvy ones, may adapt their behavior strategically in order to make the algorithms show the type of information that they are most interested in (Bucher (2018, p. 109). They may consciously click on specific content on Facebook or YouTube to affect which type of content shows up and is recommended. Bucher (2018) furthermore showed that the same goes for those who use social media platforms professionally to run a community or circulate content. They also adapt their tactics to make sure their information reaches a wide audience.

Such pro-active approaches underline that professional communicators have agency in relation to social media algorithms. At the same time, these communicators still have to operate within the possibilities and limitations which the social media algorithm gives them. Digital Influencers on Instagram can be perceived as “playing the visibility game,” where “algorithms structure, but do not unilaterally determine user behavior” (Cotter, 2019, p. 896).

### **4.3 Algorithmic power**

If we perceive the algorithms as a structuring power to which social media users react and adapt their behavior (see Chapter 1), the question becomes how social media users such as Influencers react to the algorithmic power. These reactions can fall on a continuum stretching from reacting to the algorithms as intended by the platform designers and owners, to on the other hand resisting and even fighting back against the influence of the algorithm. The determining factor in where Influencers are positioned on this continuum is how they perceive the locus of control (Lefcourt, 2014), i.e., the extent to which they believe that they themselves are in the driving seat versus believing that outside forces determine their behavior. On the one extreme of the spectrum, Influencers see the locus of control of their behavior as external and perceive themselves as powerless. On the other extreme of the spectrum, the locus of control is internal as Influencers perceive themselves as having agency and power to make their own decisions, which might go against the actions intended or incentivized by the algorithm.

When Influencers adapt their posting tactics to what is rewarded by the algorithms, this indicates that Influencers have limited power in

relation to algorithms. Examples of this could be the timing of the posts, use of specific hashtags, tagging people, or posting emotional content, which evokes feelings in other social media users.

The power relation between Influencers and the algorithms is more balanced when Influencers engage in behavior that gets rewarded by the algorithms but was not intended by the algorithm designers. Search engine optimization would be an example of using knowledge about algorithms to optimize outcomes (e.g., Velkova and Kaun, 2021). The participation in Pods is another example. Pods are groups of social media users, who agree to like and comment on each other's content in the expectation that the algorithms will then further promote this content. Here, the members of the Pods "game the system" by "acting on knowledge about the algorithmic system to effect certain outcomes" (Cotter, 2019, p. 899). Based on interviews with 16 Influencers, O'Meara (2019, p. 9) describes Pods as "organized efforts to manipulate the algorithm that has inserted itself between content creators and their audiences and are an attempt to take control of the process whereby they are dated, measured, and assigned value."

While Pods give Influencers more power, Pod members are not independent of the algorithms. They still react to algorithms and adjust their behavior to what is rewarded by algorithms. A more extreme form of taking back control would be instances where social media users engage in resistance against an algorithm. Velkova and Kaun (2021) described a campaign by Johanna Burai, a Swedish visual designer, with the goal of getting six non-white hands to show up on the top of the search list in Google Image. This campaign is an example of intervention on the side of social media users who resist features of social media algorithms which they are not satisfied with. While this is a form of resistance, Velkova and Kaun (2021) describe it as cautious resistance, since the campaign used knowledge about the working of the algorithm to make small corrections to biases imposed by the algorithm. A more rigid form of resistance would be to leave a social platform altogether in response to dissatisfaction with the working and outcomes of the algorithm. An announced change in Twitter's algorithm in 2016 led many Twitter users to threaten to resign from using the platform, using the hashtag #RIPTwitter (DeVito et al., 2017). This would be the ultimate example of taking back control from algorithms, either by "freeing" themselves from the demands of the platform or by putting pressure on social media organizations to adapt the working of the algorithm to avoid losing users.

These different responses describe the way social media users might react and gain agency toward the structuring power of social media algorithms. However, it has to be noticed that feeling in control does not always equate with actual control. Apte (2020) argues that social media architectures might give a user a feeling of autonomy, while in reality they

are implicitly controlled in their actions. This can be understood when seeing social media platforms as a nudge-based choice architecture. Choice architectures shape the context in which decisions are made and thereby influence the decisions. Minor changes in choice architecture, so-called nudges, can have big consequences for people's behavior. A nudge can be defined as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic perceptions" (Thaler and Sunstein, 2008, p. 6). Nudges do not literally force people to take certain actions but preserve people's freedom to choose. Still, the real intentions of the nudges are often not made explicit and therefore it is debatable whether nudging lives up to the requirements of informed consent. Drawing a parallel with nudging, Apte (2020) describes Facebook's choice architecture as a form of soft control that creates an autonomy-control dichotomy: on the one hand, users have explicit control, as they can decide what to like, who to follow, and who to add or remove as friends. On the other hand, the platform incentivizes certain behavior, like remaining active and monitoring the comments and shares which posts receive.

#### 4.4 Research questions

Against this background, this chapter addresses two research questions. To understand the dependency relation between Influencers and Instagram's algorithms, the analysis first looks into how Influencers perceive their own power in relation to the algorithms, assessing how dependent they are on the algorithms, what resources they are willing to offer in order to take advantage of the algorithms, and whether they have agency to resist the algorithms.

Research question 4.1:

*What is the perceived power of Influencers vis-à-vis Instagram's algorithms?*

The second research question is narrower and looks into the legitimacy of the power of Instagram's algorithms. As described in Chapter 1, dependency is not necessarily perceived as problematic, as long as the power that others have over one is considered legitimate. The second part of the analysis will look into instances where the acceptance of algorithms' power is discussed, to understand which principles and standards are applied to make such assessments.

Research question 4.2:

*Which criteria are used to assess the legitimacy of the power of Instagram's algorithms?*



#### 4.5 Method

Inspired by Cotter (2019), a thematic analysis was conducted of online articles in which Influencers and other professional communicators discuss changes to Instagram's algorithms. The purpose of this analysis is not to quantify how often certain views on the power of professional communicators vis-à-vis Instagram's algorithms are expressed, nor to compare groups of professional communicators. The analysis makes use of the strengths of thematic analyses by systematically describing common themes in the way professional communicators discuss the algorithms and examining what this reveals about the power relations with the algorithms and their perceived legitimacy.

The analysis is based on 96 online articles about Instagram's algorithms and changes in the algorithms written by Influencers, marketing experts, journalists, and other professionals in 2019. These articles are aimed at people who use Instagram professionally as they inform about the working of the algorithms and discuss strategies and tactics on when, what, and how to post on Instagram. Typical titles of these articles are *How the Instagram algorithm works (and where your strategy needs to shift)*; *15 expert secrets for Influencers to beat the Instagram algorithm* and *Influencers are okay with the recent changes on Instagram*. The analyzed articles were published in 2019, when Instagram discussed no longer making visible how many likes an image received. Several articles address this change, but the main focus of the analysis is the reflections on Instagram's algorithms and changes in the algorithms. Not all articles explicitly refer to Influencers as their target audience. However, such articles are still included in the analysis, when they addressed similar themes as the articles for Influencers and were relevant for individual professional Instagram users.

The articles were collected in April and May 2020 based on keyword searches on Google Search. In the first step, the generic keywords *Instagram algorithm 2019* and *Instagram algorithm change 2019* were used. This was followed by more specific Google searches, based on words that appeared in some of the articles that were found by using the generic terms (such as *Instagram algorithm 2019 influencer* or *Instagram algorithm 2019 resistance*). For all searches, the time period was set from 1 January 2019 to 31 December 2019. Up to 200 results were checked for each search term, stopping when suggested articles were no longer relevant. Each article was checked manually to make sure that it deals with Instagram's algorithms and is aimed at individual professional Instagram users. This resulted in 70 relevant articles.

During the analysis, it became clear that Instagram Pods play an important role in Influencers' resistance to algorithms. Therefore, additional Google searches were undertaken for online articles published

in 2019 where marketing experts, journalists, Influencers, or other professionals write about Instagram Pods. These searches used keywords like *Instagram algorithms pods 2019* or *Instagram engagement groups 2019* and resulted in 26 articles about Instagram Pods, such as *IG pods: How your friends will help your brand grow on Instagram* or *Can Instagram Pods help you beat the algorithm?*

The coding of the materials started with the development of a codebook. This codebook was developed based on previous research on perceptions of algorithms reviewed in this book but was fine-tuned during coding in an iterative process. The final coding scheme consisted of four main categories:

- *Awareness*: Used for statements showing recognition of the presence of algorithms on Instagram and revealing that this awareness influences behavior on Instagram. The category contains sub-codes that identify which characteristics of the algorithms professionals are particularly aware of; negative and positive reactions to the algorithms; where they gain their knowledge about the algorithms from; and what specific advice they offer other professionals on how to handle these algorithms.
- *Metaphors*: This code is used for statements in which the professionals use metaphors about the algorithms and how to handle them. For example, “I can help you beat the Instagram algorithm.”
- *Transparency*: This upper category is used for statements concerning whether the professionals perceive Instagram algorithms to be transparent. The category contains sub-codes that describe rumors or myths about Instagram algorithms (e.g., describing the algorithm as mysterious, demystifying the algorithms by debunking myths).
- *Trust and distrust*: This category includes instances indicating that the professionals trust or distrust the Instagram algorithms. This includes a sub-category for statements where professionals describe the Instagram algorithms as fair (e.g., referring to them as neutral, objective, or unbiased) or unfair (e.g., by describing them as prejudiced, discriminatory, biased, or rigged).

