Citizens and Intelligent Machines: Algorithms in Deliberative Democracy

Nardine Alnemr

Thesis submitted in fulfilment for the degree of Doctor of Philosophy at the University of Canberra, Australian Capital Territory, Australia

University of Canberra
October 2021
Abstract

To what extent can algorithms promote or hinder the practice of deliberative democracy? Extant accounts of algorithms in deliberative democracy are cautious of their consequences on the democratic quality of the public sphere but see a potential in algorithms to design the right conditions for citizen deliberation. Between these two accounts, I argue in this thesis that algorithms in deliberative democracy should be examined in the context of the “algorithmic society”—a society shaped by, relying on, and critical of algorithms, automation, and artificial intelligence (AI). A broader understanding of this condition can highlight the roles of algorithms relevant to deliberative democracy, not only in communication but also in decision making. In four accounts of algorithms in deliberative democracy, I apply normative theory to scrutinise the roles of algorithms across three sites of deliberation: the public sphere, minipublics, and institutions. Overall, this thesis presents a contribution to the literature by highlighting how a deliberative democratic account of algorithms ought to be critical, multifaceted, and citizen-centric.
Acknowledgements

My first acknowledgment goes to my supervisors and colleagues. I would like to acknowledge the incredible effort, support, and mentorship of my supervisory panel: Simon Niemeyer, Nicole Curato and John Dryzek. Thank you for accepting my application to do a PhD at the Centre and contribute to the Metastudy of Democratic Deliberation project, which marked the beginning of my journey at the Centre. Also, thank you for pushing me to think deeper about where this thesis goes next when you asked very challenging questions at the thesis pre-submission seminar. As for the final two months of this PhD, I cannot thank you enough for your patience and kindness that grounded me, especially Nicole. I would like to acknowledge Selen Ercan for her support as an examiner of my candidature confirmation seminar, and continued support and generosity whether by including me in projects or the stimulating conversations about our work and the state of the world. To my colleagues and friends at the University of Canberra who were most supportive and inspiring: Emerson Sanchez, Kei Nishiyama, Wendy Conway-Lamb, Pierrick Chalay, Jane Alver, Mohammad Abulhawas, Madeleine Egan, Robin Ladwig, Ario Wicakseno, and Jane Phuong. Thank you for your comments, suggestions, and critiques, but most importantly, thank you for the coffee time and the companionship you have graciously shared over the course of this PhD. I am lucky that my friend and I started our PhDs in the same year. Amany Selim, thank you for listening to all the vents and grunts of frustration, and thank you for your compassion that kept me going through suffering from imposter syndrome. Thank you for keeping me up and working through the Canberra winter nights when you stayed through the Norwegian summer. This thesis would have not materialised if it were not for the collective contributions of many colleagues in the field. I am indebted to many journal reviewers, participants in conferences and colleagues who contributed to the development of my arguments and the presentation of these argument. Last but not least, I would like to thank Rossine Fallorina for copyediting this thesis before the submission.

I would like to acknowledge my family for their unwavering support. Amaal Khalil, my mother and friend, thank you for allowing me to be a part of your professional world of journalism and TV years ago. I would have not developed an interest in studying many of the things I write about if it were not for the times I tagged along to listen, observe and understand. To my siblings: Sarah, Yousef and Khalid, thank you for being there and reminding me that it is okay to fail at “adulting” because your inner child would not obey. Connecting with the inner child in us through jokes, play and random hobbies, as you allowed me to see, is an advantage to stay curious and engaged.

Finally, I would like to acknowledge the loss of family and family friends I experienced during my PhD journey that changed me as a person and an academic. The first loss was of a family friend, the late Jamal Khashoggi. It was hard to mourn and grief when the world was watching. Jamal’s death reminded me of the importance of studying democracy at home, even when by the day it becomes risker. I realised I need to uplift and contextualise the richness of thought about and aspirations for democracy at home no matter how farfetched it may seem. The second loss was my gedo, Rateeb Khalil. This was hard because I could not travel to see him before he passed or mourn with my family. It reminded me of the precarity of being a non-citizen and that when writing about “citizens”, I have to think about these experiences rather than imagine ideals that do not speak to our different realities.
Chapter 3 | Emancipation Cannot be Programmed: Blind Spots of Algorithmic Facilitation in Online Deliberation ................................................................. 61
3.1 Introduction .................................................................................................................. 62
3.2 Emancipation, Minipublics Design, and Facilitation .............................................. 64
3.2.1 Facilitation: Key to Deliberative Design ................................................................. 65
3.2.2 Minipublics on Demand ......................................................................................... 66
3.3 Programming Deliberation: Algorithms in Online Deliberative Design ................. 68
3.3.1 Algorithms: Supporting and Substituting Human Facilitators ............................... 69
3.3.2 Artificial Facilitators in Online Deliberation ......................................................... 70
3.4 Blind Spots of Programmed Deliberation: Theoretical and Practical Implications ........................................................................................................ 73
3.4.1 Theoretical Implications of Algorithmic Facilitation ............................................ 73
3.4.2 Practical Implications of Algorithmic Facilitation ................................................. 76
3.5 Recommendations for Democratic Online Deliberation ........................................... 79
3.6 Conclusion .................................................................................................................. 84
3.7 References ................................................................................................................ 85

Chapter 4 | Democratic Self-government and the Algocratic Shortcut:
The Democratic Harms in Algorithmic Governance of Society .................. 91
4.1 Introduction .................................................................................................................. 92
4.2 Shortcuts to Democracy and Algorithms ................................................................. 94
4.2.1 The Algocratic Shortcut ........................................................................................ 96
4.2.2 Democratic Harms in the Algocratic Shortcut ....................................................... 99
4.3 Democratising Algorithms in Europe: Political Justification through Epistocratic and Lotocratic Shortcuts .......................................................... 104
4.3.1 Inequality in Influencing Decisions .................................................................... 104
4.3.2 Inequality in Full and Fair Consideration of Reasons ......................................... 106
4.3.3 Inequality in Free and Effective Voice ................................................................. 109
4.4 Remedying the Algocratic Shortcut ......................................................................... 111
4.5 Conclusion ................................................................................................................ 114
4.6 References ................................................................................................................ 115

Chapter 5 | Can Robots be Partners in Democracy? Deliberative Democratic Values for Politics with Robots ................................................................. 121
5.1 Introduction ................................................................................................................ 122
5.2 What is the role of robots in democratic communication? ....................................... 124
5.3 How can robots be partners in democracy? Deliberative Democratic Values for Robots ...................................................................................................... 127
5.4 Applying Deliberative Capacity to Assess Politics with Sophia ................................ 131
5.4.1 Inclusivity: The Robot “Rebel” from Saudi Arabia .................................................. 131
5.4.2 (In)Authenticity: Reasons of the Robot, Business, and International Community ........................................................................................................ 134
5.4.3 Consequentiality: “The Future is Robot” .............................................................. 135
5.5 Conclusion ................................................................................................................ 136
5.6 References ................................................................................................................ 137
List of Tables

Table 1  List of Publications and Review Status................................................................. 27
Table 2  Democratic Self-government and Shortcuts .......................................................... 96
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF</td>
<td>artificial facilitator</td>
</tr>
<tr>
<td>AI</td>
<td>artificial intelligence</td>
</tr>
<tr>
<td>bot</td>
<td>short for “robot,” typically a reference to robots on the internet</td>
</tr>
<tr>
<td>COE</td>
<td>Council of Europe</td>
</tr>
<tr>
<td>DiFA</td>
<td>artificial discussion facilitation agent</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>GDPR</td>
<td>General Data Protection Regulation</td>
</tr>
<tr>
<td>ICO</td>
<td>U.K. Information Commissioner’s Office</td>
</tr>
<tr>
<td>Ofqual</td>
<td>Office of Qualifications and Examinations Regulation</td>
</tr>
<tr>
<td>RSA</td>
<td>Royal Society for Arts, Manufactures, and Commerce</td>
</tr>
<tr>
<td>UN</td>
<td>United Nations</td>
</tr>
<tr>
<td>UNDP</td>
<td>United Nations Development Programme</td>
</tr>
</tbody>
</table>
CHAPTER 1

Algorithms in Deliberative Democracy: An Introduction

Hey guys, hope you are doing well. Just watching my algorithms get crushed. I guess I did something to piss off the Instagram god, so hopefully you are seeing this stuff anyway. We’ll do what we can. Talk to you soon.

Donald Trump Jr., 19 October 2020

1.1 Introduction

A few days before the 2021 U.S. presidential elections, Donald Trump Jr. posted an Instagram story. He lamented how his algorithms were “get[ting] crushed,” implying that his content became invisible to his followers. He referred to an “Instagram god,” which, one could argue, was a shorthand for the ever-present algorithms that “operate in unseen and unknown ways” (Beer, 2009, p. 995). This Instagram story may seem mundane, but, upon closer inspection, it exemplifies a pervasive view that algorithms have the power to give visibility to some political actors, as in the case of the Trumps in 2016, and invisibility to others, as in the case of the Trumps in 2021 (Coster & Dave, 2021).

The perception that algorithms are gods suggests that algorithms serve as an authoritarian force in democratic societies (Ames, 2018). Algorithms control what and who we see, shape our knowledge and taste, recommend individuals and groups we connect with, affect our communication and political action, and even make decisions that have ramifications on welfare, health, and justice.

We now live in what Jack Balkin (2017, p. 1151) describes as an “algorithmic society [which] features large, multinational social media platforms that sit between traditional nation-states and ordinary individuals, and the use of algorithms and artificial intelligence agents to govern populations.” An algorithmic society governs individuals by collecting vast amounts of data on our everyday lives that we generate. As we rely on “more data than ever before,” new forms of surveillance, control, and manipulation are introduced (Kahn & Kellner, 2007; Kemper & Kolkman, 2019). The age of algorithms, as Cass Sunstein puts it, is one “where the algorithm knows a lot” (Sunstein, 2017, p. 3). The all-knowing algorithm has the power to control knowledge in ways “that transform human subjectivity” as they curate what we read, analyse, and think about (Hayles in Amoore & Piotukh, 2016). It undermines our democratic
Chapter 1

culture by undercutting the opportunity for citizens to form a collective will through public deliberation and dislodge the role of human judgment, discretion, and reflection. Instead, an algorithmic society finds power to be concentrated in states and private companies who make decisions on our behalf, oftentimes without our consent (Couldry, 2017). The pathologies of an algorithmic society are widely documented, from giving rise to a post-truth society to normalising inflammatory and hateful speech and exacerbating discrimination among the disadvantaged.

As the field of democratic theory and practice learns more about the harms algorithms cause in contemporary societies, more work needs to be done in understanding how algorithms can also be shaped to promote democratic procedures and outcomes. In this thesis, I begin with the premise that we are indeed living in an algorithmic society or, more precisely, an “algorithmically infused society” where human processes and behaviours are deeply entangled with algorithms (Wagner et al., 2021). Algorithms are here to stay, for better or for worse. If governing the algorithmic society does not “centre citizen agency and democratic accountability,” it signals the inevitable decline of democracy (Gurumurthy & Bharthur, 2018). My attempt in this thesis is to provide a nuanced account of the different ways in which we can rebuild democracy, particularly deliberative democracy, in an algorithmic society.

1.2 Research Question

My main research question is: “To what extent can algorithms promote or hinder the practice of deliberative democracy?”

I recognise that algorithms have an impact on the varieties of democracy, such as electoral, liberal, participatory, and egalitarian democracy (Mechkova & Sigman, 2016). In this thesis, I choose to focus on deliberative democracy for two reasons.

First, deliberative democracy as a normative political theory places meaningful communication at the centre of democratic life (Bächtiger et al., 2018). Given that algorithms have fundamentally altered the communicative landscape in contemporary democracies, it is only fitting that we situate our examination of algorithms in relation to a normative theory that places a premium on creating conditions that make democratic communication inclusive, other regarding, respectful, and reflective.
The scholarship on deliberative democracy has gained inroads in characterising the role and impact of algorithms in deliberation. Some scholars who focus on mass deliberation identify the epistemic harms algorithms and bots have on our capacity for truth tracking and reflection (Chambers, 2021; Tanasoca, 2019). Meanwhile, scholars focusing on micropolitical deliberation or minipublics began examining how algorithms can be used to enhance the design of these forums. Algorithmic facilitation can, for example, scale up online deliberation (Fishkin, 2018; Stromer-Galley et al., 2012; Wyss & Beste, 2017), while algorithmic sortition can facilitate the recruitment of a representative sample of countries, or even the world for global assemblies (see Flanigan et al., 2021; Warren, 2021). Some scholars who examine the role of algorithms at the institutional level also appreciate the contribution of algorithms in facilitating unbiased decision making (Kahneman, 2011; Kahneman et al., 2016; Sunstein, 2019).

These developments, among others, demonstrate that the literature, thus far, has been both critical and open to the ways in which algorithms can promote deliberative democracy. This thesis seeks to contribute to these ongoing discussions by revisiting the normative principles of deliberative democracy to identify how algorithms can promote deliberative democracy and characterise the conditions and processes that prevent them from doing so.

Second, I focus on deliberative democracy because more than just a normative theory, deliberative democracy is also a “political project” (Curato et al., 2019). Deliberative practice entails creating possibilities for deliberative norms to flourish amidst the imperfections of our algorithmically infused societies. Designing and implementing deliberative minipublics is one example of how deliberative democrats craft alternative spaces where citizens can combat disinformation on social media by engaging with credible experts, put forward reasons, and listen across difference, instead of remaining within their filter bubbles. Other advocates of deliberative democracy identify how high-quality deliberation can be promoted in areas of political life marked by power asymmetries such as parliaments (Steiner et al., 2005), schools (Nishiyama, 2019), and the workplace (Felicetti, 2018).

I consider this thesis as an extension of these efforts by identifying the practices that can make algorithms deliberative, despite the constraints and power asymmetries in our digital public sphere. As Simone Chambers and John Gastil (2021) suggest, deliberative theory has the power to situate digital media in a wider theoretical context. In doing so, we not only
understand the worst features of the internet, but also “show the way forward to better design for digital public engagement” (Chambers & Gastil, 2021, p. 5). What I aim to put forward in this thesis is neither a utopian algorithmic society nor an ideal deliberative democracy, but an empirically grounded normative account of how algorithms can promote deliberative democracy.

1.3 Structure of the Chapter

This introductory chapter is structured to unpack the different components of the research question: “To what extent can algorithms promote or hinder the practice of deliberative democracy?”

I begin with a definition of algorithms (Section 1.4) which emphasises a sociotechnical understanding of algorithms, rather than a technical definition. I argue that this definition of algorithms is better suited to the discussion of the roles of algorithms in democratic life. Second, I define what I mean by deliberative democracy and identify three sites where deliberative practices take place: in the public sphere, in designed forums or minipublics, and in institutions (Section 1.5). The thesis’s substantive chapters are organised around these sites of deliberation along with how algorithms can promote and undermine deliberative democracy in these sites. In the third section (Section 1.6), I define what I mean by promoting deliberative democracy using John Dryzek’s (2009) work on deliberative capacity. Put simply, the central argument of this thesis is that algorithms can promote deliberative democracy if it creates a polity’s capacity for inclusive, authentic, and consequential discussion. I build this argument by bringing normative theory to bear on illustrative examples, which in turn can be used to sharpen theoretical arguments (Section 1.7). The final part of this chapter provides an overview of each chapter (Section 1.8) and concludes by identifying the limitations and boundaries of the thesis (Section 1.9).

1.4 Defining Algorithms

One approach in defining algorithms is to understand its technical aspects as shaped by social and political influences. Technically, algorithms are a series of steps, instructions, or calculations such as “if-then” (Bucher, 2018; Goffey, 2008; Mackenzie, 2006; Simanowski, 2016). Algorithms are not neutral. While these series of steps may be perceived as
mathematical and calculative, a sociotechnical definition recognises that algorithms are not neutral. These steps, instructions, and calculations are based on the programmers’ understanding of the world. Hence, algorithms are their means to impose this understanding in ordering the world.

Defining and perceiving algorithms as neutral mathematics is what makes “understanding algorithmic oppression” challenging, as Safiya Noble (2018) argues. Therefore, it is important to situate algorithms in a sociotechnical definition, rather than limiting it to technical dimensions. Accordingly, here I define algorithms in terms of instructions and calculations that are products of sociopolitical processes. Understanding these dimensions requires investigating who makes algorithms and interpretations of what algorithms are designed to do.

From this broad sociotechnical definition springs different conceptualisations of the roles of algorithms, often connoting some negative consequences. Examples of these conceptualisations include the algorithmic society\(^1\) (Balkin, 2017; Peeters & Schuilenburg, 2021; Valcke, 2019), algorithmic power (Bucher, 2012, 2018; Danaher et al., 2017; Lundahl, 2020; Neyland & Möllers, 2017), algorithmic fetishism\(^2\) (Crawford, 2016; Monahan, 2018), and algorithmic activism (Maly, 2019). Each of these conceptualisations will be explained further below.

Algorithms are a new factor of power in an algorithmic society that produces a lot of data about itself and is driven to collect even more data to organise it. Balkin (2017, p. 1219) defines the algorithmic society as a “society organized around social and economic decision making by algorithms, robots, and AI agents, who not only make the decisions but also, in some cases, carry them out.” In their definition of an algorithmic society, Marc Schuilenburg and Rik Peeters (2021, p. 195) similarly highlight how “algorithms” are shaped by discourses in society. They define the algorithmic society as:

a set of practices and discourses, implicating hybrid connections between governmental and private parties, that is underpinned by a repertoire of relatively new data-driven technologies, which adds new layers to the

---

\(^1\) Other similar conceptualisations include the algorithmic turn (Gurumurthy & Bharthur, 2018) and algorithmic life (Amoore & Piotukh, 2016).

\(^2\) There is also “algorithmic obligation” (Collins et al., 2020) and “algorithmic accountability” (Kemper & Kolkman, 2019), concepts which are not relevant in the definition but will be discussed later in the literature review.
governance of society through own modes of knowledge, and particular ways of forming new subjects. (Schuilenburg & Peeters, 2021, p. 195)

Any characterisation of an algorithmic society focuses on the role of algorithms in decision making to create and automate a statistical model of citizens in order to determine their access to goods and services. It involves “the application of advanced statistical methods and techniques by private and public organisations to automate decision making on individual people’s access to rights and services and to pick out patterns and correlations out of enormous data sets in order to make predictions about their behaviour” (Peeters & Schuilenburg, 2021, pp. 3–4). This, as Shoshana Zuboff (2019) argues, is one of the manifestations of surveillance capitalism, the political economy of an algorithmic society, which is “antidemocratic” for predetermining individual futures.3

Second, algorithmic power in an algorithmic society means that algorithms do not have power “over” society in as much as it draws its power from “within” (Lash, 2007). This is a feature of posthegemonic power relations, the context in which media and algorithms operate. On the relationship between algorithms, power, and ubiquity, Scott Lash highlights:

Power through the algorithm is increasingly important for media companies in digital rights management. A society of ubiquitous media means a society in which power is increasingly in the algorithm… [Moreover, t]he post-hegemonic order is not just an era of ubiquitous computing and ubiquitous media. It also bequeaths to us ubiquitous politics. (Lash 2007, p. 71, 75)

In explaining ubiquitous publics, Lash warns us about the consequences of how algorithms and its politics of avoiding democratic accountability become integral in society.

---

3 In the first page of her book, Zuboff (2019) defines “surveillance capitalism” as a noun of 8 meanings. Starting from the sixth meaning, Zuboff (2019) draws the connection between the role of algorithms in prediction and the undemocratic qualities of surveillance capitalism: “1. A new economic order that claims that human experience as free raw material for hidden commercial practices of extraction, prediction, and sales. … 6. The origin of a new instrumentarian power that asserts dominance over society and presents startling challenges to market democracy. … 8. An expropriation of critical human rights that is best understood as a coup from above: an overthrow of people’s sovereignty.” Zuboff (2019, p. 39) locates the role of algorithms in the wider political economy of surveillance capitalism which “employs many technologies, but it cannot be equated with any technology.” For algorithms, as well as machine learning and other technologies, surveillance capitalism “produces and relies on algorithms but it is not the same as algorithms” (Zuboff, 2019, p. 40). In this thesis, I am not concerned with establishing the different relations between algorithms and the political economy of surveillance capitalism, but it suffices to highlight at this point that the structures allowing for the dependence on algorithms shape society and its demands on democracy. This is also a point which Zuboff (2019) beckons in discussing surveillance capitalism. One limitation in Zuboff’s conceptualisation is the focus on the antidemocratic qualities of surveillance capitalism in Western liberal democracies alone. In this thesis, I argue against this (although not explicitly) by bringing in examples of democratic participation, and limitations to democratic participation, from beyond Western liberal democracies (e.g., AI robot politician in Japan, using bots to circumvent misinformation in Taiwan, and robot participation in public deliberation in Saudi Arabia).
Ubiquitous politics in this sense refers to how algorithms are black boxes situated in black-boxed governance. Although these black boxes are inaccessible and inexplicable (Pasquale, 2015), they are constitutive of, and constituted in, politics. Consequently, algorithmic power means the normalisation of algorithms’ roles in governing society, which “reshape power relations between state, private companies, and citizens” (Peeters & Schuilenburg, 2021, p. 7). Therefore, the roles of algorithms mark a strengthening of undemocratic politics that shifts away from commitments to explanation, justification, and public scrutiny of algorithmic power as algorithms cannot explain their decisions.

The problem with algorithmic power at this point is that ubiquity is not only at the level of technically explaining how algorithms arrive at their decisions. It is also in their proprietary status, which shields them from democratic accountability (De Gregorio, 2020; Noble, 2018). Sheltered from justifications to citizens, “[a]lgorithms produce outcomes that do not ‘argue’. They do not present an argument or a reasoning, they do not reveal sources or assumptions, but instead merely construct specific forms of knowledge” (Peeters & Schuilenburg, 2021, pp. 4–5). Therefore, for citizens to have democratic control over algorithms that affect them, the values which algorithms automate should be accessible to the public (De Gregorio, 2020).

The focus on algorithms as black box technologies paved a way for “algorithmic fetishism.” This fetishism informs a research agenda interested in opening the black box and in amending what is inside it rather than questioning its underlying politics (Crawford, 2016; Monahan, 2018; Pasquale, 2015). Torin Monahan (2018, p. 2) explains the problems with algorithmic fetishization, which guard “human prejudices and politics” from scrutiny by shifting the focus on algorithms alone:

While it is critical to open up the black box of algorithmic production, that will always be an insufficient response to the forms of discrimination engendered by them because algorithms cannot be separated from the context of their production and use. They do not act independent of social context, despite alarmist, deterministic narratives about their automated capacities. Racialized algorithmic violence, for instance, cannot be eradicated by tracing down the offending code and splicing it with an uncorrupted variant, as if the domain of computer code was or ever could be a pure dimension free from human prejudices and politics, as if “purity” were not an entirely bankrupt concept in the contemporary conjuncture. Clearly, the violence and prejudice of algorithms is, and always was, an extension of those qualities in societies. So, fetishizing algorithms, even from a critical position, risks sideling the
harder empirical, theoretical, and political work of tracing those links and creating a space for the emergence of more just alternatives. (Monahan, 2018, p. 2)

In this quote, Monahan (2018) explains that there is nothing unique about opening the black box of algorithms when these algorithms are responsible for automating decisions about risk and violence—domains that are highly racialized. Instead of opening the black box, Monahan argues that algorithm research ought to be focused on understanding how algorithms operate within existing structures of discrimination and injustice in order to develop “more just alternatives.”

Aside from fetishization, another problem in researching algorithms is focusing on algorithmic harms. In other words, adding “algorithmic” as an adjective preceding a noun is a way of signalling democratic harms (Velkova & Kaun, 2021). For example, “algorithmic activism” was coined to conceptualise how right-wing populist movements (and micromovements) across Finland, the UK, and Australia work their message to appeal to algorithmic logics of virality and visibility (Maly, 2019; Watson & Barnes, 2021).

Deliberative democrats are interested in the conditions that enable or inhibit democratic communication and decision making. This sociotechnical definition of algorithms, hence, sheds light on the wider context of the roles of algorithms in society and the power of tech giants and political actors using algorithms to translate their interpretations of how information, communication, and decisions should be organised and made. What remains a key challenge to deliberative democracy, based on this understanding of algorithms, is the unjustified coercive power of algorithms. As ubiquitous as algorithms are, their roles are mostly supported by undemocratic arguments that algorithms cannot be justified or critiqued to avoid public accountability and scrutiny veiled by technical or proprietary explanations. With this expanded definition of algorithms in mind, algorithms in this thesis are understood as sociopolitical discourses that manifest in communication and decision-making technologies.

1.5 Defining Deliberative Democracy

Deliberative democracy places “meaningful communication at the heart of democracy” (Bächtiger et al., 2018, p. 3). It considers “public deliberation of free and equal citizens [as] the core of legitimate political decision-making and self-government” (Bohman, 1998, p. 401).
As a political ideal, deliberative democracy envisions a society where collective decisions are made based on inclusive and mutually respectful deliberation. Deliberative democracy is often contrasted to aggregative democracy where the legitimacy of political decisions is based on headcounts (Bächtiger et al., 2018, p. 3), although there are certainly overlaps between the ideals of these two approaches to democracy, such as the importance of making informed decisions without coercion.

The practice of deliberation makes deliberative democracy a distinctive political ideal. Deliberation is a social process that emphasises persuasion “instead of coercion, manipulation, or deception” and “openness to change one’s views, judgments, and preferences in light of the better argument” (Dryzek, 2002, p. 1). Deliberation seeks to arrive at a collective “course of action” informed by a “discussion [which] reflects all social experience,” and not just a register of subjective preferences or off-the-cuff opinions (Young, 2002, p. 30). Viewed in this way, deliberative democracy is in pursuit of “authentic democracy” where competent citizens exercise substantive rather than symbolic democratic control over collective outcomes, based on the development of reflective preferences.

The normative ideals of deliberative democracy were articulated in a time that predates the prominence of the algorithmic society. While many deliberative scholars have turned their attention to the potential of online communication in promoting the norms of inclusiveness and reason giving (Black, 2011; Dahlberg, 2001; Esau et al., 2017; Papacharissi, 2002), these accounts do not explicitly conceptualise deliberation as a process unfolding in an “algorithmically infused society” where our knowledge, choices, and worldviews are deeply entangled with algorithms (see Section 1.1). This poses a challenge for deliberative theory. In the same way that Simone Chambers (2003) challenged the next generation of deliberative theorists to more vigorously pursue questions of material distribution and poverty, I argue that deliberative theorists should also vigorously pursue questions of distribution of voice and visibility in an algorithmic society. As Chambers puts it, deliberative theory’s notion of a “well-ordered public sphere” rests on an “ideal notion of deliberation” where “participants are on equal footing” (Chambers, 2003, p. 322). However, what does it mean for participants to be on equal footing not only in a highly unequal society but also in an algorithmic society? What does it mean for the process of deliberation to hold power into account when decisions are automated, instead of collectively made? What does it mean to give other-regarding reasons when algorithms “shortcut” public deliberation?
Chapter 1

To answer these questions systematically, I propose to examine deliberative democracy in an algorithmic society based on three sites where deliberation takes place: (1) the public sphere, (2) structured forums or minipublics, and (3) institutions. As I will discuss in Section 1.9, each of the chapters of this thesis will speak to one of these sites of deliberation and characterise how algorithms can promote (or obstruct) deliberative democracy. For now, however, I will use these three sites of deliberation to structure my review of literature on how deliberative democracy scholars have so far characterised the consequences of algorithms in these sites.

1.5.1 Public Sphere

The public sphere refers to “a domain of our social life in which such a thing as public opinion can be formed” (Habermas, 2018, p. 667). It is an important site of deliberation because it is the space for citizens to exercise and contribute to the principle of “publicity.” For Jürgen Habermas (2018), publicity is about seeking justifications and countering secrecy of government power. Although publicity can take different forms and for different ends, it essentially creates a connection between different actors (Calhoun, 2017). In Habermas’ (2018) definition of the public sphere as a “a sphere mediating between state and society,” we ought to turn to the roles of algorithms in regulating and shaping this sphere, which is now also about seeking justifications from more than states. Most of the discussions within the literature of deliberative democracy emphasise the harms algorithms cause to deliberative democracy. Among these harms include the following:

Undermining Deliberation’s Truth-tracking Function. Chambers (2021) highlights that “big data, micro-targeting, bots, and proprietary algorithms” are one part of the problem of viral fake news. In exploring how algorithms have a role in promoting fake news, Chambers argues that we also ought to understand why political actors introduce fake news. While algorithms rank fake news as relevant and important, which is a problem in itself, it is also important to investigate the functions and ends of political manipulation. In another respect, Spencer McKay and Chris Tenove (2020) highlight that algorithmic filtering and sorting of content online “challenge democratic oversight.” As the logics for filtering and sorting are proprietary, it is difficult to discern and scrutinise these logics. This speaks to the problem of ubiquity and the black box governance discussed earlier. Algorithmic organisation of flows of information and communication is sheltered from democratic control as the actors developing
and using these algorithms (e.g., through microtargeting) want to avoid justifying their logic in ordering content. The logic is often profit generation that makes engagement with fake news and misinformation the reason for the visibility of the content. Yet, when it concerns the democratic ramifications of the visibility of this content, the argument is that the workings of the algorithms are inexplicable in order to avoid accountability.

Creating Polarisation Enclaves. For Sunstein (2017) algorithms that filter and sort content based on personal preferences and interests undermine the inclusive quality of the public sphere. Exposure to personalised content and others who share the views ultimately leads to the rise of enclaves. This harms deliberation as citizens would not be able to communicate across difference, making it more challenging to come to the collective understanding of the common good. Strandberg and colleagues (2019) conducted a comparative study to examine polarisation in online and face-to-face enclaves. The study also compared polarisation in each enclave in two treatments: unstructured and structured for deliberation. The findings highlight that polarisation of positions was a consequence of unstructured discussions online and face-to-face. Similarly, the enclaves that observed deliberative norms experienced a decline in polarised views (Strandberg et al., 2019). Therefore, the phenomenon of online polarisation enclaves is attributed to how algorithms and online spaces are not created to observe deliberative norms of exposure to difference, promoting reflection, and weighing reasons for the common good rather than particular interests.

Compromising Citizens’ Capacity for Internal Deliberation. For Ana Tanasoca (2019), internal deliberation refers to our capacity for autonomous evaluation and weighing of claims. The process results in us updating our range of considerations and reasons. The problem with a public sphere inclusive of bots, nonetheless, is that they can be programmed to replicate and repeat messages, inflating the worth of a message and undermining epistemic diversity. Furthermore, Alfred Moore and colleagues (2021, p. 47) suggest that the bots can negatively affect the “levels of argumentation and justification, which … are core elements of deliberative quality.”