The articles were coded using qualitative data analysis software Nvivo. After coding the articles, all text elements were sorted according to the relevant sub-categories. Next, the texts in the sub-categories were further analyzed in-depth to search for common patterns in the perceived power relation of professional Instagram users vis-à-vis the algorithm, using the theoretical concepts of power, dependency, locus of control, agency, autonomy, and legitimacy.

## 4.6 Results

### *Dependence on the algorithms*

According to the online articles, Influencers and other professional communicators are dependent on Instagram's algorithms because these algorithms control access to several resources. The articles mention that algorithms affect how much users interact with individual posts, measured as engagement (understood as commenting or liking), organic reach (number of people who viewed a post), interaction rate (engagement relative to the number of followers of an account), and lifespan of a post (time before the post no longer attracts new views or engagement). This in turn spills over to the number of followers on one's account, sales of products advertised in the posts, and new customers. Since these are important resources for Influencers, there is a strong feeling of dependence on the algorithms. The feeling of dependence is strengthened by the unpredictability and lack of understanding of the algorithms. On the one hand, the articles argue that "understanding the algorithm is crucial" (Hart, 2019). On the other hand, it is highlighted that "nobody understands the algorithm" (Hart, 2019). This makes people highly attentive and sensitive toward Instagram's algorithms:

*Since the death of the chronological timeline, Instagram's constant experimentation with ways to bring you the posts on your feed has led to increasingly low reach for both normal users and Influencers. (...) Algorithms are the pain points of marketers all over the world. They change all the time, they are unpredictable, and we are dangerously dependent on them.*

(Eloise, 2019)

This strong reliance on Instagram's algorithms also becomes clear in the feelings expressed when discussing changes in algorithms, as the articles refer to frustration, confusion, panic, shock, and even hate when mentioning shifts in the way Instagram works.

Following Dahl (1957, p. 203), power is "a successful attempt by A to get B to do something that he would otherwise not do." The articles describe several tactics that Influencers could follow in order to take advantage of the way the algorithms work. The articles describe that Influencers should be willing to adapt the timing of the posts, focus posts on the most engaging audiences, post more of the high-performing content, and vary the type of content which they are posting. Another way to benefit from the way the algorithms work is to interact and engage directly with followers on a daily basis. This interaction is not presented as a goal in itself, but rather as a means to get rewarded by the algorithm with more reach and engagement of the posted content. Other

tactics which were mentioned include deleting content if it does not go viral in the first 30–60 minutes.

When Influencers adapt their strategy to such advice, they are essentially giving in to the perceived power of the algorithms. This could mean that Influencers are more willing to create posts about topics that are likely to resonate with a small group of highly engaged Instagram users, or posts which trigger an emotional reaction, or pay lip service to ideas that are popular at the time. Thus, such reactions are biased toward publishing about specific topics and opinions while ignoring others. This is less problematic if these are true expressions of the Influencers' own beliefs and interests. However, it could lead to not addressing certain societally relevant topics or holding back or toning down opinions, which they believe will not fly due to the logic of the algorithm.

### *Power balance perspective*

Figure 4.1 shows illustrative quotes of how professional communicators perceive their own power and agency vis-à-vis Instagram's algorithms. On the top pole of the continuum, the power balance is totally tipped in favor of the algorithms. The first two quotes illustrate a position where the Influencers are totally dependent on the algorithms. A change in the algorithm is presented as working directly against the interests of the professional communicators. It forces them to humbly ask their followers to actively make sure that their posts are seen. The third and fourth quotes reveal a different perspective. Here, the power balance is still tilted in favor of the algorithms, but now the algorithms are also seen as offering opportunities. The users can now influence with their own actions whether the algorithm works in favor of or against their interest. The professional communicators can benefit if they play along and adapt to the type of posting behavior, which the algorithms reward. In the fifth and sixth quotes, we see a shift in the power balance, acknowledging that professional communicators have agency and do not have to be dependent on the algorithms. If influencers follow the tips offered in the articles in which these quotes appear, they can "beat the algorithm." They can bring their posts to the attention of their followers and other Instagram users despite the perceived negative influence of the algorithms. The two quotes at the bottom pole of the continuum reflect instances where the Influencers have agency and are not seen as dependent on the algorithms. Influencers are encouraged not to focus on the algorithms, which they do not control, but instead to focus their energy and efforts on how to provide the type of content that their followers love and want to engage with. These quotes put the perceived dependence on the algorithms, which many professional communicators feel, in a different light. Complaints

**Algorithms > Influencer**Dependence

“Then, a shift in algorithms *pulled the rug from under me*” (Baggs, 2019)

“It was a time of widespread panic, of *Influencers begging* followers to “turn notifications on” so users would be alerted whenever they posted, and never miss a thing despite the pesky algorithm” (Kircher, 2019)

--

Opportunity

“Instagram’s algorithm develops faster than the shelf life of a cake in the Spindogs kitchen, presenting a huge *opportunity* for brands to reach customers far and wide.” (Headlam, 2019)

“On Instagram, hashtags are an integral tool for getting your content in front of the right eyes—that is, the eyes most likely to engage with it. They *can help* create a virtuous circle where, as your post gets more attention, the algorithm will make sure even more eyes see it.” (Standberry, 2019)

--

Beating the algorithm

“How to survive and *outsmart* the Instagram algorithm” (Barnhart, 2019)

“Instagram pods can *help you beat the Instagram algorithm* and get your content in more users’ feeds. In some cases, they may be able to help boost your content onto Instagram’s “Explore” page, which can give you viral exposure” (Grossman, 2019)

--

Take back control

“*Instead of trying to find a way around* the Instagram algorithm, you may instead want to invest in finding ways to engage the users who are most likely to be interested in your products.” (Sherman, 2019)

“The Instagram algorithm does not make any significant changes regularly. The algorithm likely did not change. *You cannot put the blame for your content performance on anyone but yourself.*” (Herman, 2019)

**Influencer > Algorithms**

Figure 4.1 The relation between Influencers and Instagram’s algorithms as a balance of power.

about the algorithms are here seen as a form of scapegoating or a cheap excuse for bad performance of one’s posts, for which one should take responsibility oneself. In sum, the quotes reveal that there is not one dominant perspective in the perceived power balance between Instagram algorithms and Influencers. Dependence on algorithms is only one perspective, and on the other side of the spectrum, we find a perspective where the Influencers are in control. Still, the articles did not reveal any examples of open resistance. None of the articles encouraged Influencers to leave the Instagram platform as a reaction to the perceived negative influence of changes in the algorithms.

*Perceived versus real autonomy and control*

The articles reveal an interesting paradox about the agency of Influencers vis-à-vis Instagram's algorithms: Influencers can feel in control and have agency as long as they follow the posting strategies which the algorithms reward. This means posting more actively, more direct communication with followers, and posting more content that resonates with the wishes and interests of followers. This raises the question of how much Influencers are really in control. Having agency on Instagram does not mean that one can take a longer break from posting and replying. Bringing up topics or opinions that are not seen as relevant and timely will probably have negative consequences for the performance of their accounts. There are clear demands that they have to live up to and boundaries within which they have to operate:

*As long as you are creating engaging, relevant, and timely content, the algorithm is actually an advantage to you.*

(Lua, 2019)

A similar paradox is visible in the discussion of Instagram Pods. These online groups of Instagram users, who help to boost each other's content by liking and commenting on the content of the other members of the Pod, are presented as a way to game the algorithms. Posts with many initial likes and comments from other Pod users are expected to give a leg up and gain larger visibility than posts that did not receive such support. However, there are also several downsides about the Pods, which make it questionable to which degree they actually give power and control to Influencers. First, taking part in Pods also requires a lot of resources, since time needs to be spent liking and commenting on posts by other members of the Pod. Second, being part of a Pod might force Influencers to like posts with questionable quality or posts that one does not genuinely agree with, only because this is part of the deal of being a member of a Pod. Third, while Pods might increase the engagement of one's posts, a large part of this engagement will be from other members of the Pod, rather than from the target group who the Influencers actually want to reach:

*If your goal for being on Instagram is only to get as many likes, comments, and followers as possible, then this isn't really a problem for you. However, if your motivation for being on Instagram is more profound than that (and it really should be), then being in an Instagram pod may not be as beneficial as you thought.*

(Chappell, 2019)

Furthermore, Pod members are warned that Instagram might suspend their accounts as their actions might be perceived as violation of the Terms of Service. If their engagement suddenly reveals a high spike, or if the number of likes their post receives is much higher than the number of followers, Instagram could pick this up as a red flag. Ultimately, this could result in a shadow ban, which means that Instagram restricts access of other users to one's profile. In sum, attempts to by-pass or gain control over Instagram's algorithms by joining a Pod simply make Influencers dependent in different ways. Instead of joining a Pod, focusing on pursuing one's authentic goals and creating true connections are presented as truly liberating and a genuine alternative:

*Next time you log into Instagram, consider why you're on the platform in the first place and try creating content that furthers your original purpose for joining. There are no easy answers for growing your following on Instagram, but shortcuts may ultimately hurt you in the long run.*

(Chappell, 2019)

### ***Legitimacy criteria***

Given the perceived power of Instagram's algorithms, the question becomes whether this power is seen as legitimate. If power is seen as legitimate, people will be more willing to accept being in a dependent position. Four types of criteria are used by professional communicators when discussing whether the power of Instagram's algorithms is legitimate or illegitimate: the character of the people behind the algorithms; the values embedded in the algorithms; the target group who benefits from the algorithms (who is it for?); and the outcomes produced by the algorithms.

### *Character*

The personal character of professionals can be a source of legitimacy for an institution. However, none of the analyzed articles was positive about either the executives behind social media platforms or the programmers behind the algorithms. As this quote from 2017 reveals, being clever can even be seen as disqualifying:

*"A stupid algorithm created by really smart people" (...) "these extremely book smart—but clearly not street smart—IG executives."*

(Tori, 2017)

Thus, viewing the professionals behind the algorithms as having a high IQ does not increase the legitimacy for the algorithms and even seems to

undermine their legitimacy. The articles do not refer to other personal characteristics of the people behind the algorithms, such as their integrity, community spirit, or trustworthiness.

### *Values*

The values which professional users refer to when assessing whether Instagram's algorithms and changes in the algorithms are just, include fairness, non-commerciality, equal opportunity, and openness to newcomers. These criteria are very similar to the ones that are applied when assessing the legitimacy of news media or other institutions with a public function. Apart from these values, which should be embedded in the algorithms, the discussions also mention values that are not directly embedded in the algorithms but are more fundamental. A lack of predictability, consistency, and transparency make the professional communicators critical toward the algorithms and less willing to accept their power.

### *Who is it for?*

Legitimacy can be claimed with reference to the people who benefit from the actions and services provided. When professional communicators discuss whether Instagram's algorithms have positive or negative effects, they often refer to the general users and whether (changes in) the algorithms make using Instagram a better or worse experience for them.

However, there were also cynical reactions, where claims that changes are made to benefit the user are met with skepticism. This became clear when Instagram proposed to remove the possibility to see how many likes particular posts received. The motivation behind this change was presented as a desire to make using Instagram a psychologically healthier experience and to remove the pressure to constantly compare oneself to others. As the following quote illustrates, some users did not believe that Instagram cares about its users, but instead is only interested in what benefits the platform:

*“An agent for big-name influencers anonymously told Hack he was cynical about Instagram’s ‘mental health’ justification, and suspected removing the ‘likes’ count would give the company more freedom to change the algorithm and “roll-out aggressive marketing tools.”*

(Purtill, 2019)

This quote exemplifies that although social media platforms and their algorithms can derive legitimacy if they are perceived to improve the user



experience, there is skepticism that the platforms look beyond their own interests.

Some of the analyzed articles expressed a concern that changes in Instagram's algorithms will benefit established users who already have a large following and a strong organization behind them, while it will be increasingly difficult for citizens with less followers to get their message across:

*In finding out how the Instagram algorithm works, one of the things you'd primarily consider is that encouraging users to spend considerable time is Instagram's topmost priority; undeniably, Instagram users come across more ads the longer the time they spend on the platform. In that case, an Instagram account that fulfils this priority stands the chance of raking in benefits. Broadly, this means in one way or the other, Instagram rewards brands or businesses that have established an enormous presence by maximizing the time they spend on the platform.*

(Forster, 2019)

#### *Outcomes*

As the quote above illustrates, there is concern that changes in the algorithms lead to more inequality of access on social media and have the negative consequence that debates on Instagram will be biased toward established interests. When assessing the legitimacy of algorithmic power, another important criterion is the effect of algorithms on people's opportunity to express themselves freely and connect to other Instagram users. This points to the ideals of community, shared experience and allowing everyone to bring their own perspectives into the broader societal debate. These worthwhile outcomes are central in discussions about removing likes from Instagram.

*I actually think it's the best idea that Instagram's put forward in a long time. Sometimes, it feels like a piece of content on Instagram is now only as good as the amount of engagements it gets. And for a social networking app that prides itself on the creativity of its users, that's the complete opposite of what should be happening. This app should be about high-quality content that inspires, uplifts, motivates or brings awareness.*

(Matt Benfield in Hensley, 2019)

*I think Instagram started as an amazing platform where people (could) be creative and share their perspectives, but then things like "likes" and "feed aesthetics" took over and genuinely ruined the platform. (...)*

(Mina Gerges in Hensley, 2019)

#### **4.7 Conclusion**

Instagram's algorithms have power over professional communicators like Influencers, because they control access to important resources like engagement and views. This power is strengthened by the fact that the algorithms are non-transparent and unpredictable. Perceptions of the power of Influencers toward the algorithms range from total dependence; to seeing the algorithm as opportunities which one can collaborate with; to perceiving a competition with the algorithms, which can be won; to independence and autonomy. There were no examples of full-blown resistance to the algorithms by leaving the platform. Even in the instances where Influencers can be seen as having the upper hand over the algorithms, there are still limits to their autonomy as they need to adhere to posting strategies, which are rewarded by the algorithms, or, when joining a Pod, replace dependence on the algorithms with the need to adhere to the rules of the Pods.

The power balance between Influencers and algorithms presented above has important democratic implications. There is a general concern that the way algorithms perform plays into the interest of organizations with large resources and established opinion leaders, while civic voices experience greater difficulty in getting their views across on the platform. The posting strategy, which is rewarded by the algorithms, tends to favor mainstream perspectives. Influencers are encouraged to focus their posts on the most engaged audience, post more high-performing content, or delete content that does not perform well. This could mean that certain important topics are not addressed and that there is a limited range of opinions, which can be expressed. The need to build a strong personal brand can also mean that there is limited room for acknowledging doubt or changing one's beliefs. Advice on how to by-pass or game the algorithms does not seem to remedy these negative democratic consequences since it comes with the danger of insincere communication. More engagement with followers is one such strategy, but this engagement is primarily presented as a means to get more likes and views, rather than an end in itself. Likewise, being part of a Pod can force Instagram users to like or express support for posts, which they do not honestly approve of. The analysis also showed voices that counter these negative democratic consequences, expressing concern about the lack of diversity or advocating for genuine engagement with followers or true connection. However, these voices are a minority in the analyzed material.

Criteria for assessing the legitimacy of algorithmic power to some extent mirror criteria used to assess the legitimacy of traditional institutions such as the news media. When discussing whether changes in the algorithms are positive or negative, the analyzed material refers to "traditional" democratic values, such as fairness, equal opportunity, and

transparency. Other important criteria are how algorithms affect target groups like individual users and civic voices. At the same time, there was also a difference from the criteria used to assess the legitimacy of the news media, as there were hardly any references to the personal character of the professionals behind the algorithms.

Using the described criteria, the legitimacy of Instagram's algorithms is generally negatively assessed. The debate is dominated by a strong sense of frustration, lack of trust, and even cynicism about Instagram's algorithms and the Instagram platform more broadly. If the algorithms were viewed as more legitimate, Influencers would probably feel less desire to by-pass or game the algorithms, and instead be willing to accept that those are the rules of the game. Instagram's algorithms seem to derive their power from the threat they pose to users who might lose their followers. This is very different and ultimately less sustainable than developing trust in the algorithms and convincing users about positive embedded values.

# 5 Algorithmic Power, Trust, and Legitimacy

## 5.1 Algorithmic power

This final chapter reflects on the findings of the different chapters in light of the aims informing the book. What lessons did the empirical chapters provide about the power, trust, and legitimacy of algorithmic gatekeeping and the challenges it poses for social media platforms, audiences, and professional communicators? Based on these lessons, the chapter suggests concrete initiatives to strengthen the trustworthiness and legitimacy of news algorithms.

The first aim of this book was to make the power of algorithmic gatekeeping tangible. The analysis of the role of YouTube's search and recommendation algorithms in spreading and correcting misinformation (Chapter 3) nuances technologically deterministic discourses, which dominate debate around social media algorithms. Challenging purely pessimistic views on algorithmic gatekeeping, the analysis showed that YouTube's search and recommendation results shield viewers from directly life-threatening misinformation. Still, not all information suggested by YouTube is supported by scientific consensus. This shows how algorithms play a mediating role and affect the distinction between legitimate and illegitimate perspectives. In line with the filter bubble argument, after watching a misinforming video, it is more likely to recommend a video confirming this misinformation than a video correcting it. The analysis also showed some unintended effects of algorithmic gatekeeping: after watching a video correcting misinformation, the algorithms at times recommend videos that directly contradict scientific consensus. Underlining the fact that algorithmic power is situated in the relationship between platforms and the people using these platforms, small differences in search terms or videos watched are consequential for the perspective provided in the subsequently recommended videos.

While Chapter 3 focused on the *mediating* power of algorithms in selecting and prioritizing information, Chapter 4 analyzed the *structuring* power of algorithms over the behavior of professional communicators.

Influencers and other professional communicators are dependent on algorithms for access to their followers. Therefore, they adapt their actions to the type of behavior that these algorithms reward. The non-transparent and unpredictable character of these social media algorithms leads to frustration and even cynicism, but no examples were found of professionals considering leaving the platform. Resistance to algorithmic power is confined within the structures provided by the algorithm and does not lead to true autonomy. The structuring powers could have democratic consequences, as they may affect the type of issues Influencers address, the range of voices available, and sense of community on these platforms.