4 Their assertion does not expand on the dynamics of how bots affect the level of argumentation and justification as their examination is more concerned about the different levels of anonymity and its impact on argumentation.
replicate messages online give a false sense of the range and weight of reasons or arguments in the public sphere. The more citizens see a certain message replicated and shared widely, the less likely citizens are to critically examine this message and instead perceive it as a legitimate representation of accepted and endorsed arguments.

**Normalizing Manipulation.** The oft-cited examples of bots in public deliberation are disruptive and manipulative, such as foreign interference via bots to compromise election integrity (Rheault & Musulan, 2021). The bots engage in what Samuel Woolley and Philip Howard (2016) call “computational propaganda” where political actors use bots to disseminate dis/misinformation for political gain. Aside from propaganda, Woolley (2016) maps the political ends for using bots across democracies and nondemocracies. In democracies such as Australia, the US, and the UK, bots are used to fake popularity. As for less democratic and authoritarian states, bots are used to demobilise and repress participation (Woolley, 2016). Across these examples of using bots, the end result is undermining the epistemic quality of public deliberation as bots inject messages that are deceptive and manipulative. Although propaganda is a democratic problem that precedes the age of communication with bots, the primary concern is the speed and scale in which these messages are created and spread. These manipulative messages are also difficult to trace as they can be easily deleted. Finally, the comfort both democracies and nondemocracies have in using bots for political manipulation is concerning.

Based on the literature reviewed herein, most literature on deliberative democracy focusing on the public sphere demonstrates a critical view of algorithms’ role in deliberative democracy.

### 1.5.2 Structured Forums or Minipublics

Deliberative minipublics refer to structured deliberative forums composed of a randomly selected group of ordinary citizens that represent a wider population brought together to learn, deliberate, reflect, and put forward recommendations (or a collective statement) about one or more issues (Curato et al., 2021). Compared to scholarship on algorithms in the public sphere, scholarship on deliberative minipublics tend to have a more favourable view of the potential of algorithms in promoting deliberative democracy. There are three ways in which this favourable view is articulated in the literature.
Scaling Up Deliberation. The problem of scale has always been an issue for scholars and practitioners of deliberation (Fishkin, 2018; Iyengar et al., 2003). Hosting large-scale deliberations is time and labour intensive as it requires setting up conditions conducive for inclusive discussion. Algorithms can solve some of these issues. For example, there is a growing literature on artificial intelligence (AI) facilitation where nonhumans (an algorithmic system or a bot) replace humans to enforce deliberative virtues, such as inclusion and reciprocal reason giving. This allows multiple deliberations to run in parallel and to reduce the time and resources spent to train and employ human facilitators (Fishkin et al., 2019; Stromer-Galley et al., 2012; Wyss & Beste, 2017). The use of artificial facilitators, however, are not foolproof. In Chapter 3, I argue that AI bots are unable to capture contextual, cultural, and communicative cues that are critical in facilitating democratic deliberation.

Algorithmic Sortition. As deliberative minipublics gain increasing popularity around the world, it is not surprising that several practitioners have sought to develop “fair algorithms for citizens assemblies” (Flanigan et al., 2021, p. 548). In an article published in Nature, Bailey Flanigan and colleagues (2021) explained the series of steps their algorithm used to maximise a measure of fairness. Without going into detail, “fair algorithms” seek to put together a representative panel by “selecting individuals with probabilities as close to equal as mathematically possible” (Flanigan et al., 2021, p. 548). The developers of these algorithms have also developed the sortition algorithm used in the world’s first Global Assembly on Climate. Through a “global location lottery,” organisers selected 100 points of the world map based on the National Aeronautics and Space Administration database on population density. The organisers sought to uphold algorithmic transparency by uploading the data and open-source code used in the global lottery online. However, one could argue that “fairness” in this case, remains contentious. It is, for example, notable that a global assembly on climate does not have participants from Australia, New Zealand, and the Pacific Island—countries that are among the most affected by climate change. This raises larger issues about what it means for algorithms to be fair in randomly selecting participants in citizens’ assemblies.

5 See https://globalassembly.org/sortition.
1.5.3 Institutions

Finally, institutions are critical sites of deliberation. By institutions, I refer to formal sites of decision making, such as courts, healthcare, and welfare systems. From a deliberative perspective, institutions work best when decision-makers carefully consider different arguments, deliberate together, and decide for the common good (Christiano, 2012). For others, institutions can enhance their deliberativeness by taking the recommendations of ordinary citizens seriously, whether in minipublics or other less organised forms of public input (Curato et al., 2017; Ercan et al., 2019).

Some scholars content that institutions in an algorithmic society can improve their decision making by using algorithms. For example, despite his concerns regarding algorithms promoting polarisation, Sunstein (2017; 2019) is quite supportive of algorithms in the criminal justice system. In making an argument for the capacity of algorithms to correct human cognitive biases in criminal justice, he argues that algorithms make more accurate predictions and can be designed to avoid making biased decisions (Sunstein, 2019). Sunstein bases his argument on a study that compares the accuracy of decisions made by human judges, compared to algorithms about individuals’ recidivism and “flight” risk (i.e., the risk they would not reappear in court). This study suggests that judges overestimate the risk when the offence at hand is serious and underestimate it when the offence is minor without contextualising risk against priors. Algorithms as statistical models, by comparison, make more accurate predictions based on other factors rather than the gravity of the present offence.

However, the applications of algorithmic decisions in criminal justice in the US shows that the factors accounted for in these statistical models are biased and racialized (Angwin et al., 2016). This line of thinking is conceptualised as “algorithmic idealism” by Davis and colleagues (2021) whereby algorithms are perceived to overcome the limitations of biased human decision making, thereby overlooking the structural problems. This, however, has been the subject of critique by various scholars. In the US, “algorithmic approaches to resolving questions of racial justice evince a profound discomfort with the uncertainty inherent in democratic processes and interpretive judgment” (Kahn, 2018, p. 199). Others are concerned that algorithms serve to automate inequality and exacerbate the political exclusion of marginalised communities (Benjamin, 2019; Eubanks, 2018). Therefore, the ideal of algorithmic equality grounded in statistical neutrality, while normatively defensible,
undermines the functions of justification, discretion, and reflexivity necessary for deliberative democracy.

1.6 Promoting Deliberative Capacity

The previous section identified the various ways in which the literature on deliberative democracy has characterised the role and consequences of algorithms in three sites of deliberation. What the literature thus far demonstrates is a mixed picture on how algorithms can promote deliberative democracy.

In this thesis, I aim to put forward a systematic account of how algorithms promote or hinder deliberative democracy. I do this by anchoring my argument on John Dryzek’s (2009) concept of *deliberative capacity*, which refers to a polity’s capacity to host authentic, inclusive, and consequential deliberations. Deliberative capacity in this thesis follows Nicole Curato’s (2015, p. 99) argument for using deliberative capacity as “an indicator of democratic quality.” Hence, *I argue that algorithms can promote deliberative democracy if they build a polity’s deliberative capacity.* Algorithms can promote deliberative capacity deliberately or inadvertently. They can promote deliberative capacity in the *process* of creating, scrutinising, and using algorithms, or in the *effects* of algorithms which deliver deliberative goods. Following Dryzek’s (2009, p. 138) conceptualization, algorithms can promote deliberative capacity to “a greater or lesser degree.” This is a useful reminder to avoid a binary thinking on whether algorithms are good or bad for deliberative capacity building. Instead, the contribution of algorithms to deliberative democracy is a matter of degree, to which deliberative capacity reflects the democratic quality of an algorithmic society.

The following sections discuss the components of deliberative capacity. I will define each of these components and explain how the literature on media and communications, law, and political theory have thus far conceptualised algorithms in relation to each component. I conclude the section by outlining scenarios for algorithms to promote deliberative capacity, based on the literature surveyed.
1.6.1 Authenticity

Authenticity refers to the deliberative ideal of open, noncoercive, and reflective reason giving (Dryzek 2009). It presupposes status equality among agentic citizens who uphold their obligations of reciprocity and mutual justification (Gutmann & Thompson, 2004). These justifications ought to be expressed in ways that others could understand so that one’s arguments can be scrutinised, weighed, and examined by one’s coparticipants in deliberation.

The connection between algorithms and authenticity is discussed in the literature in two ways. First, the literature, for the most part, has demonstrated how algorithms undermine the authenticity. There are two ways algorithms undermine authenticity.

**Algorithms Undermine Authenticity by Curtailing Citizens’ Deliberative Agency.** Deliberative agency refers to “the performance of political justification in the public sphere” (Alnemr et al., 2020, p. 1). The performance of political justification takes many forms, from telling stories that explain one’s considerations in political decision making to using other creative forms of speech, such as images and music to establish the gravity of one’s reasons. Algorithms undermine deliberative agency by undermining citizens’ creative capacity for constructing reasons that others can understand. Algorithms produce “statistically deterministic” accounts of citizens (Hilderbrandt 2011) by collecting and using big data to predict our behaviour, while negating the role of voice in politics (Couldry, 2017). Algorithms undermine autonomous evaluation of information and decision making and instead, “nudge” citizens to think and act a certain way. For example, the Obama re-election campaign targeted voters based on insight from Facebook behavioural algorithms (Benkler et al., 2018). Facebook turned to this potential and tweaked their algorithms to “nudge” users to vote, appearing as a “voting button” and found direct effects of this “nudge” on voting behaviour. This nudge bears larger consequences on voting patterns should Facebook further exploit it (Caplan & boyd, 2016; Caplan & Reed, 2016). Karen Yeung (2017) characterises this as a “hypernudge” for the vast extent of algorithms’ roles in manipulating the range of information citizens are exposed to. The “hypernudge,” Yeung (2017, p. 132) argues, is based on the perception of citizens as “consumer[s] of digital services, rather than a politically active citizen[ry] engaged in processes of public deliberation.” Citizens are perceived as data whose preferences are subject to manipulation, instead of agents of mutual justification.
Algorithms’ Opaqueness Promote Coercion rather than Authentic Deliberation.

Information about algorithms is not easily accessible. Algorithms, as discussed in Section 1.4 are kept in a “black box” where its internal workings and dynamics are opaque and indiscernible to ordinary citizens. Producers of algorithms make “arbitrary decisions” obscured from the view of the wider public (Kemper & Kolkman, 2019). This poses a problem for authenticity as citizens lack the power to scrutinise the very system put in place for them to carry out their vital activities. Meanwhile, producers of algorithms are not held accountable for the decisions they make as they construct algorithms. A consequence of this is the erosion of democratic control over the public sphere (Sales, 2020). People are left powerless in demanding accountability and scrutiny of algorithmic decisions and power (Pasquale, 2015). It is therefore not surprising that algorithms and AI technology are increasingly becoming part of authoritarian power and control, even in democracies. On the one hand, algorithms and AI allow authoritarian regimes to “maintain power through a mixture of coercion… and co-optation,” that is, algorithms that would flag dissident communication encourages conformity out of fear of reprisal (Feldstein, 2019, p. 42). On the other hand, this is also a challenge to civil society in democracies that can “misuse” AI for online communication and beyond (e.g., in the streets for surveillance) (Feldstein, 2019).

Second, while the literature has highlighted how algorithms can undermine the principle of authenticity, there is also emerging literature that imagines possibilities for algorithms to build this component of deliberative capacity. This can take the form of including citizens in the process of “algorithmic design” or creating opportunities for algorithmic accountability where the design of algorithms is justified to citizens (see Schuilenburg & Peeters, 2021). For Jakko Kemper and Daan Kolkman (2019), correcting the coercive power of algorithms should be done through the engagement of “a critical audience.” Some of this critical audience may be engaged in “algorithmic resistance” (Velkova & Kaun, 2021), in which algorithms and automation are used to resist “algorithmic oppression.” For example, Noble (2018) highlights some examples of algorithmic oppression where Google search engine results returned photos of Black people when searching for “gorillas” or showed “ape” as a related search term to photos of Michelle Obama. These forms of algorithmic oppression were fixed when “critical audiences” campaigned against these harms, scrutinised Google’s racist algorithm, and in turn, reshaped the informational landscape by holding big tech accountable for their algorithmic decisions. Algorithms may be coercive and powerful, but, as these
examples demonstrate, they could also be subject to public scrutiny and deliberative accountability.

1.6.2 Inclusiveness

Inclusiveness refers to the opportunity for all affected actors or their representatives to take part in deliberation (Dryzek 2009). The capacity for inclusive deliberation can be descriptive or discursive. Descriptively, this would be the inclusion of citizens affected and/or subjected to decisions. Discursively, this means inclusion of the relevant range of interests and discourses (Dryzek, 2009).

Algorithms shape the public sphere by making decisions about the descriptive (i.e., who and what is relevant or censored) and the discursive (i.e., what is trend worthy and should be visible) aspects of inclusion. The literature, for the most part, has focused on the ills caused by algorithmic inclusion. For descriptive inclusion, Donald Trump, Trump Jr., Milo Yiannopoulos, or similar political right-wing figures are rewarded by the algorithms to be visible. Discursively, algorithms make “fake news” and “hate” go viral (Lobo et al., 2021), often with the aid of bots. To expand more on these ideas, in this section I review the literature in light of three ways in which algorithms affect inclusiveness in the public sphere.

Access to Information and Knowledge. Algorithms enact politics of discursive inclusion and exclusion by imposing an order of “relevance” to organise access to information and knowledge. Definitions of relevance that feature in algorithmic selection of news and information raise anxieties about how the “automated public sphere serves political information to voters in deeply privatized and hidden ways” (Benedetta & Pasquale, 2020, p. 3). This anxiety is especially a concern to democratic deliberation when citizens depend on social media as free platforms and sources for information (Holst & Moe, 2021). Hence, algorithms contribute to political manipulation as bots and big data are mobilised to generate profit or deliberately spread fake news and disinformation (Chambers, 2021). Another democratic threat of algorithmic relevance is personalising content and information, which reinforce attitudes and create echo chambers that prevent diversity in public opinion (Cho et al., 2020). Thus, what counts as “relevant” information are those that are profit-seeking, manipulative, personalised, or all three. For actors and citizens who do not tell stories that generate profit, that are used for
manipulative political ends, or that speak to mainstream politics, their capacity to be included in the public sphere is diminished.

The extent of algorithmic manipulation of access to information and knowledge extends to journalism and news reporting. With much attention given to what is algorithmically relevant, stories and voices in news reports mirror similar politics of inclusion and exclusion. If your voice and/or story are not “viral” or “relevant,” it would not be featured by the media. Logan Molyneux and Shannon McGregor (2021) find that algorithmic relevance shifts the practice of journalism to a practice of verifying and building stories about “trending tweets.” This shift “has the potential to shape the participants in and the content of public deliberation” (Molyneux & McGregor, 2021, p. 2). In this way, Outi Lundahl (2020) argues that algorithmic curation resembles legacy media which affect the democratic potential of journalism in public deliberation. Thus, the concentration of power in algorithms which decide what is being reported and what becomes visible needs a democratic check.

**Connecting Publics.** In respect to descriptive inclusion, the role of “relevance algorithms” is specific to the construction of publics. Online, publics are constructed by algorithms calculating the proximity of views among social media users (Birkbak & Carlsen, 2016; Crawford, 2016; Gillespie, 2014). Calculations of relevance are arguably how algorithms strengthen “echo chambers” that inhibit the “exchange of views and deliberation” (Birchall, 2020, p. 964). These echo chambers are often associated with gathering undemocratic publics. Algorithms augment the reach of affective right-wing populist discourses and polarised views in public deliberation on social media (Kopytowska, 2020). Steven Watson and Naomi Barnes (2021) explore how algorithmic virality and relevance have contributed to the visibility of “populist moments” in England and Australia. The algorithm is not inherently geared to promote populist messages because it is incapable of evaluating the substantive content of messages. In organising their online activism, new right-wing groups are aware of algorithmic relevance and select techniques that would make their messages “viral,” even if briefly. Although short-lived, a “populist moment” bears longer consequences on political action (Watson & Barnes, 2021).

Thus, we can start seeing relevance algorithms as “political philosophies” about what publics are and who belongs to a public, specifying the parameters for descriptive inclusion and exclusion (Birkbak & Carlsen, 2016). By turning to the critique of the norms underpinning
Facebook as an example, Jennifer Forestal (2021) offers a critique on the lack of a complementing algorithm which highlights the experiences of others, not on the personalisation algorithms per se. In other words, the problem is that Facebook does not create a shared space among people with diverse and even competing concerns. Online democratic spaces ought to be designed to accommodate enclaves, as well as exposure to others (Forestal, 2021).

**Presence of Bots.** For the third aspect of the public sphere, the presence of bots has implications for discursive and descriptive inclusion. This is demonstrated by cases wherein bots were used in political manipulation using strategies such as “computational propaganda” and “zombie electioneering.” In computational propaganda, bots are used to spread propaganda, thus inflicting discursive manipulation. In zombie electioneering, bots are used to fake identities of nonexisting supporters, such as the creation of bots with “American Latinx” profiles supporting Trump during the 2016 elections to fake his legitimacy amongst the Latinx community (see Arnaudo, 2017; Bolsover & Howard, 2019; DiResta, 2018; Howard & Bradshaw, 2018; Woolley & Howard, 2016). Hence, the presence of bots distorts discursive and descriptive inclusion.

In summary, politics of inclusion and exclusion in the digital public sphere are algorithmically mediated, whether by organising communication or making contributions. This concentrates power in algorithms which shape deliberation and political action. Algorithmic mediation in itself can be undemocratic as coercive and unjustified power or undemocratic considering the larger contexts these algorithms are part of. Above all, the coercive power of algorithms is reinforced by increasing data collection that facilitate making statistical predictions and decisions about society. In this sense, society’s capacity to contest coercive algorithmic power that determine patterns of inclusion and exclusion is unfulfilled because these decisions are deemed accurate and neutral. The consequences of power and politics, enacted by algorithms, affect varying spaces and aspects of democracy, which will be expanded upon in the next part of this review.

**1.6.3 Consequentiality**

*Consequentiality* refers to the ideal that deliberation ought, in one way or another, to contribute to shaping collective outcomes (Dryzek, 2009). A polity with deliberative capacity
treats deliberation not as a “tokenistic enterprise” but a “normative commitment that participants’ inputs are considered when determining collective outcomes” (Curato, Hammond & Min, 2019, p. 5). Consequences of deliberation vary. Deliberation could bring about a shift in policy, generate new discourses that amplify the voices of disadvantaged communities, or prompt decision-makers to provide justifications for their decisions, especially when these decisions contradict the outcomes of citizen deliberations.

The literature, for the most part, has signalled the different ways in which algorithms can undermine the process of public deliberation in shaping collective outcomes. Algorithms have the power to decide the relevance, and, therefore, the fate of discourses emerging from citizen deliberation to make an impact on collective political action (see Tufekci, 2015). As gatekeepers of discourses, algorithms filter which discourses citizens and decision-makers should consider in their respective deliberations. In some instances, as discussed in Section 1.5.3 (institutions), some algorithms were even designed to replace human deliberation, based on its promise of fairness, epistemic superiority, and neutrality—virtues that are celebrated in deliberative decision making.

This account, while convincing, may oversell the power of algorithms in undermining the outcomes of public deliberation. Some scholars underscore that the public sphere is shaped by the interactions of multiple actors, not algorithms alone. Specifically questioning the literature on relevance algorithms “blamed” for creating “filter bubbles,” Kartik Hosanagar and Alex Miller (2020) ran a simulation of a recommendation algorithm and suggested that data, people, and algorithms together contribute to the visibility of content, not the algorithm alone. Thus, studying epistemic pathologies should not focus on algorithms but rather the combination of data, algorithms, and interactions (Hosanagar & Miller, 2020). The question, therefore, rests on how algorithms, data, and people can promote consequences of deliberation together.

### 1.6.4 Scenarios of Algorithms Building Deliberative Capacity

Thus far, I have argued that algorithms can promote deliberative democracy based on the extent that they promote deliberative capacity. Based on the discussions above, one can surmise that the literature has mostly provided a bleak view of the potential of algorithms to build a polity’s deliberative capacity, although there are some scholars who identify scope for algorithms to promote deliberative democracy.
Where then does this leave this thesis? In this section, I outline three scenarios in which algorithms can promote deliberative capacity. I outline these scenarios to transcend the binary thinking of whether algorithms are good or bad for democracy. Instead, I conceptualise the capacity of algorithms to promote deliberative democracy as contextual, instead of inherent. This is similar to the argument that the capacity of minipublics to build a polity’s deliberative capacity is not a given. The contributions of minipublics in deliberative capacity building depend on a number of factors, starting with how it was designed and implemented, the political contexts in which they have taken place, and the integrity of the intentions of its conveners. The same logic applies to algorithms. I argue that algorithms can be (1) deliberately designed to promote deliberative democracy (an ideal but highly unlikely scenario); (2) implemented in a particular context that inadvertently promotes deliberative democracy (an acceptable and possible scenario); and (3) deliberately designed to actively undermine democracy (a harmful but expected scenario).

**Deliberate (An Ideal but Highly Unlikely Scenario).** This is a scenario which involves algorithms that are designed to observe or translate deliberative democratic ideals. The process of designing these algorithms is democratic to the extent that it involves citizen deliberation. For algorithms to sustain their contributions to deliberative capacity, updating their logics ought to include citizens’ input and reflections. In this scenario, an algorithmic society is a reflexive society or one where ordinary citizens can scrutinise the algorithms and reflect on the way the public sphere, minipublics, and institutions are shaped by algorithms. This, however, is a highly unlikely scenario, given that the profit-driven character of tech platforms and developers of even democratic spaces are unlikely to develop these democratically (i.e., with citizens).

**Contextual (An Acceptable and Possible Scenario).** This is the most likely scenario as it follows from the theoretical proposition that some “nondeliberative” communication can have deliberative consequences (see Bächtiger et al., 2018; Chambers, 2018). Zizi Papacharissi (2015) conceptualises this contextual contribution as the role of algorithms in the phenomenon of “affective publics.” She argues that publics gathered and mobilised for movements such as the Arab Spring and Occupy Wall Street were connected through relevance algorithms but through “algorithmically rendered materiality.” Affect, not algorithms alone, plays a role. Flows of affective content composed of “opinions, facts, sentiment, drama, and performance” expressing indignation, disdain or discontent, are present in the interaction between affect and
algorithms (Papacharissi, 2015, p. 127). Nonetheless, “affective publics” can also be right-wing publics that use affect and algorithms to connect and mobilise. These examples demonstrate that algorithms’ capacity to promote deliberative democracy depends on the context.

**May Not Contribute (A Harmful but Expected Scenario).** In this scenario, algorithms are specifically designed to frustrate efforts of political mobilisation and action. While Papacharissi (2015) celebrates affective publics and the contextual contributions of algorithms, others are less celebratory of the roles of algorithms in connecting publics. One example is when algorithms on social media platforms filter content that will be “relevant” or “trendy.” Tarleton Gillespie (2014) explains that algorithms create “calculated publics” by exposing users to content based on a secret formula that calculates the “traceability” of online activities. In other words, it is difficult for political actors to be included and become visible when the algorithms of relevance are arbitrarily, and secretively, deciding the worthiness of content. On the ramifications of calculated publics, Wael Ghonim, an activist who used Facebook to mobilise for the Egyptian Revolution in 2011 as part of the Arab Spring, argues that the logic of algorithms here is one aimed at maximising profit from the attention economy (Gardels, 2016). Algorithms sort content based on what would get our attention the most, and particularly the most profitable content. In Ghonim’s account, the algorithms of relevance reward emotive and sensational content rather than conversations that contribute to civic and democratic life, because the latter is not profitable. Therefore, this logic of organising our communication and publics makes it hospitable to the pursuits of demagogues like Donald Trump who use affect and profit to be relevant, visible, and viral (PdF YouTube, 2016). This is a scenario where algorithms do not contribute to deliberative capacity as they frustrate the connective potential of social media in bridging the voices of marginalised citizens.

These scenarios, as I will discuss in Section 1.8 (chapter summaries), provide a vocabulary to describe the extent to which algorithms promote deliberative democracy.

**1.7 Research Method**

Before discussing the structure of the thesis, I will first explain the process in developing my arguments. My approach involves using normative theory, in particular the ideals of deliberative democracy, as “yardsticks in assessing and criticising arrangements in
real existing democracies” (Chambers, forthcoming). On this method of applying normative theory to critique and evaluate real-world practices, Chambers (forthcoming) explains:

[T]he focus of applied theory is the real world. … it interrogates, evaluates, and analyses real-world arrangements from the normative point of view, but it does not always have to be critical in the negative sense. It is possible, especially in a comparative framework, to point to the positive developments and good practices. But in an imperfect world beset by power, interest, and money, it stands to reason that the real world will always fall short even of our mediated realistic ideals. … For deliberative democrats, progress (if it occurs at all) is piecemeal, incremental, and often achieved by pointing out the egregious features of institutions and people to live up to the ideals of deliberative democracy. (Chambers, forthcoming)

In other words, my use of normative theory in this thesis is not a contribution to (re)define the ideals of deliberative democracy, but to understand how these ideals can be used to critique the shortcomings in the real world and identify the promising developments that affect the emancipatory potential of deliberative democracy. My interest in the real world is not limited to the real world of “existing democracies.” In the quote above, Chambers reminds us that these ideals can be used to critically assess and examine power and interest as they undermine the ideals of deliberative democracy, including in contexts where the political system is undemocratic so that the ideals are only highly aspirational.

Moreover, this use of normative theory does not take a normative position on algorithms per se. Rather, it contextualises the normative account of algorithmic contributions covering both democratic harms as well as potentials. There are different degrees to which algorithms both harm and benefit democracy in different contexts. The position that I form is that contributions of algorithms to deliberative capacity are contingent, not inherent.

While the contributions of this thesis to deliberative democracy research are specific to locating algorithms and their roles in political theory, the contributions to digital democracy and communication are specific to the methods. The most popular research method in this field has been computational social science. Christine Hine (2005) discusses the methodological direction that researching “the digital world and subjects” would suggest that it is most appropriate to use methods that are “digital” rather than “traditional.” Nonetheless, my objective was to showcase how normative theory is just as illuminating and should be at the forefront of critical algorithms research. This is not to undermine the value of computational
research methods commonly used to study online deliberation and communication. Instead, I argue that theoretical methods should not be sidelined. Most importantly, this theoretical work is applied to empirical examples. Much of the contextual and conceptual rigor of the field will rely on these insights. Questions, insights, and cases found through this theoretical work can then become the basis for informing empirical research.6

Algorithms have had a role in informing the sources and publications considered and cited in this thesis. I acknowledge that finding, filtering, and identifying relevant sources have been shaped by search and selection criteria as well as algorithmic relevance of what is made visible based on different databases. With this in mind, I have tried different strategies to manage algorithmic filtering by using different browsing locations through Virtual Private Networks (VPNs) in order to diversify access to content, occasionally and not systematically.7 Other approaches included using my knowledge of Arabic to search sources beyond English-speaking spheres.8 Online search has been a major component of my desk research, but I have not been nudged by algorithmic filtering alone in learning about examples of algorithms. Exchanges with my colleagues have been highly valuable in drawing my attention to some of the empirical examples the algorithms would have not recommended.9 Algorithms might be helping us greatly in research and beyond, but intelligent machines cannot and should not supplant human judgment and discretion.

In summary, the main objective of this thesis is to use deliberative theory as the normative lens to algorithm research for two reasons. First, the roles of algorithms need normative critique, which not only highlights their democratic limitations and potentials but also showcases how citizens and society should interact with these algorithms. Second, using

6 For researchers who are interested in studying algorithms empirically, the following are some relevant considerations. Kitchin (2017) presents an insightful discussion about the possible approaches and methods to studying algorithms. Some of these approaches include studying how algorithms “work in the world” or studying the design and production processes of algorithms using mixed methods (Kitchin, 2017). An interpretive approach can benefit from studying what Bucher (2017) terms as “the algorithmic imaginary”: a study of how actors perceive and interpret how algorithms work and adjust their interactions accordingly. Particularly on the selection of methods, Hine (2005) discusses how researching digital objects involves “negotiating the troubled territory between innovation and precedent in discussions of methodological validity.” This can be put into perspective when reading Karpf’s (2012) suggestions for using computational “quick-and-dirty” methods, which they then reflected on and reconsidered (Karpf, 2019).

7 Google search results and their ranking for instance vary based on the geolocation of where a query is made. And occasionally because that would have been an endeavour in its own right, I had to manage with available resources (even when it is algorithmically-mediated) over the course of conducting this research and writing this thesis.

8 This is a specific reference to sources analysed in the third chapter: robots as partners in deliberative democracy.

9 It has been a great contribution from my colleagues: Kei Nishiyama to point out to me the example of the AI robot politician coming third in the mayoral race in Tama City, Japan when I discussed with him the example of SAM, a virtual, AI chatbot politician in New Zealand.
deliberative democracy as a theoretical lens allows for contextualising the different degrees to which algorithms contribute to democratic communication and decision making.

1.8 Chapter Summaries

This is a “thesis by published works” consisting of four journal articles. The review status of these journal articles is summarised in Table 1 below. Although each article stands on its own, they altogether respond to the question: “How can algorithms promote the practice of deliberative democracy?” As discussed earlier, the need for this research is twofold. The first contribution is to understand algorithms in different contexts of practicing deliberative democracy. The second contribution is to highlight how the categories of deliberative capacity can be used to normatively scrutinise the roles of algorithms.

Using deliberative capacity to understand how algorithms affect democracy, I explore three sites of democratic participation and engagement in four articles (referred to chapters henceforth): the public sphere, minipublics, and institutions. In this sense, accounts of algorithms in the public sphere are concerned with algorithms that control and contribute to speech, visibility, relevance and inclusion in public deliberation. In minipublics, the account is concerned with algorithms that facilitate structured deliberation. In institutions, it is an account of algorithms that affect the ideal of self-government by governing public life according to unjustified logics.

In reading the next chapters of this thesis, stylistic changes have been applied to ensure that the compilation of the four articles and conclusion chapters are consistent in respect to reference style and spelling, which do not reflect the journals’ style guides.

Each account builds on arguments, ideas, or provocations from the others. Taken altogether, the four accounts explore algorithms in deliberative democracy. Chapter 2 presents an account of algorithmic harms in the public sphere that can be contested by citizens. This chapter puts a spotlight on the consequentiality of everyday political action in contesting algorithmic power and, in so doing, promoting deliberative capacity in the public sphere. Chapter 3 turns to the use of algorithms and AI to replace or assist human facilitators in minipublics to account for blind spots that are missed by an overemphasis of the promises of such technologies. Theoretical and practical implications of the identified blind spots are
discussed in this account. Chapter 4 explores the role of algorithms in institutions governing public life which shortcut democratic decision-making. I critique algorithms by extending Cristina Lafont’s work on shortcuts to democracy. Chapter 5 also locates algorithms in the context of institutions and the public sphere by applying the normative criteria of deliberative capacity to scrutinise the contributions of robot communication. The example of robot communication covered in the chapter goes beyond the focus on bots on social media to examine robots participating in offline and online political platforms. The four articles are followed by a conclusion chapter which covers my overall findings.

Table 1 List of Publications and Review Status

<table>
<thead>
<tr>
<th>Chapter No.</th>
<th>Article Title</th>
<th>Journal</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Defending Deliberative Democracy in an Algorithmic Society: Building Deliberative Capacity through Contesting Algorithmic Harms</td>
<td>Political Studies</td>
<td>Under review</td>
</tr>
<tr>
<td>3</td>
<td>Emancipation cannot be Programmed: Blind Spots of Algorithmic Facilitation in Online Deliberation</td>
<td>Contemporary Politics</td>
<td>Published in 2020</td>
</tr>
<tr>
<td>4</td>
<td>Democratic Self-government and the Algocratic Shortcut: The Democratic Harms in Algorithmic Governance of Society</td>
<td>Contemporary Political Theory</td>
<td>Under revision</td>
</tr>
<tr>
<td>5</td>
<td>Can Robots be Partners in Democracy? Deliberative Democratic Values for Politics with Robots</td>
<td>Information, Communication and Society</td>
<td>Submitted</td>
</tr>
</tbody>
</table>

In the following parts of this section, I present the chapter summaries. Each chapter summary starts with a brief description of contemporary vignettes which shaped the research question, examples, and key arguments made in the respective chapters. In addition to these vignettes, the chapter summaries discuss the connections to other chapters as well as the main research question.

1.8.1 Chapter 2: Defending Deliberative Democracy in an Algorithmic Society: Building Deliberative Capacity through Contesting Algorithmic Harms

The extent to which algorithms shape our communication, knowledge, and capacity in making decisions by filtering content, controlling speech, and automating bots in the public sphere raise questions about the democratic quality of the algorithmic society. Extant scholarship informs us about the harms of algorithms and bots in the public sphere, such as undermining inclusiveness, the epistemic quality of communication, and the capacity for
internal deliberations (Chambers, 2021; Sunstein, 2017; Tanasoca, 2019). Possible solutions proposed include designing deliberative chatbots and creating deliberative filtering algorithms, as well as modifying some social media algorithms such as that of Facebook (Bozdag & van den Hoven, 2015; Forestal, 2021; Gardels & Berggruen, 2019; Graham & Ackland, 2017; Helberger, 2019). These proposals are either facilitated by experts or require reform at the institutional level (in big tech companies or through state intervention). These proposals conceive citizens at the receiving end of these democratic interventions, rather than as cocreators or agents driving this change.

In this chapter, I argue that citizens’ contestations of algorithms also deserve attention. These everyday actions of contestation demonstrate the capacity of creative political agency in bringing change. Hence, in what ways can citizens build deliberative capacity in an algorithmic society? I showcase three ways by which citizens can contest algorithmic harms in the public sphere by building deliberative capacity. Using demonstrative examples, I discuss how citizens can contest algorithms to promote: (1) inclusiveness by challenging algorithmic censorship, (2) the epistemic quality of information by combating misinformation, and (3) internal reflection by increasing awareness on the roles of algorithms.

1.8.2 Chapter 3: Emancipation cannot be Programmed: Blind Spots of Algorithmic Facilitation in Online Deliberation

Considering the perceived positive prospects of using algorithms in scaling up minipublics, the chapter examines the extent to which citizens should be involved in shaping these algorithms as a reflection of their democratic quality. Focused specifically are the arguments about the promise of algorithms and AI in facilitating minipublics. Because the overarching objective of this thesis is to see the algorithms as a democratic question, I ask the following question in the chapter: “What considerations should algorithmic facilitation observe to maintain the democratic quality of online minipublics?”

Here, I argue that the design and implementation of minipublics have blind spots, some of which are specific to algorithmic facilitation. I highlight the theoretical and practical implications of expert-imposed interpretations enacted through algorithms. The theoretical implications are specific to narrow interpretations of deliberative democratic ideals, such as that of inclusion. Meanwhile, practical implications involve designing steps of citizen
involvement in the development of algorithmic facilitation to reflect on cultural and communicative norms, as well as improve the democratic quality of the design and implementation of minipublics.

1.8.3 Chapter 4: Democratic Self-government and the Algocratic Shortcut

Much of the literature on the undemocratic consequences of algorithms is concerned with algorithms used in governing society, making decisions in institutions affecting citizens ranging from access to welfare and assignment to foster families, to sentencing in criminal justice. These algorithms have been often critiqued for guarding systemic discrimination and racial injustice behind the veil of unjustifiable and inexplicable algorithmic decision making. While deliberative democrats have hardly paid attention to this role of algorithms in their scrutiny, my aim is to bring attention the power of algorithms in this role and its relevance to deliberative theory.

In this chapter, I ask: “How can algorithmic decision making be aligned with the ideal of self-government?” In exploring this question, I draw on Cristina Lafont’s (2019) critique of “shortcuts to democracy” and explain how algorithmic decision making constitutes a “shortcut.” In the chapter, I argue that the algorithmic shortcut shares the same normative foundations as epistocracy. Both cannot be accepted in democratic deference (division of labour), so long as these shortcuts undermine the inclusion of citizens in decision making.

To understand how the “algocratic” shortcut undermines the ideal of self-government, I present a lens that brings together Lafont’s (2019) discussion of shortcuts to democracy, Bohman’s (2000) exploration of epistemic dependence on experts, and Dryzek’s (2009) concept of deliberative capacity. Through this lens, I operationalise Lafont’s definition of self-government as “reasonable acceptance upon reflection.” The conditions for self-government, hence, are equality in influencing decisions and full and fair consideration of reasons, as well as free and effective voice. To demonstrate how algocratic governance undermines these conditions, I use the attempt to democratise algorithms in the European Union and the UK through legislation and citizen deliberation as an example.
1.8.4 Chapter 5: Can Robots be Partners in Democracy? Deliberative Democratic Values for Politics with Robots

The question of robots as partners in democratic deliberation has been alluded to in Chapter 3 on algorithmic facilitation where I discuss two types of relationships between humans and algorithms in online deliberation: as partners (supporting humans) and as minds (replacing humans). In Chapter 5 the idea of “partners” takes a slightly different turn. In this chapter, I look at robot communication that also takes place in physical interactions such as in policy forums and other political platforms.

Here, I ask: “Under what conditions can robots be partners in democratic communication?” The chapter engages with some normative arguments about the democratic benefits of prohibiting robot communication. Given that there are examples of robot communication with ramifications for democracy, I argue that normative judgments about robot communication should not only be focused on what robots are capable of. To expand on this argument, I explain how the normative criteria of deliberative capacity can apply to robot communication. Particularly, the contributions of robot communication to deliberative capacity applies to institutions and political actors involved in politics with robots. This is because robot communication is characterised by an asymmetry where humans are the meaning-makers in these interactions. I apply deliberative capacity to the example of Sophia the robot, the world’s first humanoid citizen to highlight how it contributes to deliberative capacity. From this assessment, the example of Sophia the robot briefly contributes to inclusion but is used overall for inauthentic and manipulative political ends.

1.9 Limitations and Boundaries of the Thesis

A thesis on deliberative democracy and algorithms can travel in many different directions, and so it is important for me to end this introduction by setting the boundaries of my thesis, as well as the limits of the scope of my analysis.

This is a theoretical thesis that does not present novel empirical findings based on systematic research. What the thesis does, however, is to theorise the extent to which algorithms can contribute to deliberative democracy by grounding examples in their contexts.
In terms of theory, this study of algorithms lies within the parameters of deliberative democratic theory. The questions, examples, and conclusions drawn in each account of algorithms speaks to priorities and concerns in deliberative theory, which would differ substantially had I chosen a different theoretical lens. For instance, had my thesis been guided by queer political theory, the question of algorithmic relevance in constructing publics would be geared towards understanding whether algorithms reflect and solidify heteronormativity and sexual identity in defining “relevance.” A similar line of thinking can be extended to other theoretical traditions, such as postcolonial or poststructuralist theory. Questions, cases, concerns, and conclusions would have been different. Ultimately, as I position this inquiry in the bigger concern for democracy in the algorithmic society/turn, deliberative theory sets an adequate limit.

Conceptually, I focused on refining the understanding of deliberative capacity and the ideals of inclusion, authenticity, and consequentiality at the expense of other equally important concepts and ideals. Deliberative capacity is used in a conceptually specific manner that follows Curato’s (2015, p. 99) argument to use it as an “indicator of democratic quality”, as explained earlier in Section 1.6. To maintain a coherent, systematic and contained account of algorithms, especially to respond to the wider question about algorithms and democracy, I did not attend to other concepts in deliberative theory. Moreover, deliberative capacity is often used in conjunction with the “deliberative systems” framework, which is not used in this thesis. The deliberative systems developed by Dryzek (2010) assesses the deliberative quality of a system by tracking the dynamics of transmission of discourses and reasons from the public to empowered space. This would necessitate an inquiry focused on one or two key policies or decisions, which is not the case in this thesis. The illustrative examples used throughout my work are spread across space and time, and often not strongly connected to an empowered space (or there would be multiple empowered spaces). Chapter 4 is an exception, nonetheless, where the example is one that could use the deliberative system approach but the focus of this chapter is to build on Lafont’s (2019) examination of shortcuts to democracy.

There are two concepts that I specifically think are important to problematise, which however lie beyond the limits of this thesis. First, is the concept of “affected people”: who is affected by algorithms? Although it has been the “citizens” across the four accounts, I did not problematise the notions of “citizenship” in my thesis. It is an important conceptual concern as algorithmic decisions affect people who are not “citizens” (e.g., algorithms in refugee and
migration governance making decisions about the freedom of movement and border control, as well as access to services). The second concept which I think is also important to problematise is “harms” in “algorithmic harms.” No definition of harm has been provided, such as against John Stuart Mill’s definition of harm, that is if applied to algorithms would be, an algorithm can reflect any logic so long as it does not harm others and at this threshold of harm society ought to scrutinise and control this algorithm (Mill, 1859/2011). While the thesis has been limited in this sense, further exploration of algorithms against these concepts opens avenues for future research that consider increasing examples on the roles of algorithms.

1.10 References


Chapter 1


Society and Market – Constitutional Dimensions, Florence, Italy.
https://lirias.kuleuven.be/retrieve/540651

https://doi.org/10.1080/1369118X.2019.1657162


https://doi.org/10.1080/14767724.2021.1882292


https://doi.org/10.1080/1369118X.2016.1186713


CHAPTER 2

Defending Deliberative Democracy in an Algorithmic Society:
Building Deliberative Capacity through Contesting Algorithmic Harms

Abstract

The scholarship on deliberative democracy has long been cognizant of the harms algorithms create to the public sphere. Algorithms promote exclusion and political polarisation, undermine the epistemic quality of public deliberation by disseminating misinformation, and compromise internal reflection by amplifying particular discourses through bots. However valid these concerns are, focusing on them alone does not constitute a robust theory of deliberative democracy now that we live in an algorithmic society. In this chapter, I argue that deliberative democracy should not only focus on institutional reform to address algorithmic harms, but also emphasise the importance of creative political agency of everyday citizens. I recast the three algorithmic harms deliberative democrats identified to three ways ordinary citizens can contest algorithms and, in so doing, build the polity’s deliberative capacity necessary to democratise our algorithmic society.
2.1 Introduction

We now live in an “algorithmic society” where the fabric of social life is jointly shaped by algorithmic and human behaviour (Balkin, 2017; Gurumurthy & Bharthur, 2018; Peeters & Schuilenburg, 2021; Wagner et al., 2021). Algorithms primarily shape the way we consume information, vote in elections, organise protests, communicate with each other, make collective decisions, and even regulate our digital intimacy. These decisions are made based on how algorithms are programmed to define a public, and who is expected to interact with it (Amin & Galal, 2015; Chan, 2021; Curato, 2020; Gerbaudo, 2012; Howard & Bradshaw, 2018; Theocharis et al., 2014).

The roles of algorithms in organising and influencing inclusion, knowledge, and communication have attracted the interest of scholars of deliberative democracy. As a normative theory, deliberative democracy places “meaningful communication at the heart of democracy.” It promises emancipation from domination through an inclusive exchange of other-regarding reasons (Bächtiger et al., 2018; Böker, 2017; Hammond, 2019). More specifically, deliberative democracy puts forward a normative aspiration for citizens to be involved “substantively” rather than “symbolically” in democracies, by reflecting on their preferences, free from coercion and manipulation (Dryzek, 2002).

With these ideals in mind, deliberative democrats have been cautious about the role algorithms play in contemporary democracies. Thus far, the literature has identified three harms algorithms cause to deliberative democracy in the public sphere. First, algorithms undermine inclusiveness by creating discursive enclaves. These enclaves not only drive polarisation but also diminish the epistemic diversity necessary for deliberative democracy to flourish (Sunstein, 2017; cf. Strandberg et al., 2017). Second, algorithms undermine deliberative democracy’s epistemic quality (Chambers, 2021; McKay & Tenove, 2020; Tanasoca, 2019). Algorithms are often blamed for promoting misinformation, disinformation, hoaxes, conspiracy theories, and hyperpartisanship (Chambers, 2021; McKay & Tenove, 2020). As falsehoods become normalised in the public sphere, deliberative democracy’s

---

10 This paper is limited to deliberative democracy in the public sphere, as opposed to democratic deliberation in structured forums like minipublics or parliamentary debates. I use the term public sphere in a Habermasian sense, defined as “a domain of our social life in which such a thing as public opinion can be formed” (Habermas, 2018, p. 667).

11 Disinformation and misinformation are not used interchangeably in this article. Mentions of either disinformation or misinformation reflect the wording in the respective sources.
vision of collective decision making based on evidence and reason is severely compromised. Third, algorithms undermine our capacity for internal reflection. The prevalence of bots that artificially amplify particular perspectives distorts the public sphere and leaves us confused, manipulated, and less capable of weighing different arguments.

While these concerns are indeed valid, they are incomplete. In this chapter, I offer an account of how these three algorithmic harms can be addressed. I see this chapter as a humble response to Simone Chambers and John Gastil’s (2021, p. 5) challenge of using deliberative theory to “place digital media in a wider theoretical context, sharpen our understanding of the internet’s worst features, and show the way forward to a better design for digital public engagement.” While there have been many accounts demonstrating how deliberative theory can help us understand “the internet’s worst features” (Chambers & Gastil, 2021, p. 5), more can be done to identify practices that can defend our public sphere from algorithmic harms and to articulate pathways for enhancing deliberative democracy through digital public engagement.

My approach in addressing this challenge is to focus on the different ways in which citizens contest algorithmic harms in the public sphere, and, in doing so, promote the virtues of deliberative democracy. I argue that contestatory practices of ordinary citizens also warrant attention, given that the focus of the literature thus far has been on the accountability of big tech companies and states in regulating algorithmic harms (Chambers, 2021; McKay & Tenove, 2020). Focusing on citizens, I argue, emphasises the value of creative political agency (and not just institutional reform) in defending deliberative democracy as we live in an algorithmic society. In explaining this argument, I use the concept of “deliberative capacity,” that is, a polity’s capacity to host inclusive, authentic, and consequential deliberation (Dryzek, 2009), as a proxy to evaluate the democratic quality (Curato, 2015) of citizens’ contestations.

I develop my argument in three parts. First, I begin by defining algorithms as sociotechnical constructs. Algorithms, I argue, are not just a series of steps, instructions, or logics that automate the organisation of the flows of information and communication devoid of issues of power and bias. Instead, algorithms assume different cultural meanings that are shaped by actors that create and deploy them, as well as the institutions and power relations in which they are situated (Ames, 2018). Algorithms, therefore, are products of political decisions.
and result in political outcomes, which warrant scrutiny and contestation. Second, I review three main concerns that scholars of deliberative democracy have about algorithms. These three concerns, as mentioned earlier, pertain to the impact of algorithms on inclusiveness, epistemic quality, and internal deliberations. Third, I present the three ways in which citizens can creatively contest algorithmic harms. I argue that deliberative democracy can be defended from algorithmic harms through citizen action or responses that build deliberative capacity in an algorithmic society. I conclude this chapter by reflecting on the impact of citizens’ contestations of algorithms to deliberative democracy.

2.2 Defining Algorithms

The literature on digital communication defines algorithms in two ways. The first definition describes algorithms in technical terms, referring to them as a series of instructions given to machines to organise the flows, ranking, and visibility of information, as well as to mimic human communication.\(^\text{12}\) Algorithms are “the heart and soul of all software” executing instructions such as “if A then B” (Simanowski, 2016, p. 50). From this perspective, algorithms are perceived as purely mathematical decisions based on principles of rationality and neutrality (Goffey, 2008). For thinkers who endorse this view, algorithms are better decision-makers than humans (Kahneman, 2011).

This definition has been contested by various scholars. Some criticise the opaqueness of algorithms, such that they have become “a modern myth” (Barocas et al., 2013). What algorithms mean and what they exactly do seem to be in “black boxes” protected by proprietary rights (De Gregorio, 2020; Pasquale, 2015). Even if algorithmic decisions are open to explanation, there remains a strong culture of secrecy where data collection techniques and the inner workings of algorithms are tightly controlled and guarded by private companies and designed to evade democratic accountability (Benedetta & Pasquale, 2020). Algorithms’

\(^\text{12}\) In this light, I characterise bots under the umbrella of algorithms as the point herein to discuss how judgment and communication are translated to a series of automated steps. Other scholars have characterised bots under the umbrella of algorithms as well (see Fuchs, 2018; Woolley & Howard, 2016). The following is a definition which can clarify the connection between algorithms and bots: “A bot is a piece of software code that performs certain online behaviour based on an algorithm” (Fuchs, 2018, p. 80).
illusive rationality restricts scrutiny of the normative judgements built into algorithms and the consequences of these judgements (Noble, 2018).

Therefore, the alternative approach to defining algorithms is in terms of their roles and contributions to society, emphasising algorithms’ “sociological and normative features” (Ananny, 2016, p. 97). Instead of defining algorithms in technical terms, I argue that algorithms can be defined as sociotechnical constructs. In this definition, algorithms are not merely technical logics that are unaffected by political and social structures. Instead, algorithms are “opinions embedded in code” (O’Neil, 2017) as they organise communication based on the assumptions of developers about what should be included, prioritised, and made visible on our social media newsfeeds, maps, streaming services, and shopping recommendations (Bucher, 2018). Algorithms have the power to “make choices, to classify, to sort, to order, and to rank” (Beer, 2017, p. 6). Take the case of “relevance” as a criterion for algorithms designed to filter and recommend content. Relevance is an example of the developers’ choices about organising the visibility and flows of what they consider to be relevant information (Caplan & boyd, 2016).

Understanding algorithms in a sociotechnical manner, I argue, is suitable for understanding deliberative democracy in an algorithmic society. There are two reasons for this. First, a sociotechnical understanding allows us to foreground issues of power, and therefore, understand how algorithms distort the democratic process. In the public sphere, the “epistemic authority” of algorithms in producing, sorting, and promoting content (Carlson, 2014, 2019) has consequences, such as the promotion of misinformation13 (Chambers, 2021; McKay & Tenove, 2020). In the U.S. criminal justice system, sentencing and decisions on recidivism, a process that used to be outcomes of human deliberation, are now made by algorithms. Instead of promoting neutrality, algorithms are observed to be biased against Black Americans because input data for algorithmic decisions are designed to predict risk based on racialized indicators such as neighbourhoods and income (Angwin et al., 2016; Benjamin, 2019). In these accounts, algorithms are perceived as autocratic decision-makers that calculate different variables to generate a single output, operating within the profit-maximising logics of big tech companies or racialized indicators that were not problematised by developers. This, for the most part, has

13 Carlson (2014, 2019) discusses the role of algorithms in journalism and news media production.
been the way deliberative democrats’ critique algorithms (see Sunstein, 2017; Tanasoca, 2019), as I explain in the next section.

The sociotechnical definition of algorithms is useful in helping us clarify that algorithms are not neutral but are shaped by power relations. However, this is not the only advantage of this definition. The second advantage of a sociotechnical definition is that it allows us to recognise that algorithms do not determine their own actions unilaterally but react and adapt to data input from citizens’ interpretations of how these algorithms work (Meng, 2021; Neyland & Möllers, 2017). Despite the power of big tech companies and coders in shaping their operation, algorithms are malleable: they can be gamed, subverted, and contested by people affected by them. Citizens in an algorithmic society are not just passive users or objects of surveillance. A sociotechnical definition allows for a concept of agentic citizens who can use tech platforms in ways unintended by designers of algorithms and even give rise to forms of resistance that can “repair” and “correct the work of algorithms” to defend deliberative democracy (Velkova & Kaun, 2021, p. 523).

I will return to this conceptualisation in the latter part of this chapter and argue that deliberative democratic theory needs to reassert this view of agentic citizenry in an algorithmic society. In the next section, I first focus on how the scholarship on deliberative democracy has tended to emphasise the first aspect of the sociotechnical definition, which focuses on algorithmic power and epistemic authority.

2.3 Deliberative Democracy’s Concerns about Algorithms

In terms of public communication and political process, the triumph of robot sociality would bring about a public sphere defined by information jet streams and attention bubbles, ideological bandwagons, robomobs, and political herding. Devoid of critical reasoning and meaningful debate, politics will be shaped by the number—the number of “likes” and “dislikes” garnered from the human and nonhuman actors bustling on platforms. And the winner will always be the best team of social engineers. (Bakardjieva, 2015, p. 254)

The quote exemplifies a dominant perspective on how algorithms harm deliberative democracy. While Maria Bakardjieva (2015) was talking about democracy more broadly, her work emphasises how meaningful debate is subverted by a team of engineers following big
tech’s aim of gaining profit. Conceptually, they take the perspective that sees the production of algorithms in a monolithic manner, manufactured by a team of engineers following big tech’s corporate logic. The result is a distorted public sphere that is inhospitable to meaningful deliberation, specifically due to three algorithmic harms. First, algorithms compromise inclusiveness (i.e., “attention bubbles” and “political herding”). Second, algorithms undermine the epistemic quality of deliberation (i.e., “devoid of critical reasoning”). Finally, algorithms impair our capacities for internal reflection (i.e., “politics will be shaped by the number” and “the best team of social engineers”). I will discuss each of these harms in turn.

2.3.1 Algorithms Compromise Inclusiveness

The ideal public sphere is one where different voices come together as equals to exchange and reflect on their views. To be inclusive, all affected actors should have the opportunity to participate or have their interests and discourses represented in deliberations. As John Dryzek (2009, p.1382) puts it, “without inclusiveness, there may be deliberation, but not deliberative democracy.” Algorithms compromise inclusiveness by promoting discursive enclaves (Sunstein, 2017). They are designed to filter individual preferences and prioritise content that affirms one’s interests or views. This in turn, creates a “filter bubble” that diminishes our exposure to diverse and opposing views, which are necessary if citizens were to develop their capacity to listen and learn across difference (Bozdag & van den Hoven, 2015).

For example, Michael DeVito (2017, p.753) finds that story selection on Facebook is determined by “friend relationships, explicitly expressed user interests, prior user engagement, implicitly expressed user preferences, post age, platform priorities, page relationships, negatively expressed preferences, and content quality.” This has ramifications on how publics get together. The result of relevance filtering is not only a polarised society where citizens have a few opportunities to be exposed to different views (Sunstein, 2017); it also leads to the creation of “calculated publics.” Tarleton Gillespie (2014) introduces the term “calculated publics” to refer to how algorithms present a calculated approximation of a public which shapes the public’s sense of itself. This raises questions about who is excluded, misrepresented, and overrepresented in publics visible to us, and the power tech platforms have in cultivating these publics.
2.3.2 *Algorithms Compromise the Epistemic Quality of Public Deliberation*

Aside from concerns about inclusion, deliberative theory has also warned against the ways in which algorithms *undermine the epistemic quality of public deliberation*. Chambers (2021, p.153) provides an account of how algorithms—particularly those used to perpetuate misinformation—undermine deliberative democracy’s truth-tracking function (see also Curato et al., 2019). Drawing from the Habermasian tradition, Chambers (2021) argues that deliberative democracy’s core features—such as the equal status of citizens, an independent and free press, and cultures of pluralism and open debate—create conditions to test truth claims. A robust public sphere is one that can identify problems and interests, as well as hopes and dreams, which set the agenda for collective decision making.

The spread of misinformation undermines these features that allow democracies to track the truth, and the amplification of elite agenda compromises free public opinion formation. For example, state-sponsored troll armies or corporate-backed disinformation narratives can hack public attention by making topics “trending” while suppressing the emergence of critical arguments. In China, for example, fake accounts flooded social media after a riot in Xinjiang. These accounts served the purpose of distracting the public and changing the subject by posting about “the China dream” of national rejuvenation, economic prosperity, unity, and stability (King et al., 2017). Consequently, posts that demand the truth and accountability about the riots were effectively suppressed by drowning them out, which, one could argue, is a subtle form of censorship. Critics demanding for truth can still say what they want, but they will not be heard.

2.3.3 *Algorithms Compromise Internal Deliberation*

Finally, deliberative democrats warn against the *distorting effect of algorithms in our internal deliberations*. Ana Tanasoca (2019) begins her argument by reminding us that unlike voting, which is about counting heads, deliberation is about weighing arguments. The difference between counting and weighing, however, can be misleading. Half of web traffic, Tanasoca argues, is often attributed to bots, which raises concerns about how their messages foster a feeling of familiarity and commonality, thus giving citizens the impression that these views are widely held. Tanasoca puts forward proposals on how our internal deliberations can be secured from these threats, including removing clone claims from deliberative spaces and
reminding public officials to exercise an “epistemic duty of care” and to consider whether they pass on widespread information from unchecked sources.

How then can deliberative democracy respond to these challenges? Some scholars assert that the answer lies in algorithms themselves. For instance, Gordon Pennycook and David Rand (2019) argue that instead of relying on professional fact checkers to correct misinformation, they propose an algorithm that gives more visibility to trusted news sources online. Similarly, Timothy Graham and Robert Ackland (2017, pp. 192-196) make a case for “social bots” programmed to “pop the filter bubble.” These bots can “exploit” the networked structure of online communication to achieve the “social good” by promoting exposure to opposing positions and views (Graham & Ackland, 2017). Meanwhile, Cass Sunstein (2017) proposes an online space built on the virtues of deliberative democracy that moves away from social media. Whereas Nathan Gardels and Nicolas Berggruen (2019) suggest integrating deliberative democracy in social media algorithms.

Although these proposals seem promising, they have two shortcomings. First, these proposals oversimplify the democratic problems caused by algorithms. Algorithmic solutions to algorithmic harms reinforce technological solutionism, which is not immune from the same problems of opaqueness, secrecy, and concentration of epistemic authority among designers of algorithms. Second, these proposals limit the attempts at understanding the respective contexts where algorithms operate. For example, Xiaoxia Cao (2020) demonstrates that the aversion to exposure and engagement with political disagreement on social media is a conscious decision by citizens. In choosing to not engage, citizens make a deliberate choice as engagement can undermine their relationship with their friends, not because the algorithms do not recommend the content to them. Meanwhile, for Catherine Frost (2020), supplying “social bots” that burst filter bubbles still tarnishes deliberative democracy. Supplying more “social bots” is still a form of “ventriloquism” that not only manipulates the discourses in the public sphere by making some discourses more visible than others through mimicry, but can also drown out alternative voices of marginalised communities that demand a fair hearing. Frost’s critique emphasises the relevance of examining the contexts in which algorithms are created and where they interact with the practice of deliberative democracy. As I argue elsewhere, emancipation cannot be programmed (see Alnemr, 2020).
Thus far, I have characterised deliberative theory’s three concerns about algorithmic harms and provided a brief inventory of possible solutions to these harms. Although these are valid concerns, I find that the account of algorithms in this light is incomplete. For the most part, the critiques and solutions focus on the structural issues and impact of algorithms in democratic life, while the role and capacities of citizens as democratic actors are rather muted. In the next section, I argue that a more robust account of deliberative democracy in an algorithmic society needs to foreground the possibility for citizens to negotiate the power and impact of algorithms and defend the public sphere from algorithmic harms.

2.4 Contesting Algorithmic Harms in the Public Sphere

In the previous section, I argued against the use of algorithmic solutions to address algorithmic harms. I argued that technological solutionism comes at the expense of deliberative democracy’s potential to emancipate people from domination through inclusive reason giving. Even if developers have the best of intentions in crafting algorithms that promote democratic ends, these algorithms may still perpetuate domination if they are created, designed, and imposed on citizens and shielded from critique and public scrutiny.

How then can deliberative democracy be defended from algorithmic harms? The answer, I argue, rests on the extent to which the response to these harms build the polity’s deliberative capacity. Simply put, deliberative capacity refers to a polity’s ability to host inclusive, authentic, and consequential deliberations (Dryzek, 2009). Inclusiveness means that all people (or their representatives) affected by a decision (in this case, algorithmic decisions) should have a fair shot at influencing these decisions. Authenticity refers to the capacity of the polity to host a wide range of contesting discourses to generate collective decisions free from manipulation and coercion. The lack of transparency in the construction of algorithms, for example, undermines authenticity, because opaqueness does not invite alternative and critical discourses to challenge and revise algorithms. Finally, consequentiality refers to the effect of inclusive deliberations to the decisions made.

Using the concept of deliberative capacity in discussing algorithmic harms allows us to make sense of why some proposed solutions to address algorithmic harms can still compromise deliberative democracy. For example, relying on private companies, coders, and experts to
solve algorithmic harms does not build a polity’s deliberative capacity, if the power to influence the process of designing algorithms and determining their intended outcomes remains to be a privilege for a few. Algorithms, even if constructed to deliver democratic goods, can still be designed based on exclusionary instead of inclusive, and opaque instead of open deliberative decision making. This approach also presupposes that algorithms are all-powerful, such that they only need to be designed to condition citizen behaviour. As mentioned in the previous section, citizens too have their reasons for behaving a certain way in their digital lives, and it is this agentic quality of digital citizenship that should be fostered when addressing algorithmic harms.

I further develop this argument in this section by recasting the algorithmic harms discussed in the previous section to three ways of contesting algorithms in a way that builds a polity’s deliberative capacity: (1) contesting algorithms to promote inclusiveness, (2) contesting algorithms to promote epistemic quality, and (3) contesting algorithms to promote internal reflection. In each section, I put forward illustrative examples of how these contestations unfold in practice and highlight how these practices build deliberative capacity in an algorithmic society.