Together the analyses of the mediating power and structuring power of algorithmic gatekeepers challenge technological deterministic visions and show that algorithmic power is situated in the interplay between platforms, audiences, and professional communicators.

## **5.2 Algorithms, trust, and legitimacy**

The second aim of this book was to study trust in and the legitimacy of algorithmic gatekeeping. The general audience cannot assess or comprehend the exact working of complex algorithms. Therefore, a willingness to use them and be vulnerable to their actions requires that the algorithms are trusted. As shown in Chapter 2, only a small minority of the general population says that they trust news selected or written by algorithms. These low levels of trust are not triggered by negative connotations of the term “algorithm,” but seem to be the results of a general skepticism about their role in news selection and creation. The general population is particularly skeptical about the possibility that algorithms can provide objective, balanced, and diverse information. Still, people are more accepting of algorithms playing a role in news selection than in other areas of public life, where algorithms do not yet play such a dominant role. This could indicate that when more people become familiar with algorithms fulfilling journalistic tasks, they also become more willing to accept them. This interpretation is supported by the finding that acceptance is strongest among Generation Z, for whom algorithms have always played a role in the selection of their news diet. The general population was more willing to accept algorithms when they select news in collaboration with humans than when algorithms fulfill this task alone.

Power is legitimate when it is perceived as just and fair. When discussing whether the power of algorithms is legitimate, professional communicators assess whether the right values are embedded in the algorithms; which target groups benefit from the working of the algorithm; and which positive outcomes the algorithms produce. The character or skills of the people behind the algorithms (such as programmers or executives behind social media

platforms) do not grant the algorithms legitimacy. This is in contrast to other institutions such as the news media or bureaucracy where the character of the professionals is seen as a source of legitimacy. Important values, which grant algorithms legitimacy, are perceived non-commerciality, equal opportunity, and openness to newcomers. When assessing algorithms based on these criteria, the legitimacy of their power is primarily assessed negatively. Changes in algorithms are seen as fair and just when they are perceived to be in the interest of the general users. However, there is skepticism among professional communicators that social media platforms are concerned about the interest of others beyond themselves. Finally, the power of algorithmic gatekeeping is seen as legitimate when it contributes to the common good, in particular by fostering community, the creation of shared experience, and the opportunity for everyone to bring their own perspectives into the broader societal debate.

### **5.3 Implications**

The findings presented in this book have important implications for the legitimacy of algorithmic gatekeeping. Firstly, this legitimacy is undermined by the lack of trust in news algorithms. Even among the younger generations, who are generally more willing to accept that algorithms play a role in news selection, there is a general lack of trust in news algorithms. This is something that needs to be taken seriously, and efforts should be made to build trust. Based on the literature discussed above, a lack of understanding of the way algorithms work in combination with negative debates about filter bubbles and echo chambers may be a breeding ground for skepticism about algorithms, especially when they are directly compared to human journalists. An important step toward building trust will be to improve the performance of news algorithms. More transparency around news algorithms and their working, more possibilities for individual control of algorithms, and more willingness to acknowledge possible negative side effects might be important first steps toward building trust. Regulatory initiatives, such as the European Digital Services Act, point in this direction.

However, one has to be careful not to expect too much from more openness, explanation, and more transparency about algorithms alone. As Ananny and Crawford (2018, p. 10) have argued, “[i]t is difficult to make a complex object of study transparent, especially in ways that create the authoritative knowledge that defensible accountability requires.” As the discussion here has made clear, algorithms are inherently complex, adaptive, and opaque. At the same time, the general public lacks the deep understanding of algorithms, which would be required in order to hold algorithms accountable based on the transparency of their working. More

generally, evidence that transparency increases trust is at best mixed. Some scholars have even suggested that transparency might replace rather than foster trust (see Karlsson et al., 2014; Peifer and Meisinger, 2021; Van Dalen, 2019a). Trust is as much determined by the truster as the object of trust. Therefore, it might take a long time to build and rebuild trust among populations who are used to thinking about the selection of news and public information as a traditionally human task. As people get more experience with news algorithms, they may grow more familiar with them and develop more trust. This could especially be the case when they feel a personal bond with the social media platforms and organizations that use news algorithms.

New regulatory initiatives, such as the European Digital Services Act, point toward more regulatory oversight of algorithm-driven social media platforms. This could increase trust in algorithmic gatekeeping among the general public. However, these regulations seem to focus mainly on limiting negative consequences such as hate speech or misinformation rather than providing direction for how these algorithmic gatekeepers can make a positive contribution to the public sphere, by informing citizens, building community, or fostering deliberative debate. Strongly regulated algorithmic gatekeepers might risk becoming politicized, which could lead to polarization of trust based on people's political leaning.

Letting algorithms and journalists work together could be an important step toward building trust in news algorithms, as it could combine the strengths of news algorithms (such as selecting personally relevant content) with the perceived strength of human journalists compared to algorithms (such as being objective and balanced). The same might hold true for social media platforms where human moderators or boards of oversight could be ways to incorporate human control over algorithms. Here, journalists and other media professionals could take on a role similar to the one that elevator operators once had. When the first lifts were installed in warehouses and high-rise buildings, they were not as technically advanced as they are now. At the same time, the public was not familiar with their working. Elevator operators would make sure that elevator users arrive at the right floor and help out in case of emergency. Equally important, they would make sure the elevator users would feel at ease when using this new technology. In a similar way, journalists and other media professionals could provide oversight of news algorithms, step in when algorithms make mistakes, and be the “face of the algorithms” toward users. In this way, they could help users get familiar with algorithms fulfilling journalistic tasks and in the long run build trust in news algorithms. This is particularly important in the early stages of news algorithm development and implementation. As news algorithms become more advanced and reliable, the role of humans could be phased out more and more, just like the elevator operator has now become a relic from the past.

Without a strong role for humans at this stage; however, it might be very difficult to build trust later even if news algorithms become more advanced and reliable.

One caveat that needs to be taken into account here is that the results presented in Chapter 2 are based on research in Denmark, a country belonging to the democratic corporatist media system dominated by media outlets with a public service ethos (Hallin and Mancini, 2004). Trust in journalism is higher in Nordic countries belonging to the democratic corporatist media systems than in countries belonging to the liberal media system, polarized pluralist media system, or younger democracies (e.g., Hanitzsch et al., 2018; Van Dalen, 2019a). Fletcher and Nielsen (2019) have shown that in other media systems, social media users are generally skeptical to selection of information, both by algorithms and by humans. This suggests that letting algorithms and journalists work together could be a more effective way of building trust in some media systems than in others. In this regard, it is worrying that the character and competences of developers and platform executives are not seen as a source of legitimacy for the algorithms they design and oversee (see Chapter 4). This might limit the effect of human–algorithm collaboration on trust and perceived legitimacy.

Algorithms come to play an increasingly important role in the gatekeeping process and affect all aspects of current affairs information which reaches the audience. Thus, trust in algorithms might become a central element in people’s generalized trust in news and public information. Nowadays, trust in the media and democratic institutions more broadly is to a large extent a consequence of people’s interpersonal trust (e.g., Van Dalen, 2019a). This shows how much importance the general public attaches to the people working in these institutions when assessing the credibility of the outcomes of their workings. When algorithms take over and supplement tasks performed by humans, the importance of interpersonal trust as driver of generalized trust in the news might diminish. At the same time, trust in technology and science might become a more important driver of trust in the news. This could affect trust in news among the population at large, but also change how demographical backgrounds affect trust in the news. Surveys in the United States suggest that certain automated decisions, such as automated personal finance scores are seen as fairer by Blacks and Hispanics than by Whites. At the same time, more Blacks and Hispanics than Whites believed that news on social media accurately reflects society (Smith, 2018). This could mean that these groups are more skeptical about the decisions which human gatekeepers make than about algorithmic decision-making. If this is the case, they might become more trusting of news in general if automated decision-making replaces human biases.



A strong sense of ethics of developers as well as institutional processes to counter their individual biases could be a way forward to more legitimate and trustworthy news algorithms. Friedman and Nissenbaum (1996) have argued that preexisting biases on the individual or societal level can translate into biased computer systems. Raising awareness of these biases is an important first step toward countering them. Measures to counter such biases could on the one hand take a principled approach and define key values, which should guide the design process independent of the outcomes (DeVito, 2017). This could include values such as fairness, equal opportunity, openness, non-commerciality, or predictability and consistency. On the other hand, biases could be countered taking a consequential approach by formulating criteria or metrics that could be used to assess the outcomes of the algorithms (e.g., Mitchell et al., 2019). For example, exposure diversity has been proposed as a design principle for news recommender systems (Helberger et al., 2018). Exposure diversity is just one value, which could be operationalized as metrics to assess the performance of algorithms. Other criteria could be formulated to assess positive outcomes such as equal access of different actors, contribution toward community building, the presence of inspirational content, or the absence of misinformation or hate speech.

Discussions about which values should be embedded in news algorithms and algorithms on social media platforms raise questions about which information these algorithms should prioritize and which information should be corrected (e.g., Cobbe, 2021). There will not be much debate about the successful efforts of YouTube to ban videos about Mineral Miracle Solution, but questions such as whether it is appropriate that videos are available where religious or spiritual practices are presented as ways to cure autism, remain moot. In many countries, freedom of speech will protect this type of message, even though they go against scientific consensus. Scientific consensus would in many cases also be a difficult criterium to use to distinguish misinformation from correct information. This would require continuous updating of the algorithms which distinguish between correct information and misinformation. In many cases, it takes a long time for scientific consensus to emerge, if it emerges at all. Science is an ongoing debate, and paradigms and scientific consensus change over time. More broadly, such debates tie into the need for social media platforms to balance different interests and goals, such commercial versus public interests, the interests of individual users versus the larger community, or protecting freedom of speech versus limiting potential damaging messages. Given the enormous amount of content that is available and continuously added to YouTube and other social media platforms, actively banning content requires the automated detection of such content. Another argument in favor of automated detection of information, which should be banned, is

that this is a highly psychologically demanding task for humans to carry out (Gillespie, 2020). Nevertheless, for the moderation of content with the use of algorithms to be seen as legitimate, it would be good to keep humans in the loop in some way (see Gillespie, 2020 for concrete suggestions on how this could take shape).

Still, even if the performance of news algorithms improves and humans are involved in the process, it is questionable whether this by itself will be sufficient to build (or restore) trust in such algorithms. The literature on algorithm aversion described in Chapter 2 suggests that even when algorithms perform better than humans, people are not always willing to rely on them. Addressing this aversion requires more than only better performance of algorithms and increasing algorithm literacy could be a promising way forward. Educating the broader public about algorithms does not require detailed knowledge or a deep understanding of mathematical formulas. Logg et al. (2019)'s concept of the theory of machine could provide a useful way forward to developing such algorithm literacy programs (see also Burton et al., 2020). People make inferences about motives and future actions of other humans based on some sort of theory of mind. Although we do not necessarily understand all the neural processes in detail, our broad ideas about how the human mind works allow us to interact and connect with other human beings. Similarly, algorithm literacy training could allow people to form a theory of algorithmic mind (Burton et al., 2020, p. 227), which helps to predict the working of algorithms and why they make certain decisions. Such a model of algorithmic mind could help people predict the working of news algorithms and understand what they are capable or not capable of. In the words of Burton et al. (2020, p. 227), "human-algorithm coordination needs not human agents who grasp the code behind the algorithmic aid, but rather a high-level model of its purpose and perception." Algorithm literacy training with a focus on such a model could increase trust in algorithms as well as their perceived legitimacy.

This book started out by stressing the increasingly important role of algorithms in the gatekeeping process. The empirical chapters have shown that to understand the power of these algorithms, one needs to study their interaction with professional communicators and the audience at large. As argued in this final chapter, initiatives toward trusted and legitimate algorithmic gatekeeping should take their starting point in the humans who develop, interact with, and are affected by algorithms rather than in the algorithms themselves.

# References

- Abidin, C. (2015). Communicative♥ intimacies: Influencers and perceived interconnectedness. *Ada: A Journal of Gender, New Media & Technology*, 8: 1–16.
- Airoldi, M., Beraldo, D., & Gandini, A. (2016). Follow the algorithm: An exploratory investigation of music on YouTube. *Poetics*, 57: 1–13.
- 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.
- Andreassen, A. M. (2020, October 16). Ups! Altinget-robot gjorde Poul Schlüter til statsminister [Oops! Altinget-robot made Poul Schlüter prime minister]. *Journalisten*. Retrieved on February 16, 2022, from: <https://journalisten.dk/ups-alinget-robot-gjorde-poul-schluter-til-statsminister/>.
- Apte, D. S. (2020). Explicit autonomy, implicit control: User autonomy in the dichotomous choice architecture of Facebook. *Journal of Creative Communications*, 15(2): 165–176.
- Araujo, T., Helberger, N., Kruikeimeier, S., & De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, 35(3): 611–623.
- Auxier, B. (2020, October 15). 64% of Americans say social media have a mostly negative effect on the way things are going in the U.S. today. *Pew Research Center*. Retrieved on February 15, 2022, from: <https://www.pewresearch.org/fact-tank/2020/10/15/64-of-americans-say-social-media-have-a-mostly-negative-effect-on-the-way-things-are-going-in-the-u-s-today/>.
- Avaaz (2020). *Why is YouTube broadcasting climate misinformation to millions?* Avaaz Report. Retrieved on February 16, 2022, from: [https://secure.avaaz.org/campaign/en/youtube\\_climate\\_misinformation/](https://secure.avaaz.org/campaign/en/youtube_climate_misinformation/).
- Baggs, R. K. (2019, November 16). Why Instagram removing Like counts is the greatest change the platform has ever made. *Fstoppers*. Retrieved on June 7, 2020, from: <https://fstoppers.com/originals/why-instagram-removing-counts-greatest-change-platform-has-ever-made-428356>.
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239): 1130–1132.

- Bandy, J., & Diakopoulos, N. (2020a). Auditing news curation systems: A case study examining algorithmic and editorial logic in Apple News. In *Proceedings of the International AAAI Conference on Web and Social Media*, 14: 36–47.
- Bandy, J., & Diakopoulos, N. (2020b). # TulsaFloP: A Case Study of Algorithmically-Influenced Collective Action on TikTok. *arXiv preprint arXiv:2012.07716*.
- Bandy, J., & Diakopoulos, N. (2021). Curating quality? how twitter’s timeline algorithm treats different types of news. *Social Media+ Society*, 7(3): 1–17.
- Barnhart, B. (2019). How to survive (and outsmart) the Instagram algorithm. *Sprout Social*. Retrieved on June 7, 2020, from: <https://sproutsocial.com/insights/instagram-algorithm/>.
- Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, 181: 21–34.
- Bode, L., & Vraga, E. K. (2015). In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication*, 65(4): 619–638.
- Bodó, B., Helberger, N., Eskens, S., & Möller, J. (2019). Interested in diversity: The role of user attitudes, algorithmic feedback loops, and policy in news personalization. *Digital Journalism*, 7(2): 206–229.
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415): 295–298.
- Boulding, K. E. (1968). The legitimation of the market. *Nebraska Journal of Economics and Business*, 7(1): 3–14.
- Boulding, K. E. (1971). The legitimacy of economics. In: F. R. Glahe (Ed.). *Kenneth Boulding: Collected Papers: Economics*, 2, pp. 417–425. Boulder, CO: Associated University Press.
- Boulianne, S., Koc-Michalska, K., & Bimber, B. (2020). Right-wing populism, social media and echo chambers in Western democracies. *New Media & Society*, 22(4): 683–699.
- Bruns, A. (2019). *Are Filter Bubbles Real?* Cambridge: Polity Press.
- 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.
- Bucher, T. (2018). *If... Then: Algorithmic Power and Politics*. Oxford: Oxford University Press.
- Burgess, J., & Green, J. (2018). *YouTube: Online Video and Participatory Culture* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons.
- Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2): 220–239.
- Carrie Wong, J., & Levin, S. (2019, January 25). YouTube vows to recommend fewer conspiracy theory videos. *The Guardian*. Retrieved on February 15, 2022, from: <https://www.theguardian.com/technology/2019/jan/25/youtube-conspiracy-theory-videos-recommendations>.
- Centers for disease control and prevention (2022a). *Signs and Symptoms of Autism Spectrum disorders*. Retrieved on February 15, 2022, from: <https://www.cdc.gov/ncbddd/autism/signs.html>.

- Centers for disease control and prevention (2022b). *Treatment and Intervention Services for Autism Spectrum Disorder*. Retrieved on February 15, 2022, from: <https://www.cdc.gov/ncbddd/autism/treatment.html>.
- Chappell, D. (2019, February 14). You should think twice before joining that Instagram Pod. *Fstoppers*. Retrieved on June 7, 2020, from: <https://fstoppers.com/originals/you-should-think-twice-joining-instagram-pod-340121>.
- Chen, A. Y., Nyhan, B., Reifler, J., Robertson, R. E., & Wilson, C. (2021). *Exposure to Alternative & Extremist Content on YouTube*. ADL Center for Technology & Society. Retrieved on February 16, 2022, from: <https://www.adl.org/media/15868/download>.
- Chivers Yochim, E., & Silva, V. T. (2013). Everyday expertise, autism, and “good” mothering in the media discourse of Jenny McCarthy. *Communication and Critical/Cultural Studies*, 10(4): 406–426.
- Cobbe, J. (2021). Algorithmic censorship by social platforms: Power and resistance. *Philosophy & Technology*, 34(4): 739–766.
- Coleman, S. (2012). Believing the news: From sinking trust to atrophied efficacy. *European Journal of Communication*, 27(1): 35–45.
- Constine, J. (2018, June 20). Instagram hits 1 billion monthly users, up from 800M in September. *TechCrunch+*. Retrieved on November 5, 2021, from: <https://techcrunch.com/2018/06/20/instagram-1-billion-users/>.
- Cook, T. E. (1998). *Governing with the News: The News Media as a Political Institution*. Chicago: University of Chicago Press.
- Cotter, K. (2019). Playing the visibility game: How digital influencers and algorithms negotiate influence on Instagram. *New Media & Society*, 21(4): 895–913.
- Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. Proceedings of the *10th ACM Conference on Recommender Systems*, pp. 191–198.
- Dahl, R. A. (1957). The concept of power. *Behavioral Science*, 2(3): 201–215.
- Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., Gupta, S., He, Y., Lambert, M., Livingston, B., & Sampath, D. (2010). The YouTube video recommendation system. Proceedings of the *Fourth ACM Conference on Recommender Systems*, pp. 293–296.
- DeVito, M. A. (2017). From editors to algorithms: A values-based approach to understanding story selection in the Facebook news feed. *Digital Journalism*, 5(6): 753–773.
- DeVito, M. A., Gergle, D., & Birnholtz, J. (2017). “Algorithms ruin everything” # RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media. Proceedings of the *2017 CHI Conference on Human Factors in Computing Systems*, pp. 3163–3174.
- Diakopoulos, N. (2019). *Automating the News*. Cambridge: Harvard University Press.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1): 114–126.
- Dimock, M. (2019, January 17). Defining generations. Where Millennials end and Generation Z begins. *Pew Research Center*. Retrieved on February 16, 2022, from: <https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/>.