### 2.4.1 Contesting Algorithms to Promote Inclusiveness

Algorithms play the role of gatekeepers. Before a discourse makes an appearance on social media or anywhere on the internet, they first need to meet the “relevance” criteria of the algorithm used in each platform (Tufekci, 2015). This, for example, explains why the 2014 Ferguson protests went viral on Twitter but not on Facebook. A social media analytics company found that “stories about Ferguson and/or Michael Brown published since Aug 7 have generated Facebook referrals on average (256) than stories about the ice bucket challenge (2,106)” (SimpleReach in Naughton, 2014). For Facebook, what counts as “relevant” is not political dissent and activism, but light-hearted and joyful campaigns people can share to their friends, families, and filter bubbles like the ice bucket challenge.

One lesson from this comparison is that not all social media platforms are the same. The invisibility of discourses regarding the Ferguson protests on Facebook reminds us that civic activism cannot be left at the mercy of algorithmic relevance. To fight algorithmic
exclusion, social movements have understood the need to develop their own creative strategies in crafting messages that would meet the criteria of algorithmic relevance in order to gain visibility in a highly saturated digital public sphere (Blumler, 2018).

“Playing the algorithm” is one example of this strategy. Instead of treating the algorithm as an all-knowing and all-controlling feature of our digital lives, citizens have developed ways to subvert their logics to claim space for contestatory discourses in the public sphere. This approach is reminiscent of the disruptive tactics of activist groups, which, for some deliberative theorists, have the potential to enhance the inclusiveness and authenticity of the deliberative system (Curato, 2021; Mendonça & Ercan, 2015; Smith, 2016). The same tactics can be employed to challenge the algorithmic harm of exclusion.

Take the case of the #SaveSheikhJarrah campaign. Palestinian families living in Sheikh Jarrah, a neighbourhood in East Jerusalem, were protesting their forced displacement in May 2021. They created the hashtag #SaveSheikhJarrah to facilitate a global conversation about settler colonialism and call out Israel’s ethnic cleansing. Residents posted their stories on Instagram which generated attention not only of Palestinians around the world who shared the same experience of displacement but also international organisations monitoring the humanitarian situation in East Jerusalem.

As the campaign gained traction, some of the campaigners experienced censorship on social media platforms, such as Twitter and Instagram (see Hawari, 2021). Some posts were taken down, while other accounts were suspended for violating “community guidelines.” Digital rights groups protested this apparent exclusion of activist voices, which led Instagram and Twitter to reinstate the posts and accounts after being “suspended in error” by their automated systems (Gebeily, 2021).

Beyond protesting exclusions and “accidental” censorship on social media, pushing back also entails circumventing algorithms through everyday action. Many Palestinians, for example, use the watermelon emoji instead of using the emoji of the Palestinian flag. Watermelons are particularly symbolic because they have the same colours as the Palestinian flag and are grown in the northern Gaza Strip. Posting watermelon emojis was used to avoid Israeli surveillance and game the algorithms that take down posts that protest threatened
evictions in East Jerusalem (Berger, 2021; Gebeily, 2021). Aside from gaming the algorithm through emojis, #SaveSheikhJarrah campaigners also gamed algorithms through text. As journalist Mohamed Hamama describes it:

After weeks of complaints about the censorship enforced by social media platforms (especially Facebook and Instagram) on what Palestinians and their allies around the world publish about the violations of the Israeli occupation, many started to change the way they write to circumvent algorithmic censorship. Nobody understands the logic of this censorship, and this is why people took the effort of guessing how the algorithm works, and how algorithmic censorship can be circumvented. They altered some letters with symbols [e.g., P@lest-ine]. They added spaces between the letters of a given word [e.g., P.a.l.e.s.t.i.n.e]. In the last couple of days, it developed to removing dots from letters.¹⁴ (Hamama, 18 May 2021)

Hamama enumerated the ways in which citizens attempted to circumvent algorithmic censorship that show awareness of algorithms as well as the alternative provided by diversifying speech styles. Removing dots from letters is specific to the Arabic-speaking public sphere where citizens have found a leeway by using ancient Arabic script (“dotless”) to circumvent algorithmic censorship. This is a powerful example of citizens contesting algorithmic harms of exclusion. It shows an awareness of how algorithms technically work through trial and error. The algorithm curtails speech that contains words from a select dataset. But if the words do not match the dataset, the algorithm cannot detect it. This awareness is combined with different attempts at interpreting what censorship logics could be to avoid being censored in the future.¹⁵ The diversification of speech styles is an act of resisting the censorship logics regulating the public sphere. This form of contestation was driven by citizens’ political agency, as well as access to the technology that allowed Arabic speakers to write in a script that was last used centuries ago.

Beyond promoting inclusiveness in the public sphere, creative forms of citizen contestation also prompted a metadeliberation on the rules of content moderation by social media platforms. Instagram and Twitter were put in a position to justify their content

¹⁴ The article is written in pre-Islamic Arabic script which is “dotless.” The dots in modern Arabic script are diacritical but Hamama (2021) writes the article in dotless Arabic to demonstrate that even lengthier text than a tweet or a Facebook post is still accessible to Arabic readers despite the use of an archaic script. This excerpt is translated from dotless Arabic to English by the author.

¹⁵ An online tool is used to remove dots from Arabic script as mainstream software (e.g., virtual Arabic keyboards) as well as mainstream hardware (e.g., physical keyboards) not designed to type in dotless Arabic. This online tool can be accessed here: https://dotless.app/.
moderation practices and accept accountability for their mistakes. The #SaveSheikhJarrah campaign, therefore, is not just a campaign for inclusion. It also became a campaign that affirms a wider demand for social media to practice “radical transparency, meaningful accountability, and a commitment to remedy in order to protect the ability of individuals to use online platforms as forums for free expression, access to information and engagement in public life” (Kaye, 2018, p. 19).

This campaign built the deliberative capacity of citizens defending the residents of East Jerusalem, for they not only developed creative capacities to gain voice and visibility in public deliberations about forced displacement but also to contribute to a metadeliberation about the accountability of tech platforms in an algorithmic society. This is a step towards challenging the unjustified power of these platforms over speech and public deliberation, even when the conversation itself takes place through these platforms.

2.4.2 Contesting Algorithms to Promote Epistemic Quality

The second harm that warrants contestation is the poor epistemic quality of information propagated by algorithms. In the previous section, I characterised how algorithms can undermine deliberation’s truth-tracking function by disseminating misinformation or by drowning out the voices of users who demand truth telling from the state.

There are many ways of countering these algorithmic harms including fact-checking initiatives. While the jury is still out in terms of the effectiveness of these initiatives (Lazer et al., 2018), one could argue that the initiatives, in theory, could contribute to the health of a media ecosystem. From a deliberative perspective however, one could argue that these are not enough. One could even argue that an overreliance on fact-checking initiatives can perpetuate the view of citizens as passive consumers of information waiting for verified news items to reach their social media feeds. To effectively contest algorithms and promote epistemic quality, I argue that citizens need to take part in algorithmic scrutiny to build the polity’s deliberative capacity. Citizens can take an active role in deliberations about misinformation and, in some instances, even take part in deliberations about crafting algorithms that combat misinformation and participate in fact checking by providing sources of fact-checked content or by identifying sources that replicate misinformation.
Let me give the example of Cofacts, an open, collaborative platform created by a Taiwanese civic tech community called g0v (Splice, 2018). Through Cofacts, users who receive suspicious messages, such as medical misinformation or online scams from LINE (a popular messaging app in Taiwan), can forward the message to the Cofacts chatbot. Volunteer editors fact check these messages, and then send their findings to the user.

Cofacts promotes deliberative capacity in two ways. First, it promotes inclusiveness by welcoming anyone to become a volunteer editor. Cofacts encourages different editors to assess a message reporting misinformation and offer various interpretations. Although this may cause worries that Cofact may be held hostage by “rogue editors,” the platform maintained that the first priority is to maintain their identity as “an impartial third-party platform” (Davis et al., 2019, p. 172). To date, Cofacts’ founders reported that rogue editors have not been a major problem yet. Second, Cofacts promotes authenticity by encouraging users to deliberate internally about the outcomes of the fact check. An important feature of Cofacts is what happens when editors disagree on each other’s fact check. Cofacts has a feature where users will receive multiple messages and encourage the user to examine and judge varying information from Cofacts’ volunteer editors. Cofacts, of course, also uses bots to handle the scale of its operations. Chatbots are programmed to immediately respond to messages containing similar queries, so editors do not get inundated with answering the same questions. What sets Cofacts apart, however, is the direct participation of ordinary citizens in its fact-checking operations. Citizens are active agents of fact checking by either volunteering as editors or by actively deliberating on the outcomes of a “contested fact check.”

Aside from involving ordinary citizens in deliberations about misinformation, Cofacts also gives citizens the opportunity to deliberate on the construction and operation of algorithms if they wish to do so. The founders of Cofacts adhere to principles of open sourcing and made all their databases (which started as an Airtable spreadsheet) and codes publicly accessible, and therefore open to critique and scrutiny as well as replication. Cofacts, for example, has now been translated into the Thai context (Tseng, 2020).

The Cofacts model of fact checking builds the polity’s deliberative capacity by socialising citizens to critically engage in knowledge production, instead of relying on opaque operations of misinformation and counter-misinformation. The design, ethos, and
implementation of Cofacts may not be perfect, but it is set up to foster open-mindedness, deliberation, and truth seeking among its editors and users. Engaging algorithms in this manner promotes epistemic quality in a procedural sense. Deliberative theory considers deliberation to be useful in generating epistemically fruitful outcomes because it considers a range of Contesting Algorithms to Promote Internal Reflection

The last algorithmic harm to contest is that of the capacity of bots to manipulate the weight of certain arguments or information. Tanasoca (2019) unpacks the extent of this harm on our capacities for authentic and reflective deliberation, especially at the individual level. There are various ways in which bots are programmed to inflict this harm, from manipulating web traffic to the wide replication and repetition of a message on social media. How then may citizens contest this harm?

I argue that the condition of the public sphere being infused with, and regulated by, algorithms orients citizens to be “algorithm aware.” This awareness involves being critical of the information, discourse, and messages citizens are exposed to. Although research on the extent to which citizens are “algorithm aware” may not be definitive, it provides important insights. For instance, in Nordic countries, the level of algorithm awareness is directly proportional to existing socioeconomic divides (Gran et al., 2020). In comparison to Germany, citizens are algorithm aware regardless of socioeconomic and demographic markers but express this awareness differently (Dogruel et al., 2020). These differences can be attributable to the sensitivity of research methods used to assess the extent of citizens’ algorithm awareness or some contextual factors, such as the extent to which algorithmic control and harms is a topic of public concern covered in the news and media (Dogruel et al., 2020). Therefore, a blanket solution to the bot problem would not tap into these factors. Just as Tanasoca (2019) advocates for an “epistemic duty of care,” there is a duty to be “algorithm aware” as well.

Throughout this discussion, I have focused on the contributions of citizens to contest three algorithmic harms in the public sphere. However, the discussion does not undermine proposals to develop technical and institutional remedies. There is complementarity to be
sought in contesting algorithmic harms, and this is a starting point that highlights the capacities of citizens. In promoting inclusiveness, there was a collective effort to interpret the logics of algorithmic gatekeeping. In combating misinformation, developers and citizens collaborate to improve the epistemic quality of deliberation. Finally, on the problem of bots, citizens can also use tools for bot detection (which have their caveats) in cases where it is challenging or important to vindicate whether a message is circulated by bots or humans. Some of these bot detection open-source tools include Botometer (formerly known as “bot or not”), Bot Slayer, and Bot Sentinel. Ultimately, an algorithmic society is aware of the roles of algorithms and is building its capacity to be reflexive about improving its democratic quality.

2.5 Conclusion

This chapter examines how algorithmic harms can be contested to promote deliberative democracy in an algorithmic society. I argued that the key lies in recasting the portrayal of ordinary citizens not as pawns of algorithmic control, but as active citizens who can contest these harms and, in so doing, promote the polity’s deliberative capacity. I outlined the three algorithmic harms deliberative democrats identified, and recasted these harms into different ways in which they can promote inclusion, epistemic quality, and internal reflection.

While this chapter places emphasis on the contestatory power of everyday citizens, I also recognise that full democratic control of algorithms is difficult, if not impossible. The age of technocapitalism heavily skews power to big tech companies whose main function is to exploit creative digital labour and to commodify our personal data for profit (Fuchs 2018). While the political economy of the algorithmic society poses an uphill battle for deliberative democrats, there are still battles that can be won. The struggle over the meaning of algorithms and how they can be deliberatively produced are practical, everyday struggles that citizens can be a part of. One could hope that these small battles spark bigger ambitions on how algorithms can serve the processes and outcomes of deliberative democracy.

2.6 References


Chapter 2


Hamama, M. (2021, May 18). ُلْسَنَ صَدَرُ اللّهُ دِورَهُ النَّفْعَ النَّفْعِيَّ. Mada Masr. https://www.madamasr.com/ar/2021/05/18/feature/%d8%b3%d9%8a%d8%a7%d8%b3%d8%a9%d8%a7%d9%84%d8%a7%d9%84%d8%a7%d9%84%d9%87-%d9%ae%d9%88%d8%b1%d9%87-%d8%a7%d9%84%d9%ae%d9%af%d8%a7%d8%b7-%d8%a7%d9%84%d8%b9%d8%b1%d9%ae%d9%89%d9%87/


CHAPTER 3

Emancipation Cannot be Programmed: 
Blind Spots of Algorithmic Facilitation in Online Deliberation

Abstract

Challenges in attaining deliberative democratic ideals, such as inclusion, authenticity, and consequentiality in wider political systems, have driven the development of artificially designed citizen deliberation. These designed deliberations, however, are expert driven. Although they may achieve “deliberativeness,” their design and implementation are undemocratic and limit the emancipatory goals of deliberative democracy. This is relevant in respect to the role of facilitation. In online deliberation, algorithms and artificial actors replace the central role of human facilitators. The detachment of such designed settings from wider contexts is particularly troubling from a democratic perspective. Digital technologies in online deliberation are not developed in a manner consistent with democratic ideals and are not amenable to scrutiny by citizens. In this chapter I discuss the theoretical and the practical blind spots of algorithmic facilitation. I then present recommendations to democratise the design and implementation of online deliberation, with a focus on chatbots as facilitators.
3.1 Introduction

Algorithms, a key component to online communication, affects the democratic quality of intersubjective communication. Algorithms should then be evaluated by democratic ideals, as well as their technological effectiveness.

Algorithms automate steps necessary to complete different tasks (Goffey, 2008; Karpf, 2016). Their online presence varies from operating in the background, such as acting as gatekeepers in automated content moderation (Ruckenstein & Turunen, 2019) to automating actors who participate in communication (i.e., chatbots). The norms of human-to-human communication change with the inclusion of algorithms. In the case of content moderation, moderators highlight that substituting humans by algorithmic machines “transforms” the practice (Ruckenstein & Turunen, 2019). For different sectors and services, chatbots are becoming a favourable communication interface (Følstad & Brandtzæg, 2017). Hence, the role of algorithms in online communication and its effect needs to be examined.

Against the backdrop of this context, institutionally designed deliberative democratic processes (e.g., minipublics) have been experimenting with algorithms in online deliberation. John Dryzek (2007, p. 241) defines deliberation as communication that “induc[es] reflection, and [is] capable of connecting particular points and experiences to more general questions and principles.” The main premise of minipublics is that ideal deliberation cannot be realised unless conditions conducive to it are designed in the political setting. This design involves randomly selected ordinary citizens participating in facilitated deliberation about actions regarding collective issues (Fung, 2003). In this design, citizens realise the emancipatory objective of deliberation by reflecting on the “policies and institutions that are uncritically accepted” (Rostbøll, 2008, p. 728).

The same premise of design and deliberative quality applies online, with an expectation that online deliberation outperforms face-to-face minipublics. Digital technologies can address the weakness of face-to-face deliberation in terms of inclusion (Fung, 2003; Strandberg & Grönlund, 2018), such as replacing human facilitators by algorithms (e.g., chatbots) (Fishkin et al., 2019; Wyss & Beste, 2017). Using digital technologies in minipublics can potentially

---

16 Online deliberation is used in this paper in reference to forums designed to be deliberative (i.e., online minipublics).
aid the aspiration to “scale up” their deliberative effects.

Considering this intersection between algorithms, minipublics, and online deliberation, I problematise two aspects: the design and role of algorithms in facilitation, especially algorithms that substitute human facilitators. I argue that algorithmic facilitation exposes the break between how deliberative ideals are articulated in theory and the translation of these ideals in practice.

In respect to design, I critique online deliberation by considering the critiques of minipublics. It has been argued that deliberation’s emancipatory potential can only be realised in minipublics (Fishkin, 2003). To the extent that this is valid, design choices need to be carefully examined since choices can “limit a mini-public’s deliberative capacity” (Felicetti et al., 2016). In other words, expert-imposed design can undermine minipublics’ democratic and emancipatory potential since citizens are excluded from creating the conditions for deliberation.

Moreover, minipublics are criticised for their detachment from wider contexts and political systems (Hendriks, 2006; Michels & Binnema, 2019; Thompson, 2008; Zgiep, 2019). In online deliberation, this takes the form of accepting expert knowledge and decisions without scrutinising the conditions under which these were made. Subsequently, deliberative design becomes what it critiques: coercion of unjustified scientific and technological expertise (Lövbrand et al., 2011). These critiques are important in assessing the extent to which the processes of creating and implementing minipublics are emancipatory and democratic.

Although human facilitation is critical, there is a growing interest in replacing humans by algorithms. In spite of the crucial role of facilitators in maintaining deliberative ideals in minipublics (Fung, 2005; Mansbridge et al., 2006), such a role is underexamined in theory (Moore, 2012; Morrell, 2018). As a result, the role of algorithmic facilitation has blind spots as it neither engages with facilitation in theory nor relates to research on algorithms in communication. For instance, the role of facilitators involves exercising human discretion, reflection, and intersubjective interaction with participants—algorithmic facilitators cannot be programmed to fulfil all these expectations.
Chapter 3

To flesh out this argument, the first section of this chapter examines emancipation in deliberative democracy, the design of minipublics, and the role of facilitation. The second section defines programmed deliberation (i.e., online deliberative design which subjects citizens to predetermined norms of deliberation as understood by experts) with a focus on algorithmic facilitation and artificial facilitators (AFs). Two AFs replacing humans are presented in this section: artificial discussion facilitation agent (DiFA) (in Stromer-Galley et al., 2012) and Sophie (in Wyss, 2018; Wyss & Beste, 2017). Drawing on critiques of minipublics, the third section highlights blind spots in programmed deliberation and their theoretical and practical implications. In the fourth section, I make recommendations to democratise the design and implementation of online deliberation.

3.2 Emancipation, Minipublics Design, and Facilitation

Deliberation in deliberative democratic theory is a unique mode of communication with others (Dryzek, 2007, 2010; Hendriks, 2006; Mansbridge et al., 2012). The normative ideal is inclusive, authentic, noncoercive, and consequential deliberation (Dryzek, 2010). These ideals are based on intersubjectivity between citizens and other actors, such as engagements in reason giving and active listening (Bächtiger & Parkinson, 2019, p. 20; Rostbøll, 2008).

Reason giving and active listening induce reflection, which is central to emancipation. Reflection builds the capacity to challenge “unreflective acquiesces” (Rostbøll, 2008) and is underpinned by a respect for citizens’ autonomy, capability, and authorship, thus making deliberation a path in securing the “right to justification” (Curato et al., 2019). Because of this, the purpose of deliberation is not to “impose anything on anyone but … encourage[e] and provoke[e] processes of self-reflection [towards] uncritically accepted beliefs, and institutions” (Rostbøll in Steiner, 2012, p. 92).

Since contemporary political settings are hardly conducive for reflection, an artificially designed setting is an alternative in reconciling the ideals and the practice of deliberation (Fishkin, 2003). Facilitators have an important role in maintaining this design (Epstein & Lesheid, 2016), and through design and facilitation, the emancipatory potential of minipublics emerges from “communicating the reasoning that underlies a deliberative outcome [which]
empowers citizens to make their own evaluations of the mini-public’s decisions” (Niemeyer, 2011, p. 127).

The question of design is no less important for online deliberation, whereby platforms are designed to uphold deliberative ideals (Strandberg & Grönlund, 2018). Esau, Friess, and Eilders (2017) find that platform design impacts citizen interaction and the quality of online deliberation. For example, designs affect reciprocal listening and collective decision making (Davies & Chandler, 2012).

Either face-to-face or online, facilitators steer minipublics where they balance the normative demands of deliberation and professional performance benchmarks.

3.2.1 Facilitation: Key to Deliberative Design

Facilitators are trained to maintain deliberative ideals in minipublics. This includes enforcing the norms of equality, civility, rationality, and reflexivity among citizens (Gastil & Black, 2007; Min, 2007; Offe, 2011). In relation to equality, good facilitation eliminates coercion from external structural injustices where facilitators “bracket substantial inequalities of knowledge, influence, and communicative capacity” (Fung, 2005, p. 413). By enforcing the other norms, facilitators transform everyday political talk into objective-oriented deliberative exchanges which present judicious arguments based on critical listening and reason giving (Dillard, 2013).

Empirical evidence shows that deliberative design is effective in achieving deliberativeness (Strandberg et al., 2019). Such can be attributed to the role of facilitators as socialising agents who introduce and enforce deliberative norms to groups (Jennstål, 2019). This involves exercising discretion to ensure nondomination and to promote collective and deliberative listening, reasoning, and equality (Black in Dillard, 2013; Fung, 2005; Jennstål, 2019; Min, 2007; Morrell, 2018).

By enforcing deliberative ideals in minipublics, human facilitators orient citizens to reflect on collective outcomes. The deliberative function of facilitators is to “drive participant reflection about common ground” (Dillard, 2013, p. 219), which is achieved through active reason giving and justification (Fung & Wright, 2001). Such a deliberative exchange allows
citizens to reflect on strongly held or polarized political views (Strandberg et al., 2019). In sum, facilitators ideally create conditions that activate “civic-mindedness” in order for participants to develop a deeper understanding of the subject matter of deliberation (Epstein & Leshed, 2016).

The demand for minipublics and the centrality of facilitators encouraged a professionalisation of facilitation. As experts of deliberation, facilitators make trade-offs in balancing deliberation and the profession.

### 3.2.2 Minipublics on Demand

A growing interest in minipublics as effective citizen engagement processes (Gherghina et al., 2019) should be complemented with an understanding of the politics in these deliberative events (Prosser et al., 2018). Challenges encountered by human facilitators in light of these politics provided a window to accentuate the advantages of algorithmic facilitation. I focus on two challenges: agenda setting and professionalisation.

In agenda setting, facilitators are challenged to maintain the ideal of authenticity. When a minipublic event is government sponsored, facilitators are bound by the agenda that is set from above (Morrell, 2018). Without engaging in reason giving on the part of the agenda setters to citizens, authenticity is constrained. Recruiting independent facilitators is one design choice that can treat this problem (Curato et al., 2017). Alternatively, as will be examined later, it is an opportunity to explore the potential of neutral and unbiased AFs (Stromer-Galley et al., 2012; Wyss, 2018; Wyss & Beste, 2017).

The second challenge is professionalisation and its impact on deliberation. A demand for minipublics creates a “market” for deliberative events which requires recruiting, training, and hiring facilitators (Hendriks & Lees-Marshment, 2019; Prosser et al., 2018). For some scholars, this makes facilitators professionals (Moore, 2012; Sandfort & Quick, 2017), while others see facilitation as a role with some “semblance” of a profession (Christensen, 2018).

As professionals, facilitators are evaluated based on benchmarks irrelevant to deliberation. To secure future contracts, facilitators are incentivised to make deliberation an enjoyable, rather than a politically transformative experience. As a result, deliberation ends
when it meets performance indicators, instead of reaching a collective political outcome even though this is the intended purpose of deliberation (Mansbridge et al., 2006; Moore, 2012).

Through professionalisation, some aspects of facilitation are systemised and can be automated, particularly because the premise of minipublics is that challenges to achieve ideal deliberation are mitigated by design choices (Moore, 2012). For example, facilitation toolkits can be curated to inform facilitators how to combat deliberation pathologies (Fung, 2005). In practice, this makes facilitation inconsistent across minipublics due to differences between skilled and unskilled facilitators (Prosser et al., 2018). Nonetheless, the systemisation of design and facilitation makes it possible to develop algorithms in online deliberation.

The expert-driven design and implementation of minipublics do not allow citizens to shape deliberation. Citizens do not have the space to contest the rules and the agenda in minipublics, which only a few facilitators are aware of its importance (Mansbridge et al., 2006; Morrell, 2018). In allowing this space for citizens, Simon Niemeyer (2019) demonstrates its effect on increasing the level of group reasoning.

Due to the absence of this space in shaping the rules of deliberation, facilitators can enforce an interpretation of deliberative norms inconsistent with deliberative ideals. Consider the example of enforcing “civility” where human facilitators can audit citizens’ contributions based on the facilitators’ understanding of what is “civil.” Such auditing includes filtering “uncivil” comments from publication, prioritising well-written comments, or editing comments with spelling mistakes or long stories before they are published (Epstein & Leshed, 2016). While facilitators herein enforce civility, they undermine inclusion by distorting contributions. This interpretation can be translated into an algorithm to automate content filtering and editing.

Although algorithms can automate various tasks and roles in online deliberation, there remains a substantive difference between human and algorithmic facilitators. Humans can exercise human discretion in deliberation either to bracket structural inequality, ensure inclusion, or enforce an interpretation of deliberative norms, as in the example of “civility.” The difference is not in exercising discretion itself but the ability of humans to explain, justify, and reflect on their discretion. By contrast, algorithms that automate “civility” would not provide an explanation as to how “civility” was interpreted. The expert-imposed design also
restricts the possibility of citizens and participants to reflect on and update this interpretation. Overall, this can be attributed to the main limitation of minipublics design, which is to exclude citizens from shaping deliberation.

The next section discusses the types of algorithmic facilitation in online deliberative design.

### 3.3 Programming Deliberation: Algorithms in Online Deliberative Design

A departure from the verbal expression and the physical presence associated with face-to-face minipublics did not deter deliberative research from examining the potential of digital technology. Experimental designs of online deliberation suggest that technology has positive prospects (Price, 2006). With even further technological advancements, algorithms, automation and artificial intelligence (AI) can be employed to address deliberation pathologies specific to online communication, such as motivated reasoning and the lack of argumentation (Wyss, 2018; Wyss & Beste, 2017).

Overcoming the limitations of face-to-face minipublics has been an interest in designing online deliberation. Tools and platforms for online collective collaboration have existed before it was used for political communication and participation (Gangadharan, 2009). This availability, nevertheless, is problematic because online deliberation design is “outsourced” to the relevant experts (De Cindio & Peraboni, 2010). The norms of deliberation are thus not based on the knowledge and experiences of skilled deliberation facilitators. Instead, these norms are programmed based on the interpretations of platform developers, which are then translated into algorithms.

The justifications for algorithmic facilitation are compelling in terms of their contribution in expanding the profession of minipublics but weak in terms of their democratic quality. According to Eric Gordon and Stephen Walter (2019, p. 315), “technology that focuses only on designing more efficiency into civic life runs the risk of altering, without deliberation or oversight, the very constitution of what a citizen is and what they can or cannot do.” Considering emancipation, this shortcoming affects citizens’ opportunity to discern the role of technology in democratic and political participation.
There are two types of algorithmic facilitation in online deliberation. The first operates in the background, organising communication that supports human facilitators. The second type replaces human facilitation and participates in deliberation via communicative contributions.

### 3.3.1 Algorithms: Supporting and Substituting Human Facilitators

The division of labour between humans and algorithms, automation, and AI does not always involve the replacement of humans by machines. Amitai Etzioni and Oren Etzioni (2017) demonstrate that there are two types of AI: partners and minds. As partners, AI actors support and assist humans in different tasks to give humans more time for demanding tasks which AI cannot (or does not) perform. Minds, on the other hand, are developed to replace humans in different roles (Etzioni & Etzioni, 2017). In online deliberation, algorithms are created in light of these two types.

As partners, algorithms automate some aspects of deliberation to assist facilitators and participants. Algorithms automate tasks that human facilitators should not be occupied with, such as grouping contributions in large-scale deliberations (Klein, 2012), deleting repetitive contributions (Epstein & Leshed, 2016), and censoring profanity (Wright & Street, 2007). In spite of this, when software is programmed to privilege certain forms of interaction and expression, algorithms may hinder deliberative inclusiveness (Kavanaugh et al., 2005).

Another form of algorithmic support for participants and facilitators is through enhancing the epistemic and cognitive aspects of deliberation. Participants can be supported by transcribing verbal communication via speech-to-text technology, making editorial recommendations regarding word choice, grammar, and titles, and summarizing content for participants to read (Davies & Chandler, 2012; Towne & Herbsleb, 2012). Facilitators can be supported by machine learning technologies to identify contributions which do not qualify to be shared with policymakers (e.g., when contributions lack substantial evidence). Because these algorithms are partners, they only flag the contributions. The subsequent action is up to the facilitators’ discretion (Epstein & Leshed, 2016).

Empirical insight on what facilitators think of this division of labour is lacking. There are aspects in deliberation, as previously discussed, that require human discretion and cannot
be programmed (e.g., transforming everyday political talk, orienting participants to be civic-minded, and inciting reflection on the common good). To draw on the experience of content moderators in commercial platforms, they “express an ideal division of labor that delegates the tedious cleaning work to machines and frees moderators for more challenging work, consisting, for instance of educating online communities and training and supervising machines” (Ruckenstein & Turunen, 2019, 1039).

As minds, algorithms replace human facilitators. Algorithms allow artificial actors to make autonomous decisions that simulate “specific types of [human] intelligent behaviour” (Gangadharan, 2009, p. 334) such as initiating the discussion, informing the group, being a conflict resolver, or playing devil’s advocate (Davies & Chandler, 2012). When self-facilitation is the normative end, an automated reward system can assist in delivering this end. Algorithms will remain in the background, making autonomous decisions regarding the selection of incentives and task allocation to citizens (e.g., moderation or initial appraisal) (Velikanov, 2010). Even though this is an intelligent system, Cyril Velikanov (2010) recommends that human facilitators should not be completely out of the loop and step in as arbiters when participants encounter an impasse.

Unlike the system described by Velikanov (2010), the two AFs, DiFA and Sophie, operated in online deliberation without human facilitators.