- Dörr, K. N., & Hollnbuchner, K. (2017). Ethical challenges of algorithmic journalism. *Digital Journalism*, 5(4): 404–419.
- DR Medieforskning (2021). *Medieudviklingen 2020*. [Media development 2020]. Danmarks Radio. Retrieved on February 16, 2022, from: <https://www.dr.dk/om-dr/fakta-om-dr/medieforskning>.
- 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.
- Ecker, U. K., Lewandowsky, S., Fenton, O., & Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory & Cognition*, 42(2): 292–304.
- Eide, M., & Knight, G. (1999). Public/private service: Service journalism and the problems of everyday life. *European Journal of Communication*, 14(4): 525–547.
- Eloise, M. (2019, July 19). What Instagram removing likes may mean for influencers and our self-esteem. *Dazed*. Retrieved on June 7, 2020, from: <https://www.dazeddigital.com/science-tech/article/453171/instagram-removing-likes-could-mean-for-influencers-and-self-esteem>.
- 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 (Eds.). *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 2371–2382.
- 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, K. Inkpen, & W. Woo (Eds.). *CHI'15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 153–162.
- Eslami, M., Vaccaro, K., Lee, M. K., On, A. E. B., Gilbert, E., & Karahalios, K. (2019). User attitudes towards algorithmic opacity and transparency in online reviewing platforms. In: S. Brewster, G. Fitzpatrick, A. Cox & V. Kostakos (Eds.). *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–14.
- European Commission (2022a). *Europe Fit for the Digital age: New Online Rules for Platforms*. Retrieved on November 1, 2022, from: [https://ec.europa.eu/info/strategy/priorities-2019–2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment/europe-fit-digital-age-new-online-rules-platforms\\_en](https://ec.europa.eu/info/strategy/priorities-2019–2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment/europe-fit-digital-age-new-online-rules-platforms_en).
- European Commission (2022b). *Fighting Disinformation*. Retrieved on February 15, 2022, from: [https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/fighting-disinformation\\_en](https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/fighting-disinformation_en).
- Facebook (2020, March 25). *Combating COVID-19 Misinformation*. Retrieved on July 2, 2021, from: <https://about.fb.com/news/tag/misinformation/>.
- Faddoul, M., Chaslot, G., & Farid, H. (2020). A longitudinal analysis of YouTube’s promotion of conspiracy videos. *arXiv preprint arXiv:2003.03318*.
- FBI (2020, September 22). Foreign actors and cybercriminals likely to spread disinformation regarding 2020 election results. Retrieved on February 15, 2022, from: <https://www.ic3.gov/Media/Y2020/PSA200922>.

- Feezell, J. T., Wagner, J. K., & Conroy, M. (2021). Exploring the effects of algorithm-driven news sources on political behavior and polarization. *Computers in Human Behavior, 116*: 1–11.
- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly, 80*(S1): 298–320.
- Fletcher, R., & Nielsen, R. K. (2018a). Automated Serendipity: The effect of using search engines on news repertoire balance and diversity. *Digital Journalism, 6*(8): 976–989.
- Fletcher, R., & Nielsen, R. K. (2018b). Are people incidentally exposed to news on social media? A comparative analysis. *New Media & Society, 20*(7): 2450–2468.
- Fletcher, R., & Nielsen, R. K. (2019). Generalised scepticism: How people navigate news on social media. *Information, Communication & Society, 22*(12): 1751–1769.
- Forster, F. (2019). Instagram Algorithm 2019 – Everything you need to know. *Forster Media*. Retrieved on June 15, 2021, from: <https://forster-media.com/instagram-algorithm-2019-everything-you-need-to-know>.
- Franklin Waddell, T. (2018). A robot wrote this? How perceived machine authorship affects news credibility. *Digital Journalism, 6*(2): 236–255.
- Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems, 14*(3): 330–347.
- Fuentes, J., Hervás, A., & Howlin, P. (2021). ESCAP practice guidance for autism: A summary of evidence-based recommendations for diagnosis and treatment. *European Child & Adolescent Psychiatry, 30*(6): 961–984.
- Galan, L., Osserman, J., Parker, T., & Taylor, M. (2019). *How Young People Consume News and the Implications for Mainstream Media*. Flamingo: Reuters Institute & University of Oxford. Retrieved on February 16, 2022, from: <https://reutersinstitute.politics.ox.ac.uk/our-research/how-young-people-consume-news-and-implications-mainstream-media>.
- 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.
- Gillespie, T. (2014). The relevance of algorithms. In: T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.). *Media Technologies: Essays on Communication, Materiality, and Society*, pp. 167–194. Cambridge: MIT Press Ltd.
- Gillespie, T. (2020). Content moderation, AI, and the question of scale. *Big Data & Society, 7*(2): 1–5.
- Gillespie, T., Auferheide, P., Carmi, E., Gerrard, Y., Gorwa, R., Matamoros-Fernández, A., Roberts, S. T., Sinnreich, A., & West, S. M. (2020). Expanding the debate about content moderation: Scholarly research agendas for the coming policy debates. *Internet Policy Review, 9*(4): 1–30.
- Grossman, P. (2019, June 12). Instagram Pods and groups: How they can help your brand. *Ampjar*. Retrieved on June 7, 2020, from: <https://ampjar.com/blog/instagram-pod-guide/>.
- Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines, 30*(1), 99–120.

- Haim, M., Graefe, A., & Brosius, H. B. (2018). Burst of the filter bubble? Effects of personalization on the diversity of Google News. *Digital Journalism*, 6(3): 330–343.
- Hallin, D. C., & Mancini, P. (2004). *Comparing Media Systems: Three Models of Media and Politics*. Cambridge: Cambridge University Press.
- Hanitzsch, T., Van Dalen, A., & Steindl, N. (2018). Caught in the nexus: A comparative and longitudinal analysis of public trust in the press. *The International Journal of Press/Politics*, 23(1): 3–23.
- Hart, K. (2019, April 2). Our stupid simple guide to the Instagram algorithm. *Jumper Media*. Retrieved on June 7, 2020, from: <https://jumpermedia.co/instagram-algorithm/>.
- Headlam, R. (2019, January 1). What is the best time to post on Instagram to get more likes? *Spindogs*. Retrieved on June 7, 2020, from: <https://www.spindogs.co.uk/blog/2019/01/29/what-is-the-best-time-to-post-on-instagram-to-get-more-likes/>.
- Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, 7(8): 993–1012.
- Helberger, N., & Diakopoulos, N. (2022). The European AI act and how it matters for research into AI in media and journalism. *Digital Journalism*, 1–10.
- Helberger, N., Karppinen, K., & D'acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, 21(2): 191–207.
- Helberger, N., Van Drunen, M., Eskens, S., Bastian, M., & Moeller, J. (2020). A freedom of expression perspective on AI in the media—with a special focus on editorial decision making on social media platforms and in the news media. *European Journal of Law and Technology*, 11(3): 1–28.
- Hensley, L. (2019, May 3). Instagram wants to hide 'likes.' Here's what influencers think about that. *Global News*. Retrieved on February 16, 2022, from: <https://globalnews.ca/news/5226438/instagram-hiding-likes-influencers-react/>.
- Herman, J. (2019). Did the Instagram algorithm change? *Jenn's Trends*. Retrieved on June 7, 2020, from: <https://www.jennstrends.com/did-the-instagram-algorithm-change/>.
- Hofstadter, D. R. (1995). *Fluid Concepts and Creative Analogies: Computer Models of The Fundamental Mechanisms of Thought*. New York: Basic Books.
- Hurd, I. F. (2012). Legitimacy. In *The Princeton Encyclopedia of Self-Determination*. Princeton: Princeton University Press.
- Hussein, E., Juneja, P., & Mitra, T. (2020). Measuring misinformation in video search platforms: An audit study on YouTube. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1), 1–27.
- Juneja, P., & Mitra, T. (2021). Auditing e-commerce platforms for algorithmically curated vaccine misinformation. In *Proceedings of the 2021 Chi Conference on Human Factors in Computing Systems* (pp. 1–27).
- Just, N., & Latzer, M. (2017). Governance by algorithms: Reality construction by algorithmic selection on the Internet. *Media, Culture & Society*, 39(2): 238–258.
- Kaiser, J., & Rauchfleisch, A. (2020). Birds of a feather get recommended together: Algorithmic homophily in YouTube's channel recommendations in the United States and Germany. *Social Media+ Society*, 6(4): 1–15.



- Kaplan, F. (2004). Who is afraid of the Humanoid? Investigating cultural differences in the acceptance of robots. *International Journal of Humanoid Robotics*, 1(3): 465–480.
- Karlsson, M., Clerwall, C., & Nord, L. (2014). You ain't seen nothing yet: Transparency's (lack of) effect on source and message credibility. *Journalism Studies*, 15(5): 668–678.
- Khamis, S., Ang, L., & Welling, R. (2017). Self-branding, 'micro-celebrity' and the rise of Social Media Influencers. *Celebrity Studies*, 8(2): 191–208.
- Kim, D., & Kim, S. (2018). Newspaper journalists' attitudes towards robot journalism. *Telematics and Informatics*, 35(2): 340–357.
- Kircher, M. M. (2019, January 14). Instagrammers are freaking out over a nonexistent algorithm change. *New York Magazine*. Retrieved on June 7, 2020, from: <https://nymag.com/intelligencer/2019/01/instagram-did-not-change-its-algorithm.html>.
- Kitchin, R. & Dodge, M. (2011). *CodeSpace: Software and Everyday Life*. Cambridge, MA: MIT Press.
- Klinger, U., & Svensson, J. (2018). The end of media logics? On algorithms and agency. *New Media & Society*, 20(12): 4653–4670.
- Kruikemeier, S., Boerman, S. C., & Bol, N. (2021). How algorithmic systems changed communication in a digital society. *Media and Communication*, 9(4): 116–119.
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1): 1–16.
- Leerssen, P. (2020). The Soap Box as a Black Box: Regulating transparency in social media recommender systems. *European Journal of Law and Technology*, 11(2): 1–51.
- Lefcourt, H. M. (2014). *Locus of Control: Current Trends in Theory & Research*. New York: Psychology Press.
- Lessig, L. (1999). *Code and Other Laws of Cyberspace*. New York: Basic books.
- Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131.
- Lewis S. C. (2012). The tension between professional control and open participation: Journalism and its boundaries. *Information, Communication & Society*, 15(6): 836–866.
- Lewis, R. (2020). "This is what the news won't show you": YouTube creators and the reactionary politics of micro-celebrity. *Television & New Media*, 21(2): 201–217.
- Lexico (2021). Misinformation. *Lexico*. Retrieved on June 2, 2021, from: <https://www.lexico.com/definition/misinformation>.
- Linden, C. G. (2017). Decades of automation in the newsroom: Why are there still so many jobs in journalism? *Digital Journalism*, 5(2), 123–140.
- Lofthouse, N., Hendren, R., Hurt, E., Arnold, L. E., & Butter, E. (2012). A review of complementary and alternative treatments for autism spectrum disorders. *Autism Research and Treatment*, 2012: 1–21.
- 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.

- Lua, A. (2019). How the Instagram algorithm works in 2019: Everything you need to know. *Buffer*. Retrieved on June 7, 2020, from: <https://buffer.com/library/instagram-feed-algorithm/>
- Maares, P., & Hanusch, F. (2020). Exploring the boundaries of journalism: Instagram micro-bloggers in the twilight zone of lifestyle journalism. *Journalism*, 21(2): 262–278.
- Massanari, A. (2017). # Gamergate and The Fapping: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society*, 19(3): 329–346.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review*, 20(3): 709–734.
- McQuail, D. (2010). *McQuail's Mass Communication Theory* (6th ed.). London: SAGE Publications.
- Mehlsen, C. (2020). *Influencers – De Nye Unge Mediehuse* [Influencers – The new young media houses]. SDU: Mediernes Forsknings- og Innovationscenter. Retrieved on February 16, 2022, from [https://www.sdu.dk/-/media/images/om\\_sdu/centre/journalistik/mfi/influencerrapport.pdf](https://www.sdu.dk/-/media/images/om_sdu/centre/journalistik/mfi/influencerrapport.pdf)
- Mellado, C., & Van Dalen, A. (2017). Challenging the citizen–consumer journalistic dichotomy: A news content analysis of audience approaches in Chile. *Journalism & Mass Communication Quarterly*, 94(1): 213–237.
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 220–229.
- 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.
- Mosseri, A. (2021, June 8). Shedding more light on how Instagram works. *Instagram*. Retrieved on June 11, 2021, from: <https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works>.
- Munger, K., & Phillips, J. (2022). Right-wing YouTube: A supply and demand perspective. *The International Journal of Press/Politics*, 27(1): 186–219.
- Napoli, P. M. (2014). Automated media: An institutional theory perspective on algorithmic media production and consumption. *Communication Theory*, 24(3): 340–360.
- 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.
- Neubaum, G., & Krämer, N. C. (2018). What do we fear? Expected sanctions for expressing minority opinions in offline and online communication. *Communication Research*, 45(2), 139–164.
- Newman, N., Fletcher, R., Kalogeropoulos, A., & Nielsen, R. K. (2019). *Reuters Institute Digital News Report 2019*. Oxford: Reuters Institute for the Study of Journalism.
- Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2): 141–161.

- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2): 175–220.
- Nielsen, R. K. (2016). People want personalized 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. 111–113. Oxford: Reuters Institute for the Study of Journalism.
- O’Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2015). Down the (white) rabbit hole: The extreme right and online recommenders systems. *Social Science Computer Review*, 33(4): 459–478.
- O’Meara, V. (2019). Weapons of the chic: Instagram influencer engagement pods as practices of resistance to Instagram platform labor. *Social Media+ Society*, 5(4): 1–11.
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. London: The Penguin Press.
- Paul, K. (2020, 11 June). ‘Stop treating protesters like Coachella’: Influencers criticized for capitalizing on movement. *The Guardian*. Retrieved on December 8, 2021, from: <https://www.theguardian.com/media/2020/jun/11/influencers-protests-staged-photos-black-lives-matter>.
- Peifer, J. T., & Meisinger, J. (2021). The value of explaining the process: How journalistic transparency and perceptions of news media importance can (sometimes) foster message credibility and engagement intentions. *Journalism & Mass Communication Quarterly*, 98(3): 828–853.
- Powers, E. (2017). My news feed is filtered? Awareness of news personalization among college students. *Digital Journalism*, 5(10): 1315–1335.
- Purtill, J. (2019, July 18). ‘It’s demotivating for me’: Influencers and agencies react to Instagram hiding ‘likes’ count. *ABC*. Retrieved on February 16, 2022, from: <https://www.abc.net.au/triplej/programs/hack/influencers-and-agencies-react-to-instagram-hiding-likes-count/11322832>.
- Puschmann, C. (2019). Beyond the bubble: Assessing the diversity of political search results. *Digital Journalism*, 7(6): 824–843.
- Rader, E., & Gray, R. (2015). Understanding user beliefs about algorithmic curation in the Facebook News Feed. In: B. Begole, J. Kim, K. Inkpen, & W. Woo (Eds.). *CHI’15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 173–182.
- Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A. F., & Meira, W. (2020). Auditing radicalization pathways on Youtube. In: M. Hildebrandt, C. Castillo, E. Celis, S. Ruggieri, L. Taylor, & G. Zanfir-Fortuna (Eds.). *FAT\* ’20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 131–141.
- Rieder, B. (2015) ‘YouTube data tools (version 1.10)’, Software, available at <https://tools.digitalmethods.net/netvizz/youtube>.
- Rieder, B., Matamoros-Fernández, A., & Coromina, Ò. (2018). From ranking algorithms to ‘ranking cultures’ Investigating the modulation of visibility in YouTube search results. *Convergence: The International Journal of Research into New Media Technologies*, 24(1): 50–68.
- Saha, A., & Agarwal, N. (2015). Demonstrating social support from Autism bloggers community on Twitter. In: J. Pei, F. Silvestri & J. Tang (Eds.). *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pp. 1053–1056.