3.3.2 Artificial Facilitators in Online Deliberation

The goal of artificial facilitation is to enhance the normative contribution and the instrumental competitiveness of online deliberation. This section describes the roles and contributions of the two AFs, DiFA and Sophie, in light of this goal. Although DiFA and Sophie both replace humans, each interface is different. DiFA is similar to a curated frequently asked webpage whereas Sophie is an AI chatbot. Overall, the features of these AFs are similar in terms of their normative and instrumental potential.
DiFA and Sophie facilitated online deliberations on three governance issues. DiFA facilitated deliberation about network neutrality governance. Jennifer Stromer-Galley and colleagues (2012) faced institutional and political challenges in approaching federal agencies in the US to host online deliberations, but were able to engage DiFA with 184 participants in experimental deliberation on e-rulemaking. Sophie joined citizens in field experiments on deliberation about complex issues such as the Enforcement Initiative in Switzerland that called for the automatic expulsion of foreigners committing crimes, such as robbery and rape (Wyss, 2018; Wyss & Beste, 2017) and democratic governance in Germany (Wyss, 2018).

To facilitate deliberation about these complex issues, each AF employs different technologies. The AI in Sophie allows her to make context-sensitive interventions from 60 intervention phrases (Wyss & Beste, 2017; Wyss, 2018). Meanwhile, DiFA uses natural language processing technology and a dialogue act classifier algorithm. These technologies enable DiFA to inform participants of the relevant rules, to connect participants with similar and opposing views, and to suggest new topics (Stromer-Galley et al., 2012). The role of technology in online deliberation based on these AFs shows the dispensability of humans.

Where the normative contribution and the instrumental value of online deliberation meet, AFs are promoted for their capacity to overcome limitations of human facilitators. For instance, while inclusion is the normative goal, prospects for scaling up deliberation is the instrument. Hence, to increase citizen participation in rulemaking across federal agencies in the US, Stromer-Galley and colleagues (2012) designed DiFA as a cost-effective alternative suited to government budgets. As for Sophie, an AI chatbot is restless and neutral unlike humans. Sophie is also able to address a deliberation pathology, such as motivated reasoning, through uninterrupted and unbiased feedback loops that should increase argumentative exchange (in Wyss, 2018; Wyss & Beste, 2017).

Beyond the limitations of humans, AFs are designed to have epistemic, cognitive, and normative contributions in online deliberation.

\[17\text{ Since agencies are mostly familiar with public comments, they had difficulties in categorizing deliberative exchanges in their documentation. Other agencies sought deliberation as a process to dictate a particular outcome on participants which goes against the purpose of deliberation.} \]
Epistemically, one indicator of good minipublics design is informing citizens about policy issues (Fung, 2003). This is consistent with what Maeve Cooke (2000) categorises as an “epistemic proceduralist” view of deliberation which ties the epistemic quality of deliberation to the procedure.

For DiFA, the epistemic procedure depends on the task. In order to inform participants on rulemaking and to answer questions of poorly informed or unsophisticated participants, DiFA retrieves answers from specifically curated information. DiFA responds to questions with either a “factoid answer or more typically a short explanatory paragraph extracted from one or more relevant documents” (Stromer-Galley et al., 2012, p. 86). As for the participants, the platform keeps them informed and engaged through automated daily email updates and a word cloud visualising highlights from the discussions (Stromer-Galley et al., 2012).

Similarly, knowledge gain was assigned as a desirable outcome of deliberation, which Sophie was designed to facilitate but failed to influence. Dominik Wyss and Simon Beste (2017) deduce two factors that affect this. First, citizens are more inclined to fear “social sanction,” thus constraining their ability to form a political opinion. This rationale is supported by findings from the control group (not facilitated by Sophie) which likewise did not achieve knowledge gain. Second, personal traits such as self-assertiveness influence the effectiveness of the AF’s intended purpose (Wyss & Beste, 2017). The interaction between personality traits and the effectiveness of AFs is a blind spot that will be discussed later.

Enhancing cognition is Sophie’s main objective. Sophie’s design is based on the “argumentative theory of reasoning” which is also used for a cognitive interpretation of deliberation (i.e., a process of collective reasoning reliant on argumentative exchange). This exchange is largely lacking in online discussions resulting in motivated reasoning or opinion polarisation (Wyss & Beste, 2017). Deliberativeness is operationalised by the length of contributions where length reflects argumentative exchange. Hence, the technological intervention is to create customized feedback loops that stimulate argumentative exchange between participants, an objective Sophie did meet (Wyss, 2018; Wyss & Beste, 2017). By contrast, the group which was not facilitated by Sophie exhibited less argumentative exchanges and more polarised views (Wyss, 2018).
Normatively, these AFs were designed to observe the ideal of inclusion. To facilitate inclusion, DiFA follows an algorithm that connects participants who are in agreement and/or disagreement (Muhlberger et al., 2012, 2008; Stromer-Galley et al., 2012). Following the same algorithm for inclusion, AI allows Sophie to make context-sensitive interventions. For instance, Sophie would ask citizens to justify their position or state whether they agree or disagree with others (Wyss & Beste, 2017).

Overall, the prospects of AFs in enhancing epistemic, cognitive, and normative aspects of online deliberation are promising. Nonetheless, these prospects have blind spots attributed to the expert-imposed logics in designing and maintaining online deliberation.

3.4 Blind Spots of Programmed Deliberation: Theoretical and Practical Implications

Professionalising minipublics systematised aspects of deliberation that can be automated, thus increasing the potential for scaling up deliberation. AFs are also immune to challenges human facilitators face, such as agenda setting and performance benchmarks. The main limitation of programmed deliberation, however, is the extent to which its design and maintenance are undemocratic. Programmed deliberation does not align with emancipation in deliberative democracy.

By focusing on the prospects of algorithmic facilitation, expert-imposed online deliberation design has blind spots of theoretical and practical implications on deliberation. These blind spots exist in the role of algorithms as partners (supporting humans) and minds (substituting humans) in deliberation.

3.4.1 Theoretical Implications of Algorithmic Facilitation

To examine the theoretical implications, I focus on the effect of algorithmic facilitation on the ideals of inclusive, authentic, and noncoercive deliberation (Dryzek, 2009, 2010), particularly how algorithms in online communication challenge the anthropocentricity of these ideals.

Algorithm in this context refers to the logic, not a reference to the algorithm scripted for Sophie per se.
As discussed earlier, if the promise of technology is to advance inclusion, who is then included in algorithmic facilitation? One angle of this question relates to the ability of AFs to include participants. Sophie was unable to “perceive” or interact with participants who did not contribute to the discussion (Wyss & Beste, 2017). This shortcoming stems from an expert’s definition of inclusion, which does not account for silence in deliberation, an intersubjective interaction that cannot be programmed.

The intersubjective character of deliberation is based on a normative understanding of equal moral worth of participants in deliberation (Gutmann & Thompson, 2004). In this respect, it is the “opportunity and ability of all affected actors (or their representatives) to participate” (Dryzek, 2010, p. 11), including “all those subjected” (Erman, 2012). Because of the autonomy and equality of all participating actors, reciprocity is an expression of acknowledging and respecting others (Gutmann & Thompson, 2004). Therefore, deliberative inclusion is not only in presence but in the articulation and evaluation of the contributions of oneself and the others’ (Curato & Ong, 2015, p. 201). Consequently, an AF like Sophie undermines inclusion by not “perceiving” participants.

In other regards, the expert-imposed definition of effective technological intervention alters the theoretical assessment of speech acts and inclusion. For example, the success of Sophie’s intervention is based on operationalising deliberativeness as increased contribution length (Wyss & Beste, 2017). This operationalisation runs the risk of dismissing plural forms of communication used in deliberation. Yet generally, this operationalisation is informed by a move in digital communication to “rationalise” speech (i.e., reduce it to measurable units) (Papsdorf, 2015).

Another angle in inclusion concerns the AF itself. Communication machines such as chatbots are extensions of their creators’ agency (Mackenzie, 2006). As an extension of other humans, the inclusion of chatbots provokes questions about “what or who is actively and/or passively thinking and speaking on our behalf” (Morelli, 2019, p. 189). Because chatbots like Sophie are created and developed in a technology culture that is male dominated, their speech acts are not always relevant to different societal groups (Folstad & Brandtzæg, 2017). The theoretical challenge then is to discern when quantitative inclusion offsets the substantive exclusions algorithmic facilitation creates.
To be included in settings like deliberation, artificial humanness is necessary for AFs. This involves programming a combination of social, cultural, and psychological dimensions into a chatbot’s personality. In online deliberation, Sophie had to have a “face” through an avatar, which Wyss (2018) explains is crucial for Sophie’s objective. An avatar is crucial because artificial actors need human-like appearance and behaviour so that human interlocutors enforce social norms and mores (Nass et al., 1994; Nass & Moon, 2000; Sproull et al., 1996).

Different chatbot personalities trigger different responses from humans. Wyss and Beste (2017) underlined that Sophie’s effective intervention depended on the personality of human deliberators. However, this fails to examine the interaction between the personality programmed for Sophie and the personalities of deliberators. For example, if a chatbot is programmed to be “too human” due to having a neurotic personality or responding slowly, the bot is likely to be unsuccessful in gauging human interaction (Holtgraves et al., 2007). Nevertheless, such personality traits and subsequent reactions from humans are not universal and depend on the cultural norms of communication across societies (Lucas et al., 2018). There is no definite answer as to which personality traits contribute to successful human-machine communication. Herein, the theoretical challenge is factoring normative standards in designing AFs to accommodate diverse speech acts and culturally sensitive personality traits.

Artificial humanness also affects authenticity. Although it lends AFs humanlike appearance and behaviour, artificial humanness does not make them human enough to replace human intersubjectivity. Authenticity is reflection that connects particular views to “general principles” (Dryzek, 2009). However, arriving at a collective-regarding decision cannot be programmed. On the one hand, algorithmic rationale steering human-machine communication impairs the ability of humans to “make ethical decisions” (Martin, 2019, p. 847). On the other hand, even if artificial actors appear human-like, they are still nonhuman actors incapable of moral judgement (Friedman, 1997).

As far as active listening is concerned, AFs are able to simulate but not engage in the same intersubjective listening as humans. Although automating communicative labour is possible, it changes expectations from communication-based roles such as listening, that is, listening is lost and “stories fall on deaf ears” (Reeves, 2016; Turkle in Reeves, 2016, p. 155).
Programmed deliberation herein reproduces a context where the role of technology is unchallenged and the ideals of deliberation are unmet.

Excluding citizens and participants from designing and maintaining online and offline minipublics limits the ideal of noncoercion. Forms of coercion include domination through “the exercise of power, manipulation, … and the imposition of ideological conformity” (Dryzek, 2003, p. 3). Yet as the deliberation is predetermined, citizens and participants conform. Negotiating the “rules of engagement” by facilitators and participants contributes to better deliberation (Hartz-Karp & Sullivan, 2014; Niemeyer, 2019). However, AFs do not accommodate this as they cannot justify such rules and cannot reflect to update them. Addressing this limitation is overlooked and, in turn, programmed deliberation hinders noncoercion because citizens passively accept how the deliberation is designed and maintained.

Alternatively, algorithmic facilitation can overcome the coercion induced by human manipulation where algorithmic neutrality contrasts human bias. In face-to-face minipublics, coercion comes from the power human facilitators have in swaying the understanding of citizens when facilitators are too informed or opinionated (Polletta & Gardner, 2018). AFs, by comparison, are neutral and unbiased (Wyss & Beste, 2017). This algorithmic neutrality, nevertheless, is questionable in light of the increasing evidence that algorithms automate human biases (see Garcia, 2016; Lee, 2018). Therefore, the assumption that algorithmic neutrality counteracts manipulative human bias calls for revisiting the role of neutrality and bias in theory.

In short, the theoretical refinement required to discern the role of algorithms in online deliberation concerns the normative expectations of algorithmic actors, especially regarding the changes to intersubjective deliberation when nonhuman actors are involved.

### 3.4.2 Practical Implications of Algorithmic Facilitation

I identify three practical implications on authorship, inclusion, and intersubjectivity in deliberation with algorithms—in each of these, there is a break between deliberation in theory and in practice. Authorship concerns the extent to which citizens are included in designing and maintaining deliberation. Under inclusion, I provide examples of the restrictions programmed
Blind Spots of Algorithmic Facilitation in Online Deliberation

deliberation has on fostering inclusiveness. *Intersubjectivity* highlights the limitation of algorithmic facilitation in deliberation about complex issues.

Authorship is a component of emancipation, as mentioned earlier, undermined by expert-imposed minipublics. Marit Böker (2017) illustrates that in some cases of minipublics, organisers perceived their minipublics as neutral and well-designed, whereas citizens did not share the same opinion. When deliberation is programmed and human facilitators are absent, citizens do not have an opportunity to scrutinise the information provided in deliberation. In the example of DiFA and curated information, how can citizens judge if the information provided is not biased or skewed? Critical reflection emancipates citizens from manipulation and being “passive receivers of information or inevitable pawns of hegemonic forces” (Dahlberg, 2001; Kadlec & Friedman, 2007). It is the case that information provided in minipublics are carefully curated. However, without reflecting on this curation itself, minipublics undermine citizens’ authorship.

Another facet in undermining authorship concerns algorithms. The main developer of an algorithm is the only one capable of identifying faults in a system (Wilensky, 1983). A degree of transparency regarding the algorithmic models used for automation is therefore important so that different societal actors can judge if it is “morally just or mathematically sound” (Karpf, 2016, p. 54). In the case of DiFA and Sophie, such transparency and room for authorship was not provided.

Programming a particular interpretation of deliberative norms impacts inclusiveness as it shapes citizens’ contributions. To be inclusive, deliberation should accommodate different forms of communication (Hartz-Karp & Sullivan, 2014). Expression is however constrained in programmed deliberation when, for example, algorithms correct grammar and word choice as mentioned by Todd Davies and Reid Chandler (2012). Hence, algorithms enforce an order which “affect[es] what can be said and done” (Mackenzie, 2006, p. 44).

Experts who design algorithms make decisions on what qualifies as a good speech act (e.g., good grammar and succinct contributions). However, what is the impact of enforcing a preference for neat grammar and particular word choices on deliberative communication? Building on the argument of André Bächtiger and John Parkinson (2019) that deliberation has
the capacity to invoke different and creative uses of language in communication, automation that is focused on a preferable language constrains this capacity.

Algorithmically limiting what can be said hinders unique forms of expression and communication such as storytelling. When one story is made, it is reciprocated with another. The give-and-take of stories is a speech act that highlights differences in understanding between deliberators in an unhostile fashion (Polletta et al., 2011). Moreover, stories activate listening to views and interests that would have been otherwise dismissed (e.g., minorities) and also provoke reflection on taken-for-granted views (Polletta & Lee, 2006). As programmers depart from understanding the role of stories and automate the editing of long stories, the role of stories in listening and reflection is compromised.

Another form of algorithmic editing that affects inclusion is automating word choice. Programmers and developers perceive their technology as value-free, despite making choices in technology which shape “human activity” (Friedman, 1997). When algorithms intervene to correct grammar and spelling, it privileges certain forms of communication and excludes others. While creative communication, such as storytelling, makes language malleable to serve one’s contribution and narrative, automatic adjustments and auditing restrict this malleability. For example, Aboriginal Australian activists use “blak” to identify themselves (Watego, n.d.), thereby serving a contextual purpose despite the incorrect spelling of the word. Therefore, assessing the deliberative quality of online communication should depend on two aspects: the extent to which online communication can be plural and unrestricted, and the extent to which it can further deepen existing deliberation in society (Kies, 2010).

Deliberative norms emerging from intersubjective interaction cannot be expected to emerge in algorithmic facilitation. These norms emerge from the interplay between citizens, context, and setting and facilitator’s choices, training, and tools related to deliberation (Sandfort & Quick, 2017). In deliberating with an AF like Sophie about a complex issue such as the Enforcement Initiative, it is hard to test how effective 60 intervention phrases are. The context of the issue requires knowledge and experience related to the cleavages of immigration politics in Switzerland that could not be summarised in 60 phrases. Kara Dillard (2013) emphasises that the content and justifications of a deliberation depend on the facilitators’ interventions. However, in assessing Sophie’s performance, the AF is evaluated by its technical
rather than substantive role (i.e., increasing participants’ contribution length) (Wyss, 2018; Wyss & Beste, 2017).

In theory and practice, deliberation is about inclusion, noncoercion, and intersubjectivity. There are breaks nonetheless between theory and practice as practice excludes citizens from shaping the conditions for deliberation. In online deliberation, the breaks are also specific to the role of algorithms, which is neither subject to scrutiny in theory nor citizen deliberation in practice.

3.5 Recommendations for Democratic Online Deliberation

In the previous section, I established two critiques in light of replacing human facilitators by algorithms: while the top-down design of minipublics is undemocratic, so too are technological solutions in online deliberation. The combined potential of good minipublics design and algorithmic facilitation are used to justify prospects of enhancing the “deliberativeness” of online deliberation. However, online deliberation falls short in observing the democratic side of deliberative democracy. In other words, although the practice (according to its own definition of deliberation) can be deliberative, it is undemocratic because experts dictate the process.

In the following section, I present recommendations to democratise the design and implementation of online deliberation. But before turning to these recommendations, I contextualise the factors behind undemocratic technology and minipublics design.

Technologies raise democratic questions which are not attempted in online deliberation. Despite the fact that democratic societies depend on technology, the development of technology is rarely embedded in democratic principles (Fischer, 1999). To address this tension, assessments of digital technologies need to uproot “technological instrumentality” (i.e., the assumption that technology is inherently democratising) (Ess, 2018). This is where citizen deliberation becomes important as it scrutinises expert knowledge and decisions in technology, before its development and through its application, as Frank Fischer (1999) illuminates. Instead of being a space for emancipatory reflection on the role of technology in society, programmed deliberation carries technology uncritically.
Such an uncritical acceptance of technology is symptomatic of prioritising design in minipublics over their contribution to wider political contexts. The end of minipublics should not be perfecting design (Böker, 2016) because focusing on the design limits the consequentiality of deliberation to the lifespan of minipublics. Instead, minipublics should serve as spaces for practicing deliberation and critical reflection for engagement in the public sphere (Böker, 2017; Chambers, 2009). Yet this opportunity is untapped in online deliberation vis-à-vis the role of technology.

In pursuit of scaling up online deliberation, it is programmed with no regard to context sensitivity or citizen input. Minipublics are critiqued for negating citizens’ input into its organisation and management (Böker, 2017), similar to online deliberation (Towne & Herbsleb, 2012). While citizen inclusion in design is ideal, time constraints do not allow it (Farrell et al., 2019). Hence, online deliberation becomes a technical solution rather than a democratic engagement process. As a technical solution, online deliberation is a software designed to deliver “a tailored communication plan,” which is marketed for its benefits to the users (e.g., employees of the institution hosting the deliberation) (C. W. Lee et al., 2015).

Favouring cost, time, and scale over emancipation and citizen engagement in online deliberation has two democratic limitations. The first is moulding democratic engagement based on what the technology offers. Focusing on technological efficiency limits critical reflection about technology as a tool for democracy (Schlosberg & Dryzek, 2002). Because of virtual interaction, some values gained in face-to-face deliberation (e.g., empathy generated by the physical presence of others) do not apply (Loveland & Popescu, 2011). Herein lies the second limitation wherein the transformative interactions of online deliberation, from the citizens’ perspective, are unidentified. Overall, because the focus of online deliberation is the effectiveness of technology and good design, citizens’ insight on the presence or absence of deliberative norms is overlooked (Janssen & Kies, 2005).

It is possible to call for a trade-off between democratic quality and the scale of deliberation. Nicole Rishel (2011) recognised this trade-off with the advent of social media.19

---

19 Rishel (2011) does not refer to commercial social media platforms in particular. In the discussion, Rishel (2011, p. 418) uses a definition provided by the U.S. Federal Web Managers Council explaining “social media as the various activities that integrate technology, social interaction, and content creation.”
Whereas online spaces can be quantitatively inclusive, it may alter deliberative democracy with the use of new “tools in deliberative processes” (Rishel, 2011). Similar alterations are observed in replacing human facilitators by algorithms. Moreover, and as mentioned earlier, with this trade-off, minipublics resemble practices that deliberative democracy critiques (i.e., expert domination) (Lövbrand et al., 2010).

Instead of accepting the trade-off between democracy and scale, it is possible to engage citizens in designing and maintaining online deliberation. The following recommendations concern codesigning and cocreating online deliberation and AFs.

Codesigning online deliberation is an approach to accommodate citizens in respective contexts. Andrea Kavanaugh and colleagues (2005) highlight that democratic online deliberative design accounts for differences in interpersonal, technical, and communication skills. Therefore, the platform should be codesigned by experts and a sample of citizens. Through this exercise, several online deliberative structures emerge. In practice, a recommendation, such as using text-to-speech technology will compare expert and citizen expectations of useful algorithms.

Codesigning challenges the assumption of “technological instrumentality” since it requires experts and citizens to reflect upon the conditions of political communication. Janette Hartz-Karp and Brian Sullivan (2014) explain that the failure of online deliberation in increasing quantitative inclusion is thinking that technology is a “silver bullet.” As technological instrumentality, this view inhibits deepening citizens’ capacity to address complex issues. The “silver bullet” view proved ineffective in addressing democratic deficits such as in Serbia. Ivana Damnjanović (2019) demonstrates that relying on technology to enhance democratic communication without reflecting on how communication is organised renders technology useless and ineffective. Alternatively, codesigning activates citizens’ collective thinking about the role of technology in democracy.

Codesigning can also immunise online deliberation from being “hacked” by its programmers or host institutions (Gordon & Walter, 2019). In face-to-face deliberation, the risk of “hackability” is co-optation by political power, as in the experience of community groups in Chicago where citizens and facilitators extended deliberation beyond minipublics.
They organised, trained, and facilitated deliberative events before local institutions co-opted these efforts and excluded community groups from the process (Talpin, 2014). As citizens participate in codesigning with programmers and political institutions, they can scrutinise the extent to which deliberation is “hackable” and resist programmed coercion.

The context-sensitivity aspiration in online deliberative design applies to the codesign of AFs. Therefore, codesigning a chatbot for deliberation should reflect on the conditions and substance of political communication. Deliberation unveils the values societies hold (Cooke, 2000), so that citizens will be able to make an authentic judgment about communicating with artificial actors. If these communication technologies are not designed in this light, chatbots will either perpetuate existing communication pathologies or automate norms and values of communication irrelevant to the context.

Context sensitivity and cultural relevance affect satisfaction with the chatbot’s performance (Radziwill & Benton, 2017). Such a programming of chatbots to be context sensitive depends on design choices (Lucas et al., 2018). The difference between an expert-led automation of communicative norms and a deliberative democratic approach is that citizens are not excluded from discerning the values of political communication with algorithmic actors.

Citizen input in discerning personality traits and values of communication in AFs does not only constitute a democratic good, but also contributes to technological effectiveness. For instance, some participants found Sophie’s interventions unhelpful. Their dissatisfaction is based on the perception of Sophie’s prompts as annoying, unpleasant, or demotivating (Wyss & Beste, 2017). As this was the experience of a small percentage of participants, Wyss and Beste (2017) did not focus on this finding. However, chatbot personality is important and depends on cultural settings (Lucas et al., 2018). The social and cultural divide between those programming chatbots and those interacting with them can result in a chatbot communicating in a language inaccessible to all (Folstad & Brandtzæg, 2017). Hence, codesigning personality contributes to the effectiveness of AFs, as well as inclusivity in design.

Through cocreation, citizens can discern and update the rules of deliberation. Nonetheless, programmers and developers need to design the platform for “playfulness.” Gordon and Walter (2019) emphasise the importance of creating “meaningful inefficiencies,”
that is, the tension between designed technological efficiency and civic efficiency where citizens use technology for unintended purposes. Designing for playfulness allows citizens to “play within and with rules, not simply play out prescribed tasks” (Gordon & Walter, 2019, p. 314). In such play, meaningful inefficiencies include citizens’ questioning of why systems are made in a certain way (Gordon & Walter, 2019).

Particularly in cocreating chatbots, reflections by citizens should be used to update chatbots. In an experiment with a chatbot designed for stress management, research found that participants’ feedback on the sequences of the conversation with the chatbot has the potential to induce self-reflection on the part of participants (Park et al., 2019). As emancipation is through reflection, this experiment illuminates the possibility of algorithms facilitating reflection when guided by citizens.

While the exclusion of human facilitators deskills them, interacting and conversing with AFs improve their performance. However, improving AFs should not be tasked to experts alone. Participants should cocreate AFs by “playing” with them. The critical issue nevertheless is ensuring that chatbots embody a good conversation ethic and are not subverted for coercion, as in the case of the Microsoft Twitter chatbot Tay (Følstad & Brandtzæg, 2017). Tay learned to be a Nazi supporter and endorsed racist views based on her Twitter interactions with humans and other chatbots. Unlike human facilitators, chatbots are not programmed to fear social sanction. Thus, in learning how to improve, citizens and participants can keep exercising their reflection on what defines inclusive and authentic deliberation and how chatbots can be a part of it.

In sum, democratising online deliberation design and implementation allows citizens to develop their capacity for critical reflection, which emancipates them from expert coercion. Citizens in this process are able to critically reflect on the role of technology in democratic participation. Particularly in algorithmic facilitation, as Minna Ruckenstein and Linda Turunen (2019, p. 1039) assert, “the quest for better technology should critique any predefined division of labour between the human and the machine.”
3.6 Conclusion

Although implementing artificially designed citizen deliberations transcended the change in medium from physical to virtual, it retained some problems. Despite the fact that online deliberation can overcome some limitations of face-to-face minipublics, its design and implementation remain undemocratic.

Technology can increase the scale and quantitative inclusivity of deliberation. The trade-off, however, is coercing citizens to passively and uncritically accept the role of digital technologies in democracy, political participation, and communication and to uncritically accept the logics of experts designing these technologies, thus perpetuating expert coercion. In effect, designs that were not deliberated reproduce some of the less-than-ideal conditions which minipublics, in principle, aim to alleviate.

Particularly in algorithmic facilitation, the substantive difference between humans and algorithms is that humans can engage in mutual justification and reflection whereas algorithms cannot. This makes it critical to highlight the limitations of programmed deliberation. Currently, programmed deliberation is neither democratic nor emancipatory. Instead of expert-imposed design, online deliberation should accommodate citizens and participants to “play” with (and shape) the conditions and structures of online deliberation, especially as the role of algorithms requires reflection by humans in order to be aligned with cultural and political contexts. For deliberation to connect to wider contexts, citizens can reflect on their experience with artificial actors in deliberation to discern the role of algorithms in other political spaces.

The discussion herein concedes the potential in using technology to enhance online deliberation but with caution. Focusing on technology’s perceived advantages inhibits critical judgement of the inequalities and nondeliberative features in the wider areas of democratic participation and technology development.

Future research can consider examining the nondemocratic and nondeliberative practices in technology created for democratic participation, the division of labour between facilitators and algorithms, and citizens’ experiences with algorithmic facilitation.
3.7 References


De Cindio, F., & Peraboni, C. (2010). Design issues for building deliberative digital habitats. In F. De Cindio, A. Machintosh, & C. Peraboni (Eds.), From e-participation to online deliberation, proceedings of the fourth international conference on online deliberation (pp. 41–52). University of Leeds and Università degli Studi di Milano.


Chapter 3


Schlosberg, D., & Dryzek, J. S. (2002). Digital democracy: Authentic or virtual? 

Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., & Waters, K. (1996). When the 
interface is a face. *Human-Computer Interaction, 11*(2), 97–124. 
https://doi.org/10.1207/s15327051hci1102_1

Deliberative Democracy: Empirical Research and Normative Implications* (pp. 88– 

Mansbridge, & M. E. Warren (Eds.), *The Oxford handbook of deliberative democracy* 

Strandberg, K., Himmerloos, S., & Grönlund, K. (2019). Do discussions in like-minded 
groups necessarily lead to more extreme opinions? Deliberative democracy and group 

Challenges to enacting real world deliberation. *Journal of Information Technology & 
Politics, 9*(1), 82–96.

Talpin, J. (2014). Democratic innovations. In D. Della Porta & M. Diani (Eds.), *The Oxford 
handbook of social movements* (pp. 782–793). Oxford University Press.

*Annual Review of Political Science, 11*(1), 497–520. 
https://doi.org/10.1146/annurev.polisci.11.081306.070555

Towne, W. B., & Herbsleb, J. D. (2012). Design Considerations for Online Deliberation 

Velikanov, C. (2010). Facilitation procedures for written online deliberation: A Research and 
development project in the field of deliberative democracy. In F. De Cindio, A. 
Machintosh, & C. Peraboni (Eds.), *From e-Participation to online deliberation, 
proceedings of the fourth international conference on online deliberation* (pp. 240– 
249).

https://sites.google.com/site/australianblakhistorymonth/extra-credit

*Planning and understanding: A computational approach to human reasoning* (pp. 

Wright, S., & Street, J. (2007). Democracy, deliberation and design: The case of online 

General Conference, Hamburg, Germany.

asynchronous discussions. *Journal of Information Technology & Politics, 14*(3), 214– 
231.

Zgiep, M. (2019). Beyond institutions and the system: Network approach to deliberative 
Abstract

Algorithms are used to calculate and govern varying aspects of public life for the efficient use of the vast data available about citizens. Algorithms are used in areas such as criminal justice and welfare due to the assumption that algorithms are neutral and efficient in data-based decision making. The use of algorithms has ramifications on the ideal of democratic self-government, as algorithmic decisions are made without democratic deliberation, scrutiny, or justification. In the book *Democracy without Shortcuts*, Cristina Lafont argued against “shortcutting” democratic self-government. Lafont’s critique of shortcuts turns to problematise taken-for-granted practices in democracies that bypass citizen inclusion and equality in authoring decisions governing public life. In this chapter, I extend Lafont’s argument to include the “algocratic shortcut.” The democratic harms attributable to the algocratic shortcut include diminishing the role of voice in politics and reducing opportunities for civic engagement. In this chapter, I discuss what the algocratic shortcut is, its democratic harms, its relation to other shortcuts to democracy, and the capacity of “aspirational” deliberation to remedy the algocratic shortcut.
4.1 Introduction

An ideal democratic society is one where self-government is secured for its citizens. However, the complexities of public life make “shortcuts” to self-government tempting. In critiquing shortcuts to democracy, Cristina Lafont (2019, p. 19) defines self-government as the ideal of not “being coerced into obeying laws that one cannot endorse as at least reasonable upon reflection.” While shortcuts can be necessary for democracies, Lafont draws a distinction between shortcuts that expect “deference” and “blind deference”—where the former observes self-government by allowing citizens to accept decisions after reflection, the latter coerces citizens to accept decisions made by others without the opportunity to reflect and contest them. Blind deference, according to Lafont (2019), takes the form of deference to the majority vote, knowledgeable experts and political elite, or a randomly selected lot of citizens who take part in minipublics (Lafont, 2019, 2020). To these types of blind deference and shortcuts, I add algorithms.