- 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.
- Schouten, A. P., Janssen, L., & Verspaget, M. (2020). Celebrity vs. Influencer endorsements in advertising: The role of identification, credibility, and Product-Endorser fit. *International Journal of Advertising*, 39(2): 258–281.
- Shade, L. R. (2003). Technological determinism. In S. Jones (Ed.). *Encyclopedia of New Media*, pp. 433–434. London: Sage Publications.
- Shariff, A., Bonnefon, J. F., & Rahwan, I. (2017). Psychological roadblocks to the adoption of self-driving vehicles. *Nature Human Behaviour*, 1(10): 694–696.
- Sherman (2019). Is an Instagram Pod the Latest Revelation in Advertising Your Business? Retrieved on 25 November, 2022, from: <https://www.lyfemarketing.com/blog/instagram-pod/>
- Shoemaker, P. J., Vos, T. P., Reese, S. D. (2009). Journalists as gatekeepers. In: K. Wahl-Jorgensen and T. Hanitzsch (Eds). *The handbook of journalism studies*. London: Routledge, pp. 73–87.
- Skovsgaard M., & Bro, P. (2011). Preferences, principle and practice: Journalistic claims for legitimacy. *Journalism Practice*, 5(3): 319–331.
- Smith, A. (2018). Public attitudes toward computer algorithms. Pew Research Center. Retrieved on 30 October 2022, from: <https://www.pewresearch.org/internet/2018/11/16/public-attitudes-toward-computer-algorithms/>.
- Smith, M. R., & Marx, L. (Eds.) (1994). *Does Technology Drive History? The Dilemma of Technological Determinism*. Cambridge: The MIT Press.
- Standberry, S. (2019). Is an Instagram Pod the latest revelation in advertising your business? *Lyfe Marketing*. Retrieved on June 7, 2020, from: <https://www.lyfemarketing.com/blog/instagram-pod/>.
- Steiner, C. (2012). *Automate This. How Algorithms Came to Rule Our World*. London: Penguin.
- Strömbäck, J. (2005). In search of a standard: Four models of democracy and their normative implications for journalism. *Journalism Studies*, 6(3), 331–345.
- Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of communication*, 60(3), 556–576.
- Suchman M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *The Academy of Management Review* 20(3): 571–610.
- Sundar, S. S. (2008). The MAIN Model: A heuristic approach to understanding technology effects on credibility. In: M. J. Metzger and A. J. Flanagin (Eds.). *Digital Media, Youth, and Credibility*, pp. 73–100. Cambridge, MA: The MIT Press.
- Sundar, S. S., & Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication*, 51(1): 52–72.
- Sundar, S. S., Waddell, T. F., & Jung, E. H. (2016). The Hollywood Robot Syndrome: Media effects on older adults' attitudes toward robots and adoption intentions. In: C. Bartneck, Y. Nagai, A. Paiva, & S. Šabanović (Eds.). *HRI '16 11th ACM/IEEE International Conference on Human-Robot Interaction*, pp. 343–350.
- Sunstein, C. R. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton: Princeton University Press.

- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. London: Yale University Press.
- Thurman, N. (2011). Making 'The Daily Me': Technology, economics and habit in the mainstream assimilation of personalized news. *Journalism*, 12(4): 395–415.
- Thurman, N., Lewis, S. C., & Kunert, J. (2019). Algorithms, automation, and news. *Digital Journalism*, 7(8): 980–992.
- 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.
- Tori, A. (2017, May 5). Dear Instagram, we hate the stupid algorithm – Sincerely, every user. *HuffPost*. Retrieved on 16 February, 2022, from: [https://www.huffpost.com/entry/dear-instagram-we-hate-you-the-stupid-algorithm\\_b\\_591a3b57e4b03e1c81b0083c](https://www.huffpost.com/entry/dear-instagram-we-hate-you-the-stupid-algorithm_b_591a3b57e4b03e1c81b0083c).
- Trielli, D., & Diakopoulos, N. (2019). Search as news curator: The role of Google in shaping attention to news information. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–15.
- U.S. Food & Drug administration (2019). Danger: Don't drink Miracle Mineral Solution or similar products. Retrieved on February 15, 2022, from: <https://www.fda.gov/consumers/consumer-updates/danger-dont-drink-miracle-mineral-solution-or-similar-products>.
- Van Aelst, P., Brants, K., Van Praag, P., De Vreese, C., Nuytemans, M., & Van Dalen, A. (2008). The fourth estate as superpower? An empirical study of perceptions of media power in Belgium and The Netherlands. *Journalism Studies*, 9(4): 494–511.
- Van Dalen, A. (2012). The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism Practice*, 6(5–6): 648–658.
- Van Dalen, A. (2019a). Journalism, Trust, and Credibility. In: K. Wahl-Jorgensen & T. Hanitzsch (Eds.). *The Handbook of Journalism Studies*. London: Routledge, pp. 356–371.
- Van Dalen, A. (2019b). Rethinking journalist–politician relations in the age of populism: How outsider politicians delegitimize mainstream journalists. *Journalism*, 22(11): 2711–2728.
- Van Dalen, A. (2021). Når algoritmer styrer nyhedsstrømmen. YouTube- anbefalinger under folketings-valget i 2019. [When algorithms steer the news. YouTube recommendations during the 2019 parliamentary elections]. *Politica*, 53(2): 168–187.
- Velkova, J., & Kaun, A. (2021). Algorithmic resistance: media practices and the politics of repair. *Information, Communication & Society*, 24(4): 523–540.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380): 1146–1151.
- Wallace, J. (2018). Modelling contemporary gatekeeping: The rise of individuals, algorithms and platforms in digital news dissemination. *Digital Journalism*, 6(3), 274–293.
- White, D. M. (1950). The “gate keeper”: A case study in the selection of news. *Journalism Quarterly*, 27(4), 383–390.
- Wu, S., Tandoc Jr. E., & Salmon, C. T. (2019). Journalism reconfigured. *Journalism Studies*, 20(10): 1440–1457.

- York, B. (2018, January 16). This is why I hate Instagram's algorithm. *Thought Catalog*. Retrieved on June 7, 2020, from: <https://thoughtcatalog.com/bailey-york/2018/01/this-is-why-i-hate-instagrams-algorithm/>.
- YouTube (2019). Continuing our work to improve recommendations on YouTube. *YouTube Official Blog*. Retrieved on February 15, 2022, from: <https://blog.youtube/news-and-events/continuing-our-work-to-improve/>.
- 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.
- Zarouali, B., Makhortykh, M., Bastian, M., & Araujo, T. (2021). Overcoming polarization with chatbot news? Investigating the impact of news content containing opposing views on agreement and credibility. *European Journal of Communication*, 36(1): 53–68.
- 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. Journal on Internet Regulation*, 5(1): 1–16.

# Index

- algorithm averse 19
- algorithm literacy 69
- algorithmic power 1–2, 12, 14–15, 49, 61, 63–64
- algorithms in public life 2–3, 15, 18, 22, 27, 64
- anthropomorphism 18
- Apple News 1
- approval of news algorithms 14, 16–17, 20–23
- automatically selected news 19–21, 29
- automatically written news 19
- autonomy 15, 50–51, 53, 57, 61, 64
  
- character 12, 47, 58, 64–65, 67
- confirmation bias 4, 6
- confirms people’s worldview 4, 6
- correcting misinformation 36–42, 63
- credibility 19, 48
  
- deliberation 4
- dependence on the algorithms 54–55, 61
  
- echo chambers 5–6, 8, 20–21, 65
- elevator operators 66
- ethical guidelines 14
- extremist 1, 37–38
  
- Facebook 1, 5–7, 11, 16, 18, 22–23, 33–34, 37, 43, 46, 49, 51
- filter bubble 5–6, 37, 43, 63
- The Filter Bubble: What the Internet is Hiding from You* 5
- fourth estate 13
- Frankenstein complex 20
  
- gatekeeping 1–5, 8, 10–14, 63–67, 69; algorithmic gatekeeping 2–5, 8, 10–14, 63–66, 69; journalistic gatekeeping 2–3, 13–14
- generations 1, 15, 18, 21–23, 28–29, 46, 64–65; Boomer generation 29; Generation Z (Gen Z. 23, 28–30, 46, 64
- Google 1, 5, 7, 18, 50, 52; Google News 7
  
- Hollywood Robot Syndrome 20
  
- Influencers 3–4, 11, 15–16, 46–59, 61–62, 64
- Instagram 1, 4, 11, 15–16, 35, 45–49, 51–62
  
- legitimacy 2, 4, 11–13, 15, 46–47, 51–53, 58–59, 61–65, 67, 69
- locus of control 49, 53
  
- machine heuristic 17, 19, 21, 24
- McCarthy, Jenny 35, 39, 40–42, 44
- mediation 15
- mental model 17–18, 29
- mental shortcuts 17
- Miracle Mineral Solution 34, 36, 39, 41, 43–44
- misinformation 1, 3–4, 15–16, 31–34, 36–45, 63, 66, 68
- moderation 1, 33, 34, 69
  
- Pariser, Eli 5
- personalization 7, 21
- Pods 50, 52–53, 56–57, 61
- polarization 2, 6–7, 66
- public sphere 4, 11–12, 14
  
- qualitative analysis 15, 53

- quantitative content analysis 15
- regulation 13, 34, 48, 66; government
  - regulation 14; regulatory initiatives 65–66
- resistance to algorithms 12, 50, 52, 56, 61
- right-wing content 9–10, 35, 37
- robot 18–20, 24; robotics 18, 20
- scientific consensus 35–36, 63, 68
- selective exposure 7
- structuring power 10–11, 45, 49–50, 63–64
- Sunstein, Cass R. 6, 51
- supply-and-demand framework 8–10
- survey 8, 14–17, 20, 22–24, 29, 67;
  - survey-embedded experiments 15, 17, 22
- technological determinism 4–6, 8
- thematic analysis 47, 52
- theory of machine 69
- TikTok 1, 46
- transparency 18, 53, 59, 62, 65–66
- trust 2, 11–22, 24–25, 28–29, 33, 36, 47, 53, 59, 62–67, 68, 69
- Twitter 7–8, 16, 18, 32, 50
- values 11, 28, 34, 58–59, 61–62;
  - democratic values 61
- YouTube 1, 4, 6–7, 9, 10, 15–16, 30, 33–44, 46, 49, 63, 68; YouTube
  - Data Tools 39; YouTube recommendations 33, 36–38, 40, 42–43; YouTube search 15, 37, 39–40, 43; YouTube’s recommendation algorithm 4, 15, 36–38, 40, 43–44, 63





# Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>