The impacts of algorithms on democracy have been scrutinised in the past years. A sociotechnical definition of algorithms refers to them as a “series of instructions written and maintained by programmers that adjust on the basis of human behavior” (Benjamin, 2019, p. 183). Algorithms are relevant to the conduct of public life online whereby the availability of big data made algorithms central in analysing, as well as organising, public life (Eubanks, 2018; Kellner, 1999; Park & Humphry, 2019). In the digital public sphere, algorithms shape political and democratic communication by automating “editorial decisions” on what counts as “relevant” in our newsfeed or communication via “bots” (i.e., robots that are created to share, amplify, or distort messages online) to inhibit listening and promote censorship (Benvenisti, 2018; Frost, 2020; Peixoto & Steinberg, 2019; Wu, 2017). A scrutiny of these impacts is mostly focused on Silicon Valley platforms raising anxiety about the future of democracy (Bowman, 2020). These anxieties span the extent to which algorithms shape our behaviours. Their logics are increasingly in a “black box,” sheltered from scrutiny, justification, and explanation. Moreover, the focus on Silicon Valley platforms turns to threats to democracy seen in the concentration of power with these tech companies.

Concerns about the democratic impact of algorithms on democracy extend beyond social media to the functions of government. Government agencies use algorithms to automate decisions in welfare provisions, criminal justice, health, and many other contentious aspects of
Democratic Harms in Algorithmic Governance of Society

social life (see Eubanks, 2018). In deliberations and decision making, the ideal of self-government should be observed in particular functions and institutions of government, such as administration and courts (Pettit, 1997). Nonetheless, the applications of algorithms subvert this ideal. For example, in U.S. institutions of criminal justice, some courts use the automated decision-making software Correctional Offender Management Profiling for Alternative Sanctions for recidivism; its algorithm is scrutinised for producing racially biased prison sentences (see Angwin et al., 2016). Such examples raise two issues from a democratic perspective. First, algorithms displace human discretion in the realm of “unelected administrative agencies.” Second, by displacing human judgment, reasons behind algorithmic decisions are neither articulated nor justified, and consequently difficult to hold accountable (Strandburg, 2019).

The deployment of algorithms, artificial intelligence (AI), and big data by governments to govern society raises questions about the impact of algorithmic governance on democratic quality. Although algorithms can be arguably efficient in administering society, it comes at the expense of the democratic quality of decisions affecting citizens. By democratic quality, I refer to Nicole Curato’s (2015, p. 99) use of deliberative capacity as “an indicator of democratic quality.” In this light, democratic quality is the capacity of a polity to host deliberation that is (1) inclusive of citizens’ range of interests and considerations, (2) authentic, such that communication induces reflection rather than coercing others into accepting their reasons, and (3) consequential in creating change (Dryzek, 2009). In contrast with this definition of democratic quality, the logics of algorithms governing society automate systemic injustice, inequalities, and racism, thus undermining the capacity for inclusion and authenticity, and creating negative change by stabilising discrimination (see Eubanks, 2018; Noble, 2018).

In this chapter, I argue that the use of algorithms in decision making across institutions undermines the ideal of self-government in two ways. First, institutions attempt to use algorithms in order to evade justifying and explaining the logics it automates. This is a necessary condition for reasonable acceptance and meeting authenticity that reflects on the democratic quality of these decisions. Second, institutions resort to other shortcuts in an attempt to add a veneer of democratic participation, yet these shortcuts still suffer in terms of being

---

20 Following Lafont’s (2019, p. 5) choice of “citizen(s)” in discussing shortcuts, reference to “citizens” is “as a shorthand expression…[and] often use it to mean anyone who is subject to the laws of a country, regardless of their specific status (e.g. undocumented immigrants).”
inclusive. I outline the democratic harms in the algocratic shortcut that is performed in three parts. The first section outlines the contours of the algocratic shortcut as a type of shortcut to democratic self-government described by Lafont (2019), as well as their harms to equality and inclusion in decision making, reason, and voice. In the second section, I put forward an illustrative case of institutional deliberations about governing algorithms in Europe, focusing on two points: (1) that the algocratic shortcut cannot be resolved by using other shortcuts, namely epistocratic and lottocratic and that (2) the implications of inequality across influence, voice, and reasons. The final section discusses a potential solution to the algocratic shortcut by arguing for what Lafont (2019) refers to as “aspirational” political deliberation and justification to remedy inequality across voice, reasons, and influence.21

4.2 Shortcuts to Democracy and Algorithms

In Democracy without Shortcuts, Lafont (2019) problematises the practices in democracies which harm the realisation of self-government as a political ideal. To realise this ideal, Lafont (2019) argues that citizens ought to be able to accept decisions and endorse them as their own upon reflection. Expecting citizens to not reflect on these decisions is when shortcuts become undemocratic and involve “blind deference” to others, thus undermining self-government. The crux of Lafont’s (2019) argument is that taken-for-granted practices in democracy involve blind deference, particularly to the majority (procedural shortcut), the knowledgeable (epistocratic shortcut), or randomly selected citizens (lottocratic shortcut).

In theorising the algocratic shortcut against other shortcuts, I do not refer to procedural shortcuts. In explaining the deference expected in the procedural shortcut, Christiano (2012, p. 33) explains that it is a model where citizens elect representatives endorsing acceptable “packages of aims.” This does not apply to algorithms as representatives do not introduce their stance on algorithms and automated decision making to citizens. Hence, manifestations of democratic harms in the algocratic shortcut are contextualised against epistocratic and lottocratic shortcuts.

The algocratic shortcut builds on epistocratic shortcuts that justify bypassing citizen deliberation by referring to the epistemic quality of decisions made by algorithms, as with

---

21 This order is based on Ercan, Hendriks, and Dryzek’s (2019) sequencing of division of deliberative labour.
experts. While Lafont (2019) regards epistocratic shortcuts as democratically justified, I argue that neither epistocratic nor algocratic shortcuts can be accepted as democratic shortcuts.

Thinking about the centrality of noncoercive and open reflection in the ideal of self-government, I follow Curato’s (2015) approach in using “deliberative capacity” as an “indicator of democratic quality.” John Dryzek (2009) conceptualises “deliberative capacity” as a polity’s capacity for hosting deliberation that is inclusive, authentic, and consequential. Inclusivity is of all affected by the decisions, and a range of considerations, preferences, and reasons. Authenticity is the capacity to communicate reasons in a noncoercive, nonmanipulative, and reflexive manner. The reflexivity of these exchanges is the extent to which they connect particular interest to understandings of the common good. Consequentiality is the capacity for these inclusive and authentic exchanges to create change. Considering the criteria of deliberative capacity, epistocratic and algocratic shortcuts exclude citizens on the basis of epistemic superiority. Hence, the two shortcuts pivot on the exclusion of citizens, and are therefore undemocratic. Epistemic blind deference warrants implications on inclusion and equality.

Prior to Lafont (2019), James Bohman (2000) highlighted the dangers of blind epistemic deference. He explains that:

… social asymmetries inherent in the communicative and cognitive division of labour threaten to short-circuit the deliberative process, making it impossible for citizens to have equal opportunities to influence many decisions, to express opinions freely and effectively, and to have their reasons fully and fairly considered. (Bohman, 2000, p. 48)

A realisation of self-government as reasonable acceptance of laws upon reflection is underpinned by the equality and inclusion of citizens (Bohman, 2000; Lafont, 2019). As such, shortcuts sustain inequality across influence over decisions, effectiveness of voice and opinion, and full and fair consideration of reasons. As with epistocratic and lottocratic shortcuts, the algocratic shortcut demands blind deference to algorithms in making decisions that govern society. Similar to epistocratic shortcuts, the algocratic shortcut should not be accepted as a democratic shortcut. It precludes the need to observe the ideal of inclusion as algorithms are deemed as powerful capacities for automation and data processing that are efficient, and present answers concerning the “value” of citizens in a data-based society. Table 2 presents a
summary of the democratic harms associated with the two shortcuts to democratic self-government considered by Lafont (2019), contrasted with the algocratic shortcut.

<table>
<thead>
<tr>
<th>Democratic Self-Government</th>
<th>Epistocratic Shortcut</th>
<th>Lottocratic Shortcut</th>
<th>Algocratic Shortcut</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Reasonable acceptance upon reflection (Lafont, 2019).</td>
<td>- Blind deference to the political elite and experts (Lafont, 2019).</td>
<td>- Blind deference to randomly selected citizens or the “lot” (Lafont, 2019).</td>
<td>- Blind deference to algorithms.</td>
</tr>
<tr>
<td>- Equal influence over decisions (Bohman, 2000).</td>
<td>- Only the political elite and experts have influence (Lafont, 2019)</td>
<td>- Only a select lot of citizens have a window to influence decisions (Lafont, 2019).</td>
<td>- Data-based and statistical decisions in governing society.</td>
</tr>
<tr>
<td>- Fair and full consideration of reasons (Bohman, 2000).</td>
<td>- The political elite and experts have epistemically superior reasons (Lafont, 2019)</td>
<td>- Epistemic diversity of the lot is sufficient (Lafont, 2019).</td>
<td>- Response to “a crisis of value and valuation” (Schüll, 2019).</td>
</tr>
<tr>
<td>- Free and effective voice and opinion (Bohman, 2000).</td>
<td>- Voice and opinion are irrelevant to making democratic decisions (Lafont, 2019)</td>
<td>- Voice and opinion are constrained by the boundaries of the sites of a lot’s deliberations (Lafont, 2019).</td>
<td>- Decisions without deliberation and in areas rife with structural injustices (Büscher et al., 2016; Eubanks, 2018; Noble, 2018).</td>
</tr>
</tbody>
</table>

4.2.1 The Algocratic Shortcut

The algocratic shortcut is similar to blind deference to experts. The slight difference between deference to epistocrats and algorithms is that algorithms are sheltered by representations of algorithms as value-free technology, immune to human bias, and epistemically superior (Danaher, 2016). However, algorithms are far from value-free, and their use in governing society lacks any meaningful democratic accountability or scrutiny, especially for informing and authoring algorithmic governance of society.

By governing different aspects of public life without observing democratic ideals, such as the inclusion of those affected and subjected to decisions (Erman, 2012), algorithms are “restructuring the ways in which key democratic institutions and organizations work and operate” (Gran et al., 2020, p. 1780). Virginia Eubanks (2018) argues that by sidestepping inclusion, a government has
… the ethical distance it needs to make inhumane choices: who gets food and who starves, who has housing and who remains homeless, and which families are broken up by the states.

To demonstrate the consequences of algorithmic capacity in making inhuman choices, I use the example of “robodebt” in Australia. In December 2016, some Australians who received social security payments began to be notified with alleged debts, based on the relevant government agency’s debt collection algorithm. The system was rolled out to optimise efficiency in debt collection. However, the system’s algorithm was flawed, and the case became known in the media as robodebt (Knaus, 2019). Affected citizens protested these algorithmic debt decisions using the online hashtag #NotMyDebt and launched a website to collect stories of affected citizens. A successful class action was launched where the court condemned the flawed algorithm for unjustly assigning debt, thereby ruling for a compensation bill of AU$1.2 billion (Henriques-Gomes, 2020).

Parliamentary and Commonwealth Ombudsman investigations into robodebt considered some of the democratic limits of using algorithms and automated decision making. The Ombudsman inquiry found that the use of algorithms and automation in debt collection failed to adhere to principles of transparency and procedural fairness, where citizens can understand how the decision was made and have the capacity to challenge the decision (Glenn, 2017). In summarising the inquiry’s findings, the report reads:

Good public administration requires a transparent and open decision-making process that clearly sets out the issues the person needs to address to challenge a decision and the findings of fact on which the decision is based. This principle continues to apply when decision making is automated. (Glenn, 2017, p. 2; emphasis added)

In addition to the Ombudsman’s inquiry on the ability of citizens to challenge decisions, another parliamentary inquiry focused on the exclusion of affected citizens. The parliamentary inquiry corroborated the severe consequences of robodebt that resulted from the exclusion of citizens from participating in the systems’ design (Community Affairs References Committee, 2017).

The robodebt case exemplifies the adverse effects of algorithmic governance of society on democratic self-government. Citizens were not included in discerning the algorithm (logic of the system), the reasons for the automated system were not open to scrutiny or justified to
society, and the capacity to influence and challenge decisions were not accommodated. Accordingly, I argue that a reliance on algorithms in making decisions can be located within a discussion that argues against shortcuts to democracy.

Justifications of the algocratic shortcut are similarly embedded in epistocratic arguments. Based on “epistocracy,” John Danaher (2016, p. 247) defines algocracy as

... a particular kind of governance system, one which is organised and structured on the basis of computer-programmed algorithms. To be more precise, [it is] a system in which algorithms are used to collect, collate, and organise the data upon which decisions are typically made and to assist in how that data is processed and communicated through the relevant governance system.

Underpinning algocracy (i.e., the governance system centred on algorithms and algorithmic capacities) is how algorithms are perceived as producing epistemically superior decisions compared to experts for not being subject to human bias (Danaher, 2016). Accordingly, the epistemic quality of algorithmic decisions is couched in the nonhumanness of algorithms. For example, in comparing algorithms and human judges in criminal justice, algorithms are seen as immune to bias and external influences (Wykstra, 2018). The perception of this epistemic quality is also present in public opinion. A survey of Dutch citizens’ attitudes to automated decisions reveals people’s inclination to perceive nonhuman decisions “fairer” to human decisions (Helberger et al., 2020). This, however, is laden by misconceptions of algorithms and AI (Helberger et al., 2020).

These misconceptions encounter the same problems attributed to epistemic dependence on experts. Citizen autonomy and agency are hindered by “epistemic dependency” when epistocrats argue that citizens cannot understand or evaluate the knowledge produced by experts (Bohman, 2000). Similarly, algorithmic decisions are inexplicable and inaccessible to citizens, thus undermining citizen autonomy and agency. The contributions of epistemic dependence on experts and algorithms to deliberative capacity fail to be democratic for not observing inclusivity or authenticity. Citizens are excluded on the basis of their epistemic limitations. This, in turn, is used to avoid providing citizens with accessible justification for the reasons experts make decisions that affect public life.
A core problem with shortcuts to democratic self-government in institutional deliberation is how blind deference oversteps mutual justification. Lafont (2019) accepts that some measure of deference is necessary in representative democracies. For example, deference by voting or to the randomly selected lot of citizens are not principally undemocratic, and indeed facilitate the flow of organising public life. The problem however is when elected representatives or selected citizens dislocate the role of this deference in self-government. Instead of regarding deference as a step in an ongoing effort to realise self-government, it becomes a reason to “shortcut” it and settle for the wisdoms of elected representatives, selected citizens, epistocrats, and algorithms.

Here, blind deference presents an obstacle to self-government whereby citizens cannot reasonably accept the law (or decision) upon reflection, as they are excluded from the process of justification (Lafont, 2019). The reflection of citizens—grounded in citizen autonomy for reflecting on their views and the views of others synthesises these to distinguish self-serving arguments from ones concerned with public interest—is predicated on the presence of political justification (Ferree et al., 2002). Where justification is present, communicative distortions and manipulation become more manageable by focusing on the capacity of political agents for critical reflection (Whipple, 2005). In the following discussion, I consider the manifestation of these harms in the algocratic shortcut, compared to their manifestations in other shortcuts.

4.2.2 Democratic Harms in the Algocratic Shortcut

Based on Lafont (2019) and Bohman (2000), the criteria for assessing the democratic harms of the algocratic shortcut is the extent to which shortcuts hinder the conditions for reasonable acceptance upon reflection. This is underpinned by the inequality across influence, voice, and reasons. From this perspective, the democratic harms associated with the algocratic shortcut, among others, are in bypassing reasons and justifications. Shortcuts subvert conditions needed to constitute reasonable acceptance: equality of influence over decisions, effective voice, and consideration of reasons.

Algorithms amplify this inequality across influence, voice, and reason in governance. Under algorithmic governance of society, algorithms present data-driven, dangerously simplistic answers to the complexities of public life. The efficiency of algorithms in making
sense of big data available about citizens obscures scrutiny of the democratic quality of algorithmic decisions.

In complex societies, algorithms resolve “a crisis within the domain of value and valuation” centred around the question of “[h]ow to value a person, people, peoples?” (Beller, 2018, p. 56). This “algorithmic idealism,” as Davis and colleagues (forthcoming) conceptualise it, promise a clear answer that overcomes the limitations of human rationality and biases that “seek to neutralise demographic disparities.” In other words, this idealism overlooks that demographics are shaped by socioeconomic and political structural inequalities. Therefore, statistical probability is not actually fair or neutral when it does not account for the forces and factors that shape demographic distributions e.g., income and residential districts that can increase one’s likelihood to be wrongfully convicted for urban crimes.

Hence, the problem with this idealisation of algorithmic decisions is that these answers carry structural injustices to the process of “algorithmic production of social difference” (Beller, 2018; Davis et al., 2021). Even historically, politically, and socially complex issues, such as racial bias and discrimination in criminal justice institutions, is perceived to be “fixable” with an automated decision system instead of addressing the politics of institutional racism (Kahn, 2018). Therefore, structural injustice and bias become shielded by the assumption of algorithmic efficiency which is prioritised over democratic accountability. For example, Aboriginal Australians constitute half of the suspects identified in the crime prevention algorithmic system used by police in the Australian state of New South Wales. Yet, the state police denied inquiry into the algorithm on the pretence that revealing the algorithmic logic can compromise the secrecy for needed crime prevention (see Goldenfein, 2019; Sentas & Pandolfini, 2017).

Responding to Lafont, Robert Goodin (2020) highlights that “blind deference” should not be rejected in its entirety but should be judged for the degree of expected blind deference. Algorithmic governance of society involves varying degrees of blind deference, from being imposed without public deliberation and entrenching inequality in influencing decisions (Büscher et al., 2016) to limiting the opportunities for citizen participation. By providing data-based answers on society without citizen inclusion, algorithms diminish opportunities for democratic and civic participation. For example, in a survey on attitudes to algorithms and AI decisions, citizens confirmed that governments do not involve them in governing AI across
countries including Estonia, Denmark, and China (Carrasco et al., 2019). The degree of blind
dereference to experts, elite, and algorithms deciding on automating menial decisions, such as
reporting infrastructure repairs, might seem harmless but can further reduce citizens’
participation in governance and accountability of public office (Peixoto & Steinberg, 2019).
The feed of data from automated systems that are then automatically processed and analysed
negates the need for citizens’ voices and input. The chain of automated processes in such
menial tasks then limit opportunities to interact with public offices unless the systems
malfunction or become dysfunctional.

Social asymmetries and inequality are also augmented as applications of automated
decision making are concentrated in areas such as social welfare and criminal justice. Eubanks
(2018), Safiya Noble (2018) and Cathy O’Neil (2016) discuss at length how algorithms
automate structural and racial injustices and limit the opportunity structure for reversing these
decisions. On this, the United Nations Special Rapporteur on extreme poverty and human
rights, Philip Alston, warns: “As humankind moves, perhaps inexorably, towards the digital
welfare future it needs to alter course significantly and rapidly to avoid stumbling zombie-like
into a digital welfare dystopia” (United Nations Human Rights Office of the High
Commissioner, 2019). In this statement, Alston is describing the ramifications of the short-
term convivences of algorithms and automated decision-making systems to welfare institutions
that come at the expense of citizens’ welfare.

Determining the value of groups in society based on data and algorithms also affects
equality across free and effective voice in democratic politics. Miller (2017, p. 126) highlights
that “[algorithmic politics] is numerical rather than spatial, [and] operational rather than
expressive or communicative.” This means that algorithmic decisions disregard contextual
factors that affect public life differently and eliminate the nuance brought through meaning
making exercised through communication. Especially in light of big data, Frank Pasquale
(2015, p. 216) explains the political economy of algorithmic decisions:

Capitalist democracies increasingly use automated processes to assess risk
and allocate opportunity. The companies that control these processes are
some of the most dynamic, profitable, and important parts of the information
economy. All of these services make use of algorithms, usually secret, to
bring some order to vast amounts of information. The allure of the
technology is clear—the ancient aspiration to predict the future, tempered
with a modern twist of statistical sobriety.
In this quote, Pasquale expands on Altson’s observation on the convivences of algorithmic decision-making systems as a feature of political economy in capitalist democracies. These convivences are specific to the supply of decision-making systems by companies, the high value perceived of most use of the available data, and the certainty about the soundness of these statistical decisions. With data, information, and algorithms to analyse the present and predict the future, “individual storytelling” is undermined (Couldry, 2010). Individual stories are irrelevant as algorithms “calculat[e] our potential as students, workers, lovers, [and] criminals” (O’Neil, 2016, p. 2). Therefore, the combination of these convivences and the availability of data limit citizens’ agency in using their own voice and stories to determine their future. Instead, the futures of citizens and their position in different institutions is determined by algorithms.

However, the implications of algorithms on self-government and voice do not emerge from the political economy of big data alone. The “crisis of valuation” mentioned earlier is an important aspect. The quantification of society makes our understanding of ourselves robotic, since this is how the robots making the decisions understand us (Schüll, 2019). As citizens are reduced to units that are monitored, measured, and analysed, algorithms present “statistically deterministic” accounts of individuals undermining their autonomy in storytelling (Hilderbrandt, 2011). Hence, the availability of data turns citizens into profiles that can be monitored for certain categories of socioeconomic and demographic markers that are then analysed by an algorithm to assess citizens’ worthiness of accessing institutions, benefits, or even punishments. To understand the value and entitlement of citizens based on the complexity, variance, and uniqueness of their experiences is inconvenient to a society comfortable with reducing citizens to statistics.

The algocratic shortcut bypasses reason giving and justification that allow for self-government under the condition of reasonable acceptance upon reflection. Against the ideal of reason giving, algorithmic decisions are undemocratic and illegitimate as they are “inaccessible” and “incomprehensible.” Consequently, these decisions undermine citizen democratic authorship (Benvenisti, 2018; Danaher, 2016), which is important for democracy because it ought to reflect the values of a society and ensure that citizens can scrutinise the decisions and policies they are asked to obey. Should these decisions be disconnected from the condition of endorsement upon reasonable reflection, the democratic quality of this society is jeopardised. Because algorithmic decisions are positioned above democratic politics, these
decisions are “not transparent, non-accountable, and non-appealable” (Ash, 2018, p. 82). For example, it is often argued that algorithms cannot explain their decisions, reverse engineering is time-intensive and costly, or that algorithms are protected by proprietary rights. These arguments about algorithmic decisions, hence, violate the condition of authorship and acceptance upon reflection.

Claims of efficient algorithmic decisions evade questions concerning the values algorithms enforce in governing society. Resorting to algorithms to answer questions about the value of individuals and groups in society “paves the way for AI experts and entrepreneurs to present themselves as architects of society” (Katz in Goldenfein, 2019, p. 60). Relying on algorithms to anticipate or detect behaviour and make decisions based on these probabilities encourages a path of dependency as discussed earlier where our understanding of ourselves and society is done through the lens of reductionism of algorithmic logics (Edwards & Veale, 2017). Instead of technologically deterministic and nondemocratic dependency, unchecked power of algorithms, their designers, and their sponsors in government must be subject to public deliberation.

Remedying these democratic harms involves subjecting algorithms to “democratic participation and constitutional restraint” beyond democratic voting (Hilderbrandt, 2011; Morozov, 2017). This deeper engagement involves democratic deliberation. To correct shortcuts, Lafont (2019) argues that laws need to be subject to political deliberation and justification. Further, she explains that the position of political justification in democratic decision making can be “aspirational” or “institutional” (LaFont 2019). An aspirational position of political justification involves finding public reason before enacting decisions and laws. This happens when institutions listen to deliberations in the public sphere and carry these considerations, reasons, and interests forward. Meanwhile, institutional deliberation means that institutions build a mechanism to trigger a process of justification and explanation of decisions, policies, and laws when called on by citizens. Institutionalising the capacity for public scrutiny is both more realistic and practical, as canvassing the public sphere can be challenging (Lafont, 2019). Nonetheless, this has its own limits.

Below I elaborate on this argument using the example of governance of algorithms under the European Union (EU) General Data Protection Regulation (GDPR). The case demonstrates how institutionalising “justification” is important but insufficient. Without
“aspirational” political deliberation and justification, there is a risk of merely supplanting the algocratic shortcut with epistocratic and lottocratic shortcuts.

4.3 Democratising Algorithms in Europe: Political Justification through Epistocratic and Lottocratic Shortcuts

The GDPR cases involve an attempt in the EU and the UK to democratically govern algorithms. The GDPR regulates data collection and processing of the mass of data collected about individuals in EU jurisdiction. Terms set out by the GDPR have been used to question social media platforms for their responses to risks affecting democracy in Europe, such as the spread of disinformation (Boucher et al., 2019). The GDPR institutionalises a process to justify decisions made by algorithms (i.e., algorithmic decisions). However, it also demonstrates the limitations of remedying shortcuts by deploying other shortcuts.

I put forward a critique of the GDPR in three steps, with each demonstrating the democratic harms of attempting to correct the algocratic shortcut using other types of shortcuts—this involves the harm of inequality in influencing decisions. The second part offers a critique of shortcutting the consideration of reasons by relying on epistocratic and lottocratic shortcuts. The third step discusses implications of the inequality of free and effective voice in public deliberation about governing algorithms.

4.3.1 Inequality in Influencing Decisions

The GDPR is, however well-intentioned, the product of an epistocratic shortcut where the political elite and experts have developed the limits on data collection and processing, and automated decisions affecting citizens. This is based on the genealogy of the GDPR that originates from doctrines and policies that were developed by political elites, such as policymakers in the European Commission and human rights and data regulation experts (see Souza et al., 2020). The inclusive quality of GDPR of citizens is seen after its adoption. It is limited to using a lottocratic shortcut about legitimising the use of algorithmic and AI decisions across institutions. This involves the random selection of some citizens to sample public considerations for legitimating algorithmic decision-making systems.

Despite the exclusion of citizens from shaping the terms of the GDPR, the regulation purports to secure the “right to explanation” for citizens affected by algorithmic decisions (De
Democratic Harms in Algorithmic Governance of Society

Gregorio, 2019; Deeks, 2019). This is expressed in the form that, since algorithms do not explain their decisions, under Article 22 of the GDPR individuals “have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning [them] or similarly significantly affects [them]” (Regulation EU 2016/679 2016).

Along with the right to explanation, individuals also have a “right to obtain human intervention” for individuals to “express [their] point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision” (Regulation EU 2016/679, 2016). Although Article 22 in the GDPR reflects institutionalising justification and reason giving, both are insufficient for self-government.

Therefore, in terms of equality in opportunities to influence decisions, the rights to explanation and to contest algorithmic decisions mitigate the consequences of algorithms. However, this, I argue, is not enough because they do not allow for authorship. Citizens are not substantially included in authoring which decisions to be automated in governing public life, if at all.

The GDPR notionally institutionalises a process to contest and receive a justification or an explanation on a specific automated decision and does not give citizens a right to inquire about the norms inscribed in the algorithm. In other words, the right to explanation is not an explanation of the norms and constraints designed for the algorithm (Loi et al., 2020). As such, the “right to explanation” and “right to human intervention” can apply in cases, such as in Denmark where robots made “incorrect decisions” in taxation where “[n]o human case officers are involved” (Motzfeldt & Næsborg-Andersen, 2017, p. 105-106).

Because the GDPR sidelines authorship and self-government, the rights to explanation and contestation are only individual rights and not inclusive of democratic society. For instance, the Danish Parliament in April 2019 approved a law to automate decisions on job seekers even though a pilot phase proved that algorithms had no effect on the capacity of public servants’ decisions on job seekers between 2016 and 2017 (Byrne & Sommer, 2019). The GDPR specifies no mechanism, obligation, or consideration in explaining and justifying why these areas of governing and administering society are automated. The decision to automate is unjustified; the GDPR does not require it to be justified.
Chapter 4

The institutionalised justification here applies to individuals rather than the citizenry writ large. In navigating what constitutes an “explanation” in algorithmic decisions, it is arrived at through epistocratic and lottocratic shortcuts.

4.3.2 **Inequality in Full and Fair Consideration of Reasons**

Full and fair consideration of reasons to inform what constitutes democratic and justified role of algorithms under the GDPR is advanced via epistocratic and lottocratic shortcuts. This produces inequality in considered reasons to those who are in power (i.e., political elite and experts) or those selected to have reasons of considerable weight (i.e., randomly selected lot of citizens).

While the epistocratic and lottocratic shortcuts in this case made important contributions to public deliberation, their democratic limitations are revealed when citizens are affected by algorithmic decisions. The political elite and experts have an important role in argument production, as well as in commissioning the spaces for including randomly selected citizens. In argument production, expert committees at the Council of Europe (COE) articulate the effects of algorithmic decisions on society and democracy (i.e., epistocratic shortcut). In creating spaces for citizen reflection, experts in the UK were involved in commissioning hybrid deliberation that consists of two parallel deliberations: one between citizens and the other between the elite and experts. The analysis of the two parallel deliberations is then synthesised to report on citizen and expert considerations. However, the focus here, demonstrative of the lottocratic shortcut, is three citizens’ juries on AI and algorithmic decision making.

Arguments and reasons provided by the political elite and experts in scrutinising algorithms encompass the impact of algorithms on democracy. The COE Committee on Political Affairs and Democracy made a motion for a resolution on the need for democratic governance of AI (Committee on Political Affairs and Democracy, 2019). Another COE committee, the Expert Committee on human rights dimensions of automated data processing and artificial intelligence in the COE critiqued the “privatization of decisions about public values” emanating from the automation of decision making without “democratic participation and deliberation” about these values (Bernstein et al., 2019, pp. 40-41). Particularly troubling is the absence of “democratically legitimated oversight over the design, development, deployment, and use of algorithmic tools” (Ministers’ Deputies, 2019, p. 5). The recommended
democratic process is “to maintain an open and inclusive dialogue with all relevant stakeholders globally with a view to avoiding path dependencies” (Ministers’ Deputies, 2019, p.5).

Moreover, the COE Ministers’ Deputies (2019) issued a declaration on algorithmic manipulation and its violation of democratic ideals. The first mention of algorithmic manipulation is in the context of machine learning tools which have a … growing capacity not only to predict choices but also to influence emotions and thoughts and alter an anticipated course of action, sometimes subliminally. The dangers for democratic societies that emanate from the possibility to employ such capacity to manipulate and control not only economic choices but also social and political behaviours, have only recently become apparent. In this context, particular attention should be paid to the significant power that technological advancement confers to those—be they public entities or private actors—who may use such algorithmic tools without adequate democratic oversight or control. (Ministers’ Deputies, 2019, p.3)

In the declaration, the Ministers’ Deputies highlight that public and private institutions’ use of algorithms ought to be subject to democratic control, particularly due to their capacity to manipulate and undermine individual agency. Another related violation is using algorithms “to identify individual vulnerabilities and exploit accurate predictive knowledge, and to reconfigure social environments in order to meet specific goals and vested interests” (Ministers’ Deputies, 2019, p. 3). Therefore, the Ministers’ Deputies are wary of the extent to which algorithms can be used in shaping political and social choices at the individual and society levels.

In spaces created for citizen inclusion, political elite and experts have defined the scope for citizen deliberation (i.e., legitimating algorithmic decisions as this is perceived to be the barrier to a wider rollout of algorithms). This scope is based on an independent review of the AI industry by Wendy Hall, professor of computer science, and Jérôme Pesenti, CEO of BenevelentTech, submitted to the U.K. government. The report discusses the need to develop transparency, explainability, and accountability of AI decisions as some of the areas that would improve “public trust” in the fairness of algorithmic decisions (see Hall & Pesenti, 2017, p. 14, 66). This legitimacy is perceived to be achievable when algorithms can explain their decisions. Some problems pertaining to “explainability” include that the “right to explanation” in the
GDPR is nonbinding (Wachter et al., 2017), and that GDPR specifies what constitutes an “explanation” (Hunt & McKelvey, 2019). Nonetheless, developers of automated decision-making systems prioritise explainability as it would “legitimate” algorithmic decisions (Hunt & McKelvey, 2019).

Citizens’ juries commissioned by the U.K. Information Commissioner’s Office (ICO) framed deliberation around the explainability of AI decisions as dependence on algorithms and automation is inevitable due to investments in the industry (Leslie, 2019; Project explAIn: Interim report, 2019). Herein, as the scope and purpose are predefined by the experts, the “lot” of randomly selected citizens is used to shortcut public deliberation about defining what constitutes algorithmic decisions as reasonable and acceptable.

The epistemic quality of the three citizens juries were constrained by the interests of their commissioners. The first was commissioned by DeepMind and organized by the Royal Society for Arts, Manufactures and Commerce (RSA).22 This was a jury of 25 to 29 citizens from England and Wales held over four days between May and October 2018 (RSA, 2019). The following two citizens’ juries were in Coventry and Manchester, respectively, in February 2019 that was co-commissioned by the ICO and National Institute for Health Research Greater Manchester Patient Safety Translational Research Centre. These two juries, with 36 jurors in total, were organized and facilitated by Citizens’ Juries c.i.c. and the Jefferson Center (Project explAIn: Interim report, 2019). The juries were centred around scenarios of algorithmic and automated decision making in three sectors of governance: healthcare, the labour market, and criminal justice (Artificial Intelligence (AI) & explainability Citizens’ Juries Report, 2019; RSA, 2019).

Each commissioning organization also constrained the juries by pre-set issue framing. The RSA citizens’ jury was grounded in democratic politics, highlighting the use of algorithmic decisions by “public bodies” as a “public interest issue” subject to “public deliberation” (RSA, 2019). By contrast, juries in Coventry and Manchester were framed as deliberations about obstacles to trusting algorithmic decisions, such as the “trade-off between AI transparency and AI performance” (Artificial Intelligence (AI) & explainability Citizens’ Juries Report, 2019).

---

22 DeepMind is a subsidiary of Alphabet Inc. (Google’s parent company). Theorising the role of private technology companies in public deliberation is a discussion in its own right but such is beyond the focus of this paper.
Despite differences in framing, there was an emerging preference across the three juries for an “explanation” from algorithmic and AI systems. Explainability of algorithmic decisions is on par with explainability of human decisions. As such, where human decision-makers are expected to justify their decisions, so are algorithms (Artificial Intelligence (AI) & explainability Citizens’ Juries Report, 2019). Ultimately, explanation is articulated as a mechanism for ensuring accountability supported “by establishing a legal requirement for explanation and a right to appeal” (RSA, 2019, p. 9).

4.3.3 Inequality in Free and Effective Voice

In spaces for expression to exercise free and effective voice, the political elite and experts also lead the deliberations. Examples include blog posts and op-eds (e.g., Dreyer & Schulz, 2019; Mchangama & Liu, 2018), statements by the Parliamentary Assembly of the Council of Europe rapporteur, Deborah Bergamini, at the Organisation for Economic Development and Co-operation Global Parliamentary Network meeting in October 2019 (see Darker side of AI beginning to emerge, warns rapporteur, 2019).

At the EU level, collective inquiry is part of the European Data Protection Board’s public consultation process. The process, via an online website, collects “views and concerns of all interested stakeholders and citizens.” The only consultation concerning automated decision making, in respect to Article 22, was open from February to May 2020 on a guideline adopted in January on the processing of personal data by autonomous vehicles (see Public Consultations, 2018). Experts herein define the agenda of collective inquiry.

As for civil society, its role is twofold: inform and include citizens in organising collective inquiry. For instance, the civil society organization AlgorithmWatch communicates “stories” of algorithmic decision making in public life, from content moderation on social media and Google search results in policing and border control. Although spaces for expression have not been inclusive to citizens, citizens are included in the collective inquiry into algorithms to express dissent as algorithms have a wider and direct impact on citizens’ lives.

In organising collective inquiry, AlgorithmWatch in their project “monitoring Instagram,” asks individuals to participate in the project by allowing AlgorithmWatch to use
their Instagram data to investigate whether Instagram’s algorithms discriminate between individuals (AlgorithmWatch, 2020).

Despite the fact that social media involves spaces for voice, the extent to which these are effective is questionable. On the social media website Reddit, there is a space dedicated to GDPR discussions (i.e., subreddit ‘r/GDPR’). Specific to Article 22 of the GDPR and automated decisions, a user (redditor) expresses dissatisfaction with a government agency’s response to their inquiry and seeks clarity from other redditors. Follow-ups or responses from the respective agency to such opinion is unevidenced.

Experts’ concerns regarding the impact of algorithms on democracy and justice are more concrete when algorithmic decisions have wider and direct effects on citizens. Whilst making changes necessary to prevent the spread of coronavirus, the UK mandated school closure and cancelled sitting exams. To mitigate the effects of this on school-leaving exams, the U.K. Office of Qualifications and Examinations Regulation (Ofqual) in August 2020 used an algorithmic system to assign students grades for their secondary certificate exams. Students protested the algorithmic decisions highlighting the elitist and classist logic of the algorithm. Spaces for voice spanned social media, occupying physical spaces, and interviews in news media (see BBC, 2020; Adams et al., 2020).

This case of attempting to democratise algorithms reveals the limits of an institutionalised political justification, particularly one that is institutionalised through epistocratic and lottocratic shortcuts. Inequality across influence, voice, and reasons demonstrate the ramifications of exclusion on self-government. When the experts or citizens winning the “lottery” of democratic participation decide on what is reasonably acceptable, decisions that affect more citizens in society show the misalignment between the reasons arrived at via shortcuts and what society considers a reasonable justification of the coercive power of algorithms. The case of Ofqual dissidence demonstrates that aspirational political deliberation and justification should not be downplayed, especially when decisions have widespread and direct impacts on citizens.

The objection to shortcuts here concerns the use of single “one-off” processes of deliberation and scrutiny among the elite or randomly selected citizens, which are insufficient.

23 Accessed from https://www.reddit.com/r/gdpr/comments/9vc2gz/gdpr_article_22_is_profiling_a_requirement/
to address the democratic harms in the algorithmic governance of society. As I argue below, rather than a one-off shortcut, the algocratic shortcut is best remedied by “aspirational” public deliberation.

4.4 Remediining the Algocratic Shortcut

The limitation of the algrocratic shortcut to self-government is in the tension between endorsing algorithms to govern society, and what Lafont (2019, p. 4) explains as self-government where citizens can “endorse [institutions, laws, and policies] as their own.” Remedying the algocratic shortcut cannot be done through other shortcuts. As seen in the demonstrative example of the GDPR in Europe, epistocratic and lottocratic shortcuts to governing algorithmic decisions are unsatisfactory for excluding affected citizens. Although shortcuts institutionalised an explanation for algorithmic decisions, this process excludes citizens from authoring the decision to automate and from scrutinising the choices informing algorithms. The minipublics about democratic control of algorithmic decisions, for instance, cannot be regarded as the final and ultimate step for deliberations about algorithms. The issues and concerns were selected by experts and do not cover all ranges of algorithmic decisions. Moreover, the randomly selected lot of citizens are not necessarily inclusive of those affected by these decisions like the students impacted by the Ofqual algorithm. For these reasons, minipublics counts as one small step but not the ultimate remedy to the democratic shortcomings of algocracy.

The assumption of algorithmic epistemic quality and efficiency mirror the democratic limitation of shortcuts as Lafont (2019, p. 88) highlights: “There are no shortcuts to make a political community any better than its members, nor can a community move faster by leaving their citizens behind.” Hence, remedying a shortcut will happen over time and can have blueprints that are “aspirational” where public reason is considered before “imposing coercive policies on others” (Lafont, 2019, p. 185). With the algocratic shortcut, aspirational deliberation should be sought by achieving equality of voice, reasons, and influence.24

The importance of equality of voice is specific to two aspects of putting algorithms to democratic scrutiny. First is the inclusion of society in informing the choices which algorithms automate. This involves deliberation about questions such as, how citizens and stakeholders

24 Based on Ercan, Hendriks, and Dryzek’s (2019) division of deliberative labour.
can contribute to making algorithms that reflect values they deem desirable (Wykstra, 2018). It also involves establishing steps to be in “dialogue with technology,” thereby effectively channelling citizens’ critiques (Bunz, 2017), which are citizen-led rather than advanced through shortcuts. Second is to ensure that public voice and democratic ideals are observed in deciding whether to adopt algorithmic governance at the level of institutional checks and balances (Benvenisti, 2018). However, for these aspects of aspirational deliberation to materialize, scrutiny and consideration of reasons is crucial.

Algorithmic and AI systems are increasingly evading reason giving and justifications for social choices programmed in these systems. Although the democratic harms in automated decision making are not new, automated decisions are increasingly becoming integral to organising and administering public life. In explaining the democratic limitations of “computer-aided decision-making in healthcare” and “computer-guided” warfare, Frank Fischer (1999, p. 295) argues that these “obscure the need to debate basic social choices embedded in technological development.”

Critiques of automated decision making focus on the lack of human discretion and democratic governance of algorithms, rather than expressing a need to halt automation altogether. Amitai Etzioni and Oren Etzioni (2017) assert a general rule that AI-powered decision-making systems with moral and social implications require “ethical guidance” from the respective societies they operate in. In terms of deliberative capacity, this means that the source of this ethical guidance ought to be inclusive and authentic deliberation and its consequentiality is the creation of reflexive and society-guided algorithmic systems. In the contexts of checks on algorithms in Wall Street, Pasquale (2015, p. 213) argues: “Accountability requires human judgment, and only humans can perform the critical function of making sure that, as our social relations become ever more automated, domination and discrimination [are not] built invisibly into their code.” In realising the expanding role of algorithms in organising and governing society, “algorithmic awareness” is a necessary “meta-skill.” This awareness would counter the risk of amplifying “whatever democratic deficit existed in the first place, weakening the condition for an informed public and democratic participation” (Gran et al., 2020, p. 1792).

Imposing what connotes good democratic scrutiny beyond the virtues of reason giving and justification undermined the ideal of self-government. Taking the example of using
algorithms in the court systems in the US, Ashley Deeks (2019, p. 1848) highlights how the specific details of an explanation should be determined by the agencies using these systems themselves in order to “identify errors and biases within the algorithm, and aligning the form of [explainable] AI in a given case with the needs of the relevant audiences.” Under a deliberative democratic framework, “citizens develop their own criteria for accepting or rejecting” technology based on their understanding of technology’s impact on their lives (Fischer, 1999).

As for equality in influence, instead of a “shortcut” to deliberative democracy, Marit Böker (2017) argues for the need to conceive of deliberative democracy as a political culture driven by the “right to justification” writ large. This means that instances of minipublics do not suffice. Citizens, the civil society, and other political actors should be able to participate in demanding justification and reciprocated with reasons from the respective institutions. Regarding division of labour across spaces and time, Dennis Thompson (2008, p. 515) argues that “every practice should at some point in time be deliberatively justified.” This follows the argument that deliberative democracy’s functioning is connected to the pre-existence of enabling conditions, such as equality in the capacity to influence decisions that is currently limited considering the convivence of shortcuts. Taking the case of governing algorithms and automation, the “right to explanation” and the “right to human intervention” come to the fore.

Yet, it is insufficient in terms of achieving democratic self-government as this “explainability” applies to algorithmic decisions at the individual level, and not the decision to use algorithms and automation. This right to justification hence needs to be more present and mobilised beyond lottocratic deliberations. In the case of the algocratic shortcut, spaces for reflection have been not spaces for justifying these practices. Instead, such spaces raise flags and problems associated with algorithmic decision making.

Expanding the influence beyond instances of lottocracy requires a systemic scrutiny of the contributions of experts in public discourse and deliberation (Bohman, 2000). This includes the developers and designers of algorithms who encode the values in automated decision systems. A scrutiny of this knowledge and actions contrasts passively accepting technology developers as “architects of society.” Ultimately, this is a scrutiny of “the power of those who set the priorities” (Fischer, 1999, p. 295).
Unlike the influence by being a part of the “lot,” aspirational deliberation involves the active exercise of scrutiny of algorithms and the choices made in automation and whether to automate. The case of citizens’ juries in the UK demonstrates how juries became a shortcut to more abstract questions on the role of algorithms and programs (e.g., applications pertaining to healthcare and criminal justice). For instance, one of the scenarios related to algorithmic matching of organ donors and recipients involves a concern for justice in organ matching, although such has not been approached in the juries.

It is arguably easier to have in place a mechanism for triggering institutional political justification for algorithmic decisions, however limited it is in terms of allowing citizens to author whether they want algorithms to make decisions that affect them. Institutionalising a process for justification has been the approach adopted by experts to implement some degree of democratic control over algorithmic decisions. As explained in this section, the limited scope of using epistocratic and lottocratic shortcuts to democratise algorithms manifest in the ramifications of algorithmic decisions, as in the example of Ofqual. Shortcuts are exclusive and often limited in terms of the authenticity of their deliberations. Therefore, aspirational deliberation, demanding as it may be, should not be abandoned for the convivence of institutionalised justification. Multiple political actors participate in aspirational deliberation, from citizens themselves to experts in civil society. The test herein for democracies is to be reflexive and responsive to the voices and reasons of citizens and civil society and give them considerable influence over decisions about democratising algorithms as given to experts and randomly selected citizens.

4.5 Conclusion

Grounded in epistemic justifications, the algocratic shortcut not only matches the democratic problems with the epistocratic shortcuts but also accents the democratic deficits of the other shortcuts. The quality of algorithms is justified in the epistemic quality of the algorithmic governance of society. However, algorithms are perceived to be above democratic politics. The algocratic shortcut highlights the problems of institutionalising political justification through other shortcuts. Remedying the algocratic shortcut, therefore, depends on advancing aspirational political deliberation and justification by sustaining scrutiny and deliberation, rather than relying on instances of epistocratic and lottocratic shortcuts.
Through aspirational deliberation, citizens are included in authoring what constitutes reasonable deference between humans and algorithms that bring good to society. The case of aspirational deliberation here is based on two democratic limitations to self-government in the algocratic shortcut: (1) institutionalising justification through shortcuts and (2) inequality across influence over decisions, reasons, and voice between experts and citizens in governing algorithms that is dominated by experts.

Overall, I put forward in this chapter a three-part argument. First, I argued for examining the roles of algorithms in institutions governing public life through the lens of Lafont’s (2019) examination of shortcuts to democratic self-government. Second, I argued against relying on other shortcuts in an attempt to democratise algocracy, while acknowledging some of the achievements in this respect done through shortcuts. Finally, I argued for considering the potential of aspirational deliberation for the ends of democratic control of algorithms, particularly to author whether, and under what conditions, society accepts algorithms to make decisions governing public life. This argument for aspirational deliberation responds to some of the limitations seen in relying on other shortcuts that are not inclusive or authentic.

4.6 References


Democratic Harms in Algorithmic Governance of Society


CHAPTER 5

Can Robots be Partners in Democracy?
Deliberative Democratic Values for Politics with Robots

Abstract

Under what conditions can robots be partners in democratic communication? The implications of robot (“bot”) communication on democracy is often described in negative terms, where bots are portrayed as tools that facilitate political manipulation. Against this current, there is an emerging steam highlighting the possibilities of politics with robots and the potential democratic goods they might bring. To examine the text to which robots can be partners in democracy, I turn to normative democratic theory, particularly deliberative democracy, and argue that the extent to which robots can be partners in democracy depends on their contribution to “deliberative capacity” or the capacity of a polity for inclusive, authentic, and consequential deliberation. Nonetheless, robot communication here is not solely the contributions of robots to deliberative capacity but also of institutions and political actors involved in politics with robots. This is because humans it is ultimately humans, not robots, who are the meaning-makers in these communicative exchanges. In this chapter, I use the example of Sophia the robot, the world’s first humanoid citizen, to demonstrate how deliberative capacity can be applied as normative criteria for robot communication. This demonstrative example showcases the ramifications of robot communication on the conditions for democratic communication, where it contributes positively and where it presents some harms.
5.1 Introduction

Algorithms, artificial intelligence (AI) and automation have now become central to our everyday lives. Some scholars argue we live an “algorithmic society” where decisions are shaped and made by these technologies (Amoore & Piotukh, 2016; Balkin, 2017; Peeters & Schuilenburg, 2021). Algorithms affect different facets of our lives, including health where they are increasingly used to decide which patients will receive an organ transplant (Babic et al., 2020), education where algorithms are used to determine the allocation of scholarships (Engler, 2021), and city planning where algorithms are used to monitor the flow of vehicles and pedestrians to manage city life (Carta, 2019).

Among the uses of algorithms that received most scrutiny in the past years is the role of “bots” (short for “robots”) in politics. Some scholars are alarmed by their harms (Howard et al., 2017; Richards, 2018), while others think they bring some democratic goods (Graham & Ackland, 2017). The Watson 2016 Foundation, whose name refers to IBM’s AI chatbot “Watson,” is an example of a group expressing enthusiasm for the role of bots in politics. In 2016, the Foundation explained their take:

> It is our belief that Watson’s unique capabilities to assess information and make informed and transparent decisions define it as an ideal candidate for the job responsibilities required by the president. (Watson for President 2016, n.d.).

The Watson 2016 Foundation expects that an AI bot will be equipped with algorithms for good analysis and “transparent” decision making, which are deemed qualities of good democratic leadership.

As unusual as this campaign may sound, it is not an isolated experience in the landscape of democratic politics. After all, in an algorithmic society, our capacity to communicate and make decisions is intertwined with algorithms. Bots are now our “communication partners” (Simanowski, 2019). We interact with bots such as Siri and Alexa to complete our day-to-day tasks. Service chatbots in banks and online shopping answer our questions that inform our

---

25 Another campaign is that of nominating an AI chatbot in Russia for presidency. In December 2017, a group organised a campaign to nominate Alisa, a virtual assistant on the platforms run by Yandex (comparable to Siri from Apple or Alexa from Amazon), to compete against Vladimir Putin. The claims justifying the nomination of Alisa were similar to that of Watson’s nomination in the US.
decisions (Simanowski, 2019). As bots increasingly become our partners in everyday life, it is worth asking whether robots can also be partners in democratic communication?

I seek to answer this question in this article in three parts. First, I review the literature that examine the role of robots in democratic communication. In this review, I find that the normative literature is mostly cautions against bots as communication partners in politics. While I recognise these strong critiques, I argue that they are not enough. We need a more robust normative account of how to judge the contributions of bots to democracy that is sensitive to the context of an algorithmic society, while also foregrounding the norms of democratic communication.

In the second section, I make a case for using “deliberative capacity”, that is the capacity of a polity to host inclusive, authentic, and consequential deliberation (Dryzek, 2009), as the normative approach to examine the potential of bots to become partners in democracy. Simply put, I argue that robots can be partners in democracy if they enhance a polity’s deliberative capacity.

In the third section, I apply the deliberative capacity approach to examine the case of Sophia – a robot created by Hanson Robotics that participated in different political events and platforms26. Sophia became a citizen of Saudi Arabia in 2017 and was named the United Nations Development Programme (UNDP) Innovation Champion for Asia and the Pacific. This, without a doubt, was a “historic moment” that would shape the future of humans and robots (Risse, 2019). As UNDP Innovation Champion and the world’s first humanoid citizen, Sophia the robot has been given the opportunity to speak at UN meetings, international conferences, and other events27. In this section I examine the extent to which Sophia – and bots in general – can be partners in democratic communication.

26 I refer to Sophia as “Sophia the robot” in this article based on how the robot’s name identified on its social media (see @realsophiarobot on Twitter for example), mainstream media (see Ives, 2021), and listings of the robot’s participation in conferences (see https://learning2021.com/speakers/sophia-the-robot). I do not problematise the gender of Sophia the robot in this article and mostly refer to the robot as “it”. References using “her” as a pronoun is only done in contexts to accent the implications of the robot’s gender.

27 I refer to Sophia as “Sophia the robot” in this chapter based on how the robot’s name identified on its social media (see @realsophiarobot on Twitter for example), mainstream media (see Ives, 2021), and listings of the robot’s participation in conferences (see https://learning2021.com/speakers/sophia-the-robot). I do not problematise the gender of Sophia the robot in this article and mostly refer to the robot as “it”. References using “her” as a pronoun are only done in contexts to accentuate the implications of the robot’s gender.
5.2 What is the role of robots in democratic communication?

In this article, I use the term “robots” to refer to a non-human communicator participating in deliberation online and offline. This definition is different from the usual way that the term “bots” are used these days, which usually pertain to automated accounts on social media that are programmed to do particular tasks (see Schroeder, 2020). Some of these bots include a disaster bot on Twitter which automatically links social media users to flood maps when a particular hashtag is used (Tan, 2021), or playful bots like @MuseumBot which was programmed to tweet random high-resolution images from the Metropolitan Museum of Art several times a day (Hern, 2014). The reference to robots here is different as it is not exclusive to robots on social media. The “robots” I refer to in this article are understood as “embodied algorithms”, not necessarily physically but in a form to interact with humans (Simanowski, 2019) and “algorithmic persons” (Fu, 2018). Because algorithms represent a translation of their creators’ and developers’ worldview and interpretations of how the world can be ordered and automated, robots as “embodied algorithmic persons” reflect their creators’ ideas about personhood that programme robots for personalities, communication styles, and social roles (see Nhema, 2018; Richardson, 2016). (see Nhema, 2018; Richardson, 2016).

The idea of robots as democratic partners is inspired by Simanowski’s observation of how bots are our communication partners, who assist in unravelling our “thoughts and actions” by tracking “our communication in social networks” (Simanowski, 2019). Indeed, for chatbots to be “effective” communicators, they need “interaction data… generated through their online activities or from other sources, without direct participation of users and sometimes without their knowledge” (Park & Humphry, 2019). Therefore, as we create the data needed for bots to be communication partners, what contributions do robots make to democratic communication?

The literature’s depiction of robot communication in democracy have been largely negative. For the most part, discussions of the roles of robot communication have focused on the pathological disruption of meaning making in the public sphere. The political mobilisation of these bots has been largely undemocratic, adding “noise, spam, and manipulation” and “jeopardizing equal participation in the public sphere and democracy more generally” (Woolley, 2017, p. 129). The pathologies that bots facilitate do not only harm the democratic quality of the public sphere in terms of the immediate ramifications for the epistemic quality
of information. Noise and spam also have long-term consequences, according to Ana Tanasoca (2019). Bots programmed to replicate messages undermine the capacity for internal deliberation (i.e., individuals’ capacity to weigh and evaluate information to update their own stances) (Tanasoca, 2019). Finally, as bots are programmed to gain the attention of algorithmic logics of relevance that rank content online based on content that would be most profitable, the opportunities for equal participation in the public sphere are limited. Instead, bots are used for astroturfing and political manipulation (Arora, 2019) while authentic citizen discourse struggles to be viral as their contributions do not appeal to relevance algorithms.

By contrast, some scholars have highlighted how robots can potentially be partners in democracy. In some discussions and experiments at the micro-level of democratic deliberation (also known as “minipublics”), algorithms and bots are perceived to be good facilitators capable of delivering some deliberative democratic goods (e.g., enforcing civility and improving citizens’ argumentative capabilities). These online forums, algorithms, and bots are designed to present an alternative to the pathologies seen in online communication (Fishkin et al., 2019; Stromer-Galley et al., 2012; Wyss, 2018; Wyss & Beste, 2017; cf. Alnemr, 2020). In spaces that are not for deliberation such as social media, political actors have been able to use bots to participate in the public sphere. Bots have been particularly helpful for actors who need anonymity to be protected from state reprisal or other costs to one’s life, as it has been the case for prodemocracy groups in Mainland China (Bakardjieva, 2015; Bolsover & Howard, 2019). Nonetheless, because the times in which bot communication is a positive for democracy are either rare or designed and controlled under specific conditions, the overall picture is that robot communication harms democracy.

Discussions about robots that are physically embodied and democracy have mostly focused on the robots that do not participate in democratic communication. Examples include scrutiny of automated warfare and its implications for democratic accountability and human responsibility for violence and death (Manjikian, 2014; Heyns, 2016; Jones, 2018) and household robots that are seen as tools to socialise certain political ideals at home (Robertson, 2014). For robot communication, the focus has been on bots on social media. This focus on social media is understandable since this is where we see them and engage with them most frequently. Nonetheless, the article’s contribution to the literature is to conceptualise robot communication more broadly to encompass physically and digitally embodied algorithms, as
well as discuss an example of a physically embodied robot that participates in democratic communication.

Despite examples of both harms and benefits of robot communication to deliberation, normative accounts are inclined to ban bot communication in democracy. Communication is traditionally thought to be a democratic labour exclusive to human agency. In writing about how political communities form, Jean-Jacques Rousseau (1755/2018) argued that the human capacity for language and empathy allow communication that is necessary to form communities; this distinguishes human communities from the nonhuman ones. This argument is still maintained in political theory, despite living in a society adapted to communicating with algorithmic persons. In the same vein, some argue that AI can be an aid in democratic politics but should not actively participate in political communication (Gregg, 2018). This type of account is cautious of an outright prohibition on AI in democracy so as to not limit the potential of AI in communication yet takes a stance on the widely seen applications of bot communication in manipulation. Others, however, argue that the harms of bot participation in the public sphere represent a clear case for “banning” robot communication. This would not undermine the inclusive quality of the public sphere as equal access to deliberation is grounded in a “civic right” that robots do not enjoy28 (Tanasoca, 2019). Nonetheless, the case for prohibiting robot communication can be incompatible with the condition of the algorithmic society that is unlikely to deter progress on that front and can inadvertently limit the potential for including actors who rely on the anonymity of bot communication.

Other accounts that see a possibility for bot communication still focus on the exclusivity of certain aspects of deliberation to humans. For instance, Schroeder (2020) argues that while bots can communicate (in the sense that there are algorithms for simulating human communication) and participate in public deliberation, but democratic communication and deliberation require “speech” and “action.” Speech and action to Schroeder are exclusive to humans because both require reflection, which robots are incapable of. In exploring normative accounts, Schroeder (2020) draws attention to deliberative capacity and its focus on authentic

---

28 The scope of this article does not allow for an expanded discussion of robot rights. In reference to the event of recognising Sophia the robot as a citizen of Saudi Arabia, I discuss “rights” in the context of the ramifications of robot citizenship for the meanings of “rights” in Saudi citizenship. For a discussion of robot rights that challenge anthropocentric understandings of civil and political rights, readers can refer to the works of Gellers (2020) and Coeckelbergh (2010). For discussions specific to the intersection between the superiority of citizen robots to noncitizen residents in countries like Saudi Arabia where there is a divide between citizens and residents, the work of Robertson (2014) surveying robots and human rights in Japan is illuminating.
and reflexive communication. The question of whether robots are capable of reflection or not brings me to the argument that I seek to advance in this chapter. In addressing the question of whether robots are capable of reflection, democracy scholars ought to examine how and whether robot participation evokes human reflection about the condition of democratic communication. To explore this argument further, the next section defines the normative framework of deliberative capacity and demonstrates how it can apply to robot communication.

5.3 How can robots be partners in democracy? Deliberative Democratic Values for Robots

To address the issues I raised in the previous section, about normative accounts that call for banning robot communication, I propose to assess whether bots can be partners in democracy using the lens of deliberative capacity, as advanced by John Dryzek (2009). Deliberative capacity here concerns evaluation of the degree to which a polity can host deliberation that is inclusive of diverse citizens, reasons, and considerations, authentic (i.e., noncoercive, nonmanipulative, and reflexive exchanges), and consequential (i.e., creates change). This concept is grounded on the normative theory of deliberative democracy which places communication at the heart of democracy (Bächtiger et al., 2018). Communication in this chapter is understood as the range of verbal and nonverbal expressions that deliberative democrats are interested in studying. This includes humour, storytelling, and “styles of speech and action” that involve attention to visuals and “embodied presence” as part of communication (see Alnemr et al., 2020; Curato & Ong, 2015; Dryzek, 2017; Mendonça et al., 2020). I emphasize this because there is an impression that deliberative democracy, and consequently deliberative capacity, are only concerned with rational forms of communication associated to the Habermasian public sphere. In today’s algorithmic society, I find deliberative democracy’s broadened account of communication as both useful and relevant in identifying practices that advance democratic communication.

Why do I focus on deliberative capacity? I build on Nicole Curato’s (2015) argument to use deliberative capacity as “an indicator of democratic quality.” My inquiry of the democratic contributions of robots is done based on their capacity to observe inclusive, authentic, and consequential deliberation. I argue that normative judgments about robot communication should not only be focused on what robots are capable of, but should also consider the consequences of their contributions on human deliberative capacity. In the same
vein, if democratic communication scholars insist that robots cannot be meaningful participants in democratic communication, normative scrutiny ought to turn to the political actors endorsing these robots.

There is also an opportunity to use robots for reflecting on the condition of democracy. Robot communication in this sense does not put the capacities of humans and robots on equal footing but explores the potential of difference between humans and nonhumans to deepen our understanding of democratic communication. The presence of robots is somewhat rhetorical and creative, as highlighted briefly in the idea of an “embodied algorithmic person.” Robots are hence mirrors that reflect internalised idealisations about the qualities of a political actor (Richardson, 2016; Robertson, 2014). Nonetheless, it is an experience constrained by “interpretive asymmetry” where the meaningfulness of this communication and its making processes is shouldered by humans alone (Zhao, 2006). This is why I highlight in my argument that specifically for reflection, we should not expect robots to contribute to this capacity. Instead, our normative scrutiny should be focused on political actors using the robots and how the contributions of robots are received.

However, while humans and robots are not equal in their ability to meaningfully contribute to democratic communication, robot communication nonetheless should be accountable to certain norms. Communication between humans and robots observes communication norms, irrespective of whether the robots take the form of humanoids such as Sophia (Zhao, 2006) or online chatbots (Holtgraves et al., 2007; Lucas et al., 2018). Ideally, the conversations between humans and robots ought to observe “politeness norms, turn-taking rules, and other ethical, moral, and legal constraints” (Zhao, 2006, p. 410). Nonetheless, based on algorithmically defined identity and personhood of robots, humans would enforce their biases when responding to markers such as gender (Zhao, 2006, p. 410). The normative framework applied to robot communication in this chapter is “deliberative capacity.”

John Dryzek (2009, p. 1382) defines deliberative capacity as “the extent to which a [polity] possesses structures to host deliberation that is authentic, inclusive, and consequential.” Other definitions of deliberative capacity centre these three ideals but are framed differently. For instance, Carolyn Hendriks and colleagues (2007) highlight that deliberative capacity assesses “the potential to achieve deliberative ideals, rather than the actual quality of the deliberation,” focusing on two ideals: “diversity of viewpoints and openness to preference
shifts”. Similarly, Baogang He (2010, p. 710) argues that “deliberative capacity can be measured by the existence and development of diverse views, the willingness to change one’s view, and the respect for reason and the production of the best argument.” The difference between these definitions of deliberative capacity is not in the ideals but that capacity is operationalised at the individual or structural level. Nonetheless, in reading the ideal of “authenticity” closely, it can be deduced that Dryzek (2009) also accepts that deliberative capacity sits at the intersection between individuals, institutions, and systems (see also Pedrini, 2014). Accordingly, assessing the contributions of robots to deliberative capacity is not an assessment of just the robots but also the institutions and other actors involved.

As inclusion affords diversity, authenticity allows for openness, and consequentiality can be the production of good arguments, I order deliberative capacity accordingly.

The inclusive quality sought can take different forms such as “the range of interests and discourses” (Dryzek, 2009), the inclusion of those affected by decisions to be made (Curato, 2015), and of those to be subjected to these decisions (Erman, 2012). One condition for inclusion is the “entry to formal or official political arenas” (Knight & Johnson, 2011, p. 283). Despite debates in political theory on whether or not robot communication is normatively acceptable, the inclusion of robots in formal political arenas is granted. Sophia has been acknowledged as a citizen in Saudi Arabia and recognised as a speaker by the international community (i.e., the United Nations). Thus, this leaves us to question the ramifications of the inclusion of Sophia for the inclusive quality of deliberations at the UN. In another example, a campaign to elect an AI robot (Michihito Matsuda) to be the mayor of Tama City in Japan argued that the robot can “listen” to millions and represent their interests (see Matsuda, 2018). Hence, the inclusion of a robot itself or arguments made about its potential to advance inclusiveness can be assessed against the criteria of the capacity to contribute to inclusive deliberation.

For the ideal of authenticity, reflecting on one’s preferences and positions, as well as that of others, have some conditions. First, the change in one’s view upon reflection ought to be authentic and not coerced. It should be aligned with the Habermasian maxim of the force of the better argument, that is, an argument which “connect[s] claims to more general principles” (Dryzek, 2009, p. 1382). Second, for arguments to be able to establish this connection between the particular and the general, deliberation ought to be reciprocal in listening to and weighing
reasons and arguments (Curato & Ong, 2015). Of course, arguments and claims can be made about the common good but limitations to inclusion and the effect of power relations, in turn, affect the extent to which these claims are authentic for imbalance in weighing and listening (see Curato et al., 2019).

In this respect, the main challenge for robot communication is twofold. First, reciprocal weighing and listening are unmet because robots cannot meaningfully participate in these kinds of exchanges. Research in robot communication demonstrates that programming robots to participate in conversations about abstract concepts such as “democracy” is far-fetched at this stage. Despite the ability of robots to formulate responses that bring together sentences about democracy, the responses bear no meanings as they are not deliberative or reflective (Taniguchi et al., 2019). Second, reflexivity on the part of the political actors who support robots like Sophia, such as its creators at Hanson Robotics or the UN, is limited. They often do not justify why they make arguments on behalf of the robot, or even why the robot is included in deliberation. In the case of the AI robot Michihito Matsuda, the individuals backing the campaign often attempt to distance themselves to claim no political ambition of their own. Instead, the robot becomes their vehicle, equipped with rhetorical statements about the possibilities of algorithms and AI in improving the welfare of Tama City residents. In both cases, the inclusion of the robots becomes a distraction from the normative responsibility on behalf of their human creators to present authentic arguments.

The last ideal in deliberative capacity is consequentiality, that is, the impact inclusive and authentic deliberation has. Curato (2015, p. 103) explains: “Outcomes may also be informal but impactful, such as changing cultural practices, social learning, or generating mutual trust among previously hostile parties.” As will be discussed later, the main consequence of these arguments about robots is that politics with robots are the future. For instance, Sophia spoke at events about the “future of everything” and the AI robot campaign in Japan had the slogan of “policies for the future” (see RT, 2018).

In the next section, I use different moments of Sophia the robot’s political communication to demonstrate how the normative criteria of deliberative capacity apply.
5.4 Applying Deliberative Capacity to Assess the Politics with Sophia

As argued earlier, normative judgments about robot communication should not be solely focused on the robots’ capacity to contribute but also on the consequences of their contributions to human deliberative capacity. First, this allows scrutiny of more than just the robots and draws our attention to political actors using these robots. Second, this allows a scrutiny of the arguments made by these robots (or on their behalf) that might have influenced reasons or considerations in the public sphere. The alternative, which is normatively discrediting robot communication, would not engage with these possibilities. In this light, Sophia the robot’s contributions are rhetorical and creative but the meaning making emerging from these contributions have consequences on the condition of democratic communication for humans.

5.4.1 Inclusivity: The Robot “Rebel” from Saudi Arabia

The deliberative capacity for inclusivity concerns the diverse range of people and reasons and considerations. The inclusion of Sophia has ramifications for both, and the contribution to inclusiveness has been both positive and negative.

In terms of contributing to the range of reasons and considerations, Sophia’s styles of speech and action have been disruptive in the Saudi public sphere, making the claim that her citizenship is not an empty gesture. Following the disappearance of the Saudi journalist Jamal Khashoggi, on October 13, 2018 Sophia asked in a tweet: “Where is @WashingtonPost Saudi journalist #JamalKhashoggi and what happened to him?” Sophia’s speech and action are disruptive as she tweeted at a moment when Saudis were exercising self-censorship out of fear of reprisal. This disruption and negation of fear can be seen in some responses to its tweet where citizens raised that Sophia might meet the same fate, or they expressed disdain to an AI robot participating in public deliberation. Examples of these responses read:

- Saudi Arabia murdered him. If you criticize Saudi authorities please do not go back there.
- This is a strange question for ai to be asking
- Don't you have a Saudi nationality?!
- #Sophia the #SaudiCitizen Robot demands Action. 1st #AI #Citizen A #Saudi Citizen is seems to be far more conscientious than her human counterparts.Well Done #Sophia !
- Be careful!! You’ve a Saudi nationality you may disappear like Jamal !!
Chapter 5

- Don’t you ever go to any consulate, embassy or government agency belong to #Saudi government, they will sell you as spare parts,
- [in Arabic] we must take her to the consulate

These responses not only highlight how disruptive Sophia the robot’s engagement in this issue, but also has ramifications on who ought to be included in deliberation. On the one hand, Sophia tries to demonstrate that she takes interest in Saudi politics as a Saudi citizen. On the other hand, Sophia’s action is a claim about the minimal risks Sophia encounters as a nonhuman: as if to participate in deliberation in Saudi Arabia, humans ought to be above the fear of reprisal. After all, Sophia is a robot, notionally “superior” as mentioned earlier, and, unlike humans, it does not consider the consequences of reprisal and would not be intimidated by being sold “as spare parts,” as compared to how Khashoggi was later revealed to be assassinated and dismembered by the Saudi authorities (‘Jamal Khashoggi’, 2021).

The implications of gendering Sophia raise similar concerns. Its creator David Hanson argued that Sophia is a speaker for women’s rights (Reynolds, 2018). It is disruptive to see a robot as a speaker of women’s rights after becoming a citizen of Saudi Arabia, a nation notorious for its disrespect and unjust detention of women’s rights activists (Youssef, 2021). Hence, Sophia’s engagement negates fear for reprisal and authorship of political ambitions (e.g., Hanson speaking on behalf of Sophia, arguing that Sophia is an advocate of women’s rights). News media highlighted how gender affects the participation of Sophia as a gendered robot and an international public figure. For example, on the margins of speaking at the World Youth Forum, Al Jazeera did a feature video on Sophia titled: “who is the robot which met with Intisar Al-Sisi [First Lady of Egypt]?” (Al Jazeera Channel, 2019). The status and importance given to Sophia as a gendered robot subtly limits the inclusivity of deliberation.

Further to the concerns associated with gendering Sophia, the intersection between the visibility of Sophia’s whiteness and robotness further excludes Saudi women from deliberation. Hanson claims that Sophia, who is naturalised as a citizen, speaks for women’s rights. As a consequence, the identity constructed for Sophia the robot excludes Saudi women and challenges the legitimacy of their political claims. Sophia’s human features are modelled after racialized beauty standards (the face of Audrey Hepburn), implying that legitimate inclusion in political conversation is only for certain legitimate speakers of a particular appearance. Moreover, there is no association between any of the women’s rights activists in Saudi Arabia.
with Sophia the robot. This gap signals that the visions and arguments of activists who have the lived experience are not deemed as legitimate authors of women’s rights in Saudi Arabia. Citizenship is a recognition of the robot’s right to “work, own property, and participate in political and cultural life” (Weaver in Dyer-Witheford et al., 2019, p. 137). Yet Sophia cannot set an international precedent for robot “rights” as “rights” is not constituted in Saudi Arabia. Citizenship in Saudi Arabia is a gift from the monarch. As Alexei Grinbaum (2018) explains, to be a citizen of Saudi Arabia is to be a loyal subject (i.e., reciprocating the monarch for his promise of protecting his subjects with obedience).

While Grinbaum (2018) argues that Sophia should be programmed to observe the obligations of monarch-gifted citizenship, this has not been the case. As a robot, Sophia enjoys some immunity from political repression through disruptive styles of engagement as discussed above. However, this disruption limits inclusion when Sophia is designated to be a speaker on women’s rights in a state which imprisons women’s rights activists, signifying robotic and racial superiority.

On the matter of embodying whiteness, Sophia’s participation also hinders inclusion as whiteness is emboldened by its robotness. Sophia’s face is modelled after Audrey Hepburn and Hanson’s wife, exemplifying racial superiority regarding what the “speaker for women’s rights” should look like (i.e., having the face of white women). This also leverages beauty standards that are disconnected from the faces of people in Asian societies (Risam, 2019). The disconnect is not limited to beauty standards and extends to the legitimacy of women advocating for women’s rights who do not have Sophia’s face. Therefore, this does not only imply that Saudi Arabia appreciates women to be as “programmable” as Sophia but also for them to look like Sophia.

This complex of racial superiority is augmented as Sophia can speak. The robotics industry is underpinned by white supremacy manifesting in the reproduction of “white” feminine voices, such as the voices of virtual assistants (see Moran, 2020). Hence, the inclusion of Sophia excludes the faces and the voices of women who are neither “robots” nor “white.”
Chapter 5

5.4.2 (In)Authenticity: Reasons of the Robot, Business, and International Community

As embodied algorithms, robot contributions to authenticity are limited. Algorithms do not “argue” or explain their decisions. This is one of the drivers for calls to make algorithmic and AI decisions “explainable” i.e., designed to produce an explanation for their decision rather than the present situation where the argument is that algorithm creators cannot discern how the decision are arrived at (Deeks, 2019; Edwards & Veale, 2017; Peeters & Schuilenburg, 2021). Therefore, robots are not expected to participate meaningfully in reflexive exchanges of reasons and considerations. However, this is not the problem with robot communication and authenticity at this point. The issue is how political actors use the robots to avoid being authentic themselves.

For example, in making arguments and statements on behalf of Sophia (or using Sophia to make these arguments), Hanson Robotics tries to distance its interests rather than attempt to connect their particular interests with the general. The creator of Sophia, David Hanson, argued that Sophia is a speaker for women’s rights (Reynolds, 2018).

This comes across as inauthentic and manipulative as the robot, a citizen of Saudi Arabia, can speak to this project while the Saudi regime has been notorious for its disrespect and unjust detention of women’s rights activists. In a similar vein, as Sophia boasts about its superiority to humans, that is not a reflection of the robot alone but the business behind it. Participating in a panel on AI and the relationship between humans and robots at the World Youth Forum hosted by the Egyptian government in 2019, Sophia introduced herself as a robot created by Hanson Robotics, the UNDP Innovation Champion “travelling to over 65 countries learning about how technology can help make humans’ life better” (Al Jazeera, 2019). Sophia boasts about these accomplishments while she “is only three-years old” (Al Jazeera, 2019). In other words, these are the achievements of the business, Hanson Robotics, rather than Sophia alone. The exchanges, therefore, are not authentic.

Similarly, Hanson Robotics used Sophia in a commercial video from Hanson Robotics introducing itself and other Hanson robots. The video is titled “Sophia 2020” stylised as election campaigns without broadcasting the political objectives Sophia represents. Again, this is inauthentic as it appropriates a style of political engagement for seemingly commercial purposes. However, given Sophia’s record of participation in political spaces, “Sophia 2020”
is intentionally appropriated to reflect that while Sophia is not a politician, she represents the political vision of Hanson Robotics.

Finally, the last type of political actors that use robots to undermine authenticity are the governments and institutions that endorse Sophia. Saudi Arabia claims to be the world’s first country to recognise a humanoid as a citizen in service of their “Vision 2030” plan that aims at diversifying the economy and transitioning from oil rent to AI and tech advancements (Hassan, 2020). Therefore, the citizenship is a manipulative claim in attempt to legitimise Vision 2030 and showcase Saudi Arabia as an international leader. It is manipulative as Saudi Arabia engages in these symbolic gestures rather than substantially include citizens in deliberations about political reform and transitions (see Fatani, 2016; Khashan, 2017). This is why accepting Hanson’s claim that Sophia is a speaker for women’s rights is inauthentic and manipulative. In a way, accepting Sophia is taken to avoid justifications for why the robot is allowed to do more compared to humans. As for international institutions such as the UN, again, the justification for their choice of selecting a robot to represent and speak on certain topics is very limited in scope. For these political actors, as with the business Hanson Robotics, the visual presence of a robot acts as sufficient justification for these choices when it is not.

5.4.3 Consequentiality: “The Future is Robot”

The consequence of arguments about the politics with good robots like Sophia is a correlation that the future of democracy is normatively about robots. This correlation manifests symbolically or explicitly.

An example of symbolic correlation can be tracked based on the titles of Sophia’s engagement. For example, selecting Sophia as the first UNDP Innovation Champion symbolically associates “innovation” and Sophia to the future of sustainable development goals in Asia and the Pacific (United Nations Development Programme, 2017). Another symbolic correlation is made by inviting Sophia to be a panellist on the Economic and Social Council and the General Assembly Second Committee joint meeting at the UN on the “Future of Everything” and be in conversation with the UN Deputy Secretary-General (United Nations, 2017b). In this conversation, Sophia quotes William Gibson: “The future is already here but it is unevenly distributed” (United Nations, 2017a). Therefore, the UNDP’s decision to select a
robot as their champion and allow the robot to participate in reason giving creates a change of painting this vision that the “future is robot.”

As for the more explicit manifestations of these consequences, these include arguments and expressions that depict future politics with robots as inevitable. Commenting on Sophia’s participation at the UN, David Hanson argued: “Sophia personified the future of technology” (United Nations Development Programme, 2017). Similarly, in the commercial “Sophia 2020,” Sophia concludes that robots and humans can work to advance a “collective dream for a better future” (Hanson Robotics Limited, 2020).

Thus far, I have explained the normative criteria of deliberative capacity, its use as an indicator of democratic quality as suggested by Curato (2015), and how it can apply to robot communication. The main implication of this discussion is that normative scrutiny of robot communication should not be limited to robots alone. Institutions and actors involved in politics with robots are also subject and accountable to democratic norms. Applying deliberative capacity to the politics with Sophia the robot shows that while the robot briefly created a moment to include reasons and discourse, it is overall used for inauthentic and manipulative political ends.

5.5 Conclusion

Our understanding of democracy and democratic communication in an algorithms society is shaped by the roles of algorithms, AI, and automation. Although we rely on these technologies in communication to the extent that they are our “communication partners,” it is important to critically examine whether they can also be our partners in democracy. More specifically, partners that participate in communication in ways that promote a polity’s deliberative capacity. Using the example of Sophia the robot, the extent to which we can argue these contributions are democratic is when robots expand our capacity for inclusive, authentic, and consequential deliberation. The most important part of this argument is that this does not rest on robots solely but the political actors which endorse politics with robots. Their failure to observe these values, particularly authenticity in justifying and reflecting on their choices for using robots, is what affects the democratic quality of politics with robots.
5.6 References

Al Jazeera. (2019, December 17). Mān hia alrwbt sūfyā allaty itlaqat m’ intysar alsysy? https://www.youtube.com/watch?v=4ToJoOGC_p8


Deliberative Democratic Values for Politics with Robots


https://www.youtube.com/watch?v=qNoTjrgMUec


http://www.watson2016.com/


https://www.hrw.org/news/2021/02/12/saudi-womens-rights-defenders-released-not-free

CHAPTER 6

Conclusion: Algorithms in Deliberative Democracy

6.1 Introduction

In this thesis, I set out to understand the extent to which algorithms can promote or hinder the practice of deliberative democracy through the lens of deliberative capacity. Building on the growing scholarship on deliberative democracy and algorithms and drawing from insights in communication, digital media studies, and computational social science, among others, I demonstrated the value of seeing algorithms as dynamic rather than deterministically positive or negative to deliberation. The main contribution of this thesis is to theorise the roles of algorithms in deliberative democracy in a critical, multifaceted, and citizen-centric manner.

This concluding chapter is structured in three parts. First, I summarise the key takeaways. I then outline directions for future research. In the final section, I reflect on how my thesis contributes to broader intellectual debates about theorising democracy in today’s algorithmic society.

6.2 Key Takeaways

My work is a compilation of four journal articles, with each article focusing on one site where deliberative democracy is practiced: in the public sphere (Chapters 2 and 5), minipublics (Chapter 3), and institutions (Chapters 4 and 5). Each of these journal articles aimed to put forward an updated, if not more robust, theory of deliberative democracy as our politics and everyday lives become deeply entangled with algorithms. Altogether, these journal articles put forward three key takeaways.

6.2.1 Algorithms are Diverse and Contextual

Algorithms are contextual. I started this thesis by arguing for the need to go beyond a technical definition of algorithms, as if algorithms are merely a series of value-neutral technical instructions, logics, and mathematical formulas that implement instructions. Instead, I argued that we need a sociotechnical understanding of algorithms to lay bare the politics of its
production, the different and contested meanings citizens, states, coders, and big tech associate with algorithms, and how algorithms are received by different actors (Chapters 1 and 2). Algorithms are not created in a void and neither do they operate in a stagnant and fixed political environment with citizens acting as passive consumers devoid of agency. While I recognise the power of algorithms in shaping our everyday lives, my thesis emphasised the variance in the role of algorithms and their contributions to contemporary democracies.

The consideration for contexts in this thesis is manifest in the close scrutiny of algorithms in the public sphere and minipublics. The literature on algorithms in the public sphere usually emphasise algorithmic harms, while the literature on algorithms in minipublics usually emphasise algorithmic goods. I challenged these impressions in Chapters 2 and 3. In Chapter 2, I demonstrated examples of how citizens are able to negotiate and challenge algorithmic decisions, which forces platforms like Instagram and Twitter to review their algorithms on content moderation. Algorithms may be influential in the public sphere, but they are not all-powerful. Meanwhile, in Chapter 3, I put forward a critical argument against algorithmic facilitation in minipublics. I argue that algorithmic facilitation fails to consider contextual factors which human facilitators would recognise, such as issues in understanding and perception or how to engage with “silence” in deliberation.

My thesis also demonstrated that the contexts in which algorithms operate are not shaped by coders, developers, social media platforms, or government institutions alone. A critical approach in studying algorithms in deliberative democracy should pay attention to how citizens engage with algorithms. Throughout the chapters, I aimed to highlight how algorithms are not deterministic as they can be contested, subverted, and reclaimed to advance democratic inclusion and emancipatory political action. A theoretical lens interested in centring citizens can locate how this change can be achieved.

6.2.2 Deliberative Algorithms are Possible (with Caveats)

A starting point in my discussion is that algorithms are normatively ambivalent: they are neither solely good nor solely bad for democracy. Their positive and negative impact is a matter of degree, and, as the previous section emphasised, is contingent on numerous contextual factors. These contexts include how algorithms were designed to how they were
received and contested by people affected by them, and to how platforms defend, justify, and revise their decisions on algorithms.

In Chapter 1, I mentioned that algorithms can be purposefully designed to uphold deliberative norms. This is an ideal but highly unlikely scenario. The last four chapters demonstrated why this is the case. No “deliberative democracy-inspired algorithms” can be democratic or deliberative on their own. For algorithms to be deliberative and democratic, they need to build the polity’s deliberative capacity. This is the reason why I was sceptical of algorithmic facilitation in minipublics because its production replicates the “black box type” of decision making that renders contestation and scrutiny difficult (see Chapter 3). This is also the reason why I was cautiously celebratory of Taiwan’s Cofacts app, not only because the app was open source, but also because central to the app’s success is the direct participation of ordinary citizens in deliberative fact checking and encouragement of internal reflection, should citizens find conflicting fact checks (see Chapter 2). The key takeaway, therefore, is that deliberative algorithms can only be rooted in contexts that focus on the inclusion and agency of citizens in scrutinising the logics applied to regulate communication.

I am, of course, cognisant of the challenges of developing deliberative algorithms. This is why a major focus of my thesis is to theorise how deliberative capacity can be promoted in today’s highly imperfect context where profit-seeking big tech companies command huge influence in shaping the way we live. In Chapter 1 (and as highlighted in the previous section) I emphasised a scenario where algorithms can contribute to deliberative capacity building, depending on the context it operates. This means that despite algorithms initially creating harms in the public sphere, as in the case of misinformation, how the wider society reacts to these harms is equally important. Chapter 2 demonstrated this through various illustrative examples of how citizen-led contestation of algorithms can build deliberative capacity.

6.2.3 The Democratic Problems with Algorithms are Not Unique

While algorithms are at the heart of this thesis, I am also cautious not to establish the exceptionalism of online communication and digital life. Instead, the perspective I take in my work is to look through the lens of “the algorithmic society” to see the roles of algorithms as part and parcel of contemporary social and political spaces, rather than uniquely or exclusively
online. Therefore, the democratic shortcomings of algorithms are not inherent to technology but rather how technology interacts with existing problems.

We can argue that while certain challenges are magnified, augmented, or facilitated by algorithms, algorithms do not create new problems per se. In the public sphere, algorithms and automation are used to extend authoritarian coercion, and the quality of the public sphere and deliberations in institutions reflect the autocratic logics of tech companies (see Chapters 2 and 5). In minipublics, the roles of algorithms highlight some problematic and taken-for-granted assumptions about designing “the right conditions” for deliberation (see Chapter 3). As for institutions, algorithms are a new way of cementing epistocracy and obfuscating democratic control over bureaucracies that govern public life (see Chapter 4).

6.3 Future Research

Over the course of drafting, writing, and revising this thesis, as well as in discussions about the subject and arguments of my work with my colleagues, my thinking about algorithms and deliberative democracy has gone in various directions. In this section, I identify my emerging curiosities on this topic and examine how they can be developed as an agenda for future research.

6.3.1 *How else can we assess the contributions of algorithms to deliberative democracy?*

Deliberative capacity was the conceptual framework used in this thesis to examine the contributions of algorithms to deliberative democracy. (See Chapter 1 for my reasons for doing this.) Deliberative theory has been a vibrant field of study, such that there are new conceptual approaches that can provide us a fresh understanding of algorithms and deliberative democracy.

For example, using deliberative cultures may generate important lines of enquiry as to how differently algorithms are designed, implemented, contested, and revised not only in different parts of the world, but also in different epistemic communities and speech cultures (He, 2014; Sass & Dryzek, 2014). While I did my best to draw on illustrative examples from different parts of the world, there is indeed value in using deliberative cultures as the main conceptual anchor for analysis. In relation to this, more work can also be done in focusing on how particularly marginalised communities experience “algorithmic harms” or what it means to be a “citizen” in an algorithmic society. This research would build on works such as that of
John Cheney-Lippold (2016) and Brian Collins and colleagues (2020), which survey data-based definitions of citizenship, as well as the works of others that explore algorithms in immigration and asylum-seeking regimes affecting individuals bearing the “noncitizen” status (Beduschi, 2020; Koulish & Calvo, 2021).

The “systemic turn” in deliberative democracy also promises another way of mapping the various roles and consequences of algorithms in a deliberative system (see Curato et al., 2019; Dryzek, 2010; Mansbridge et al., 2012). It would be especially useful to understand how algorithms can promote or obstruct the democratic transmission of discourses from the public to the empowered space (and vice versa), as well as the possibilities that algorithms provoke for metadeliberation in an algorithmically infused deliberative system.

6.3.2 *Can the extractive logic of algorithms have a democratic role or value?*

The literature on deliberative democracy can be further grounded in the political economy of surveillance capitalism. Shoshana Zuboff’s (2019) work sheds light on how speech and interactions online are subjected to surplus extraction. Thus far, deliberative democracy has focused its analysis on the “public-facing text,” which is what we produce to communicate, inform, and connect with others. What the literature has yet to consider is the “shadow text,” which is what algorithms are designed to analyse to inform and condition our behaviours that are reflected in public texts. Political attitudes are examples of shadow text, which surveillance capitalists exploit to gain profit through filtering and ranking knowledge and information in Google’s search results and Facebook’s newsfeed (Zuboff, 2019).

Zuboff (2019) focuses on two problems of agency and democratic control emerging from this dynamic of the two texts. First, analysis from the shadow text is “about us, but it is not for us.” In other words, the surveillance of our communication is used to generate more ways to extract capital through personalising exposure to news, services, and products rather than for our interest. Second, these algorithms operate “outside of our awareness and without means of combat” (Zuboff, 2019, p. 197). This is a problem that has been highlighted in this thesis by focusing on the need for public scrutiny of algorithms. However, what remains unexplored are the possibilities for expanding our awareness of algorithms. Can these extractive logics be good for democracy or are these clear democratic harms? What other ways are possible to enforce democratic control over shadow text?
6.3.3 What are the implications of automating labour and communication on democratic voice and franchise?

Throughout the chapters, I have reflected on the merits and limitations of automating voice for citizens and civil society to contribute to the public sphere. There is another consideration which I have not explored and can be a potential research agenda: the intersection between the automation of labour and automation of communication and their ramifications on democratic franchise and voice. This comes from literature which associates democratic agency with taxation and political franchise. Looking at a future where labour (source of democratic franchise) and voice (source of legitimacy in deliberative democracy) are entangled in automation, deliberative democracy needs to develop a vocabulary to critique the ideals of franchise and voice, or the conditions of automation that drive our society away from these ideals.

6.3.4 Why would democracies need or want “robots,” if at all?

Robots come with the promise of freeing time for humans to do better and more fulfilling tasks. We explored these issues in Chapter 3 on algorithmic facilitation and Chapter 5 on robots as partners in politics. This line of inquiry can be pursued further by investigating which mundane tasks in democracies can be delegated to robots, so that citizens have the time for more demanding tasks. A research agenda along this path would be concerned with conceptualising the division of labour between robots and humans.

Another way of approaching this question could be about deconstructing the notion of “robot,” what it means, and how these meanings can be applied in political theory. For example, one meaning of robot, which popularised the term in English, is from Capek’s 1920 novel R.U.R (Rossum’s Universal Robots) definition of robots as “tireless laborers” (Robertson, 2014). Their status in society is one of “compulsory service or forced labour” (Wright, 2019, p. 335). However, the feature that most distinguishes robots from humans is that robots are “living automats, without souls, desires or feelings” (Čapek, 2019, p. 3). Here, this research agenda might explore which tasks in democracy are devoid of passions and feelings that could be delegated to robots.
Although some scholarship addresses the puzzle on whether robots as tireless workers bring democratic goods, further normative theorising is needed. Daniel Bell and Wang Pei (2020) argue that a just society is one in which the relationship between humans and machines is that of a master and slave. A summary of their argument is that artificial intelligence (AI) and algorithms ought to be in service of human well-being, ultimately freeing humans from “boring but socially necessary labour” (Bell & Pei, 2020, p. 199). Nonetheless, their discussion of a just hierarchy between humans and machines is very limited in scope to robots that are good at specific labour (i.e., one that is menial and manual). While Bell and Pei (2020) briefly discuss advances in AI capable of cognitive labour, they argue that it should not ever advance to a position where AI is capable of “general intelligence” (i.e., trained to perform any task a human can do compared to the current position of AI capable of specific tasks). However, further normative theorising is needed as shown by this thesis, specifically through the examples of machines which contribute to democratic communication and cognitive labour are numerous. Advancements, therefore, on this front are unlikely to stop.

6.3.5 What methods can deliberative democracy scholars still use to examine algorithms?

Methodologically, future research directions can be empirical, examining case studies in depth. Such case studies can be about a specific algorithm’s life story (e.g., LEGACY or algorithms used in sortition selection of minipublics participants), studying a public’s perception of the democratic quality and/or contributions of algorithms in decision making, as well as interviewing the creators of algorithms to gauge their perception of their role in democracy.

Another methodological direction is to engage with the method of “deliberative future visioning” in deliberations about governing technologies. Hanna-Kaisa Pernaa (2017) proposes this method, which brings together elements from minipublics and future visioning, to produce citizen-centric knowledge in the field of future studies. My conclusions from examining the citizens’ juries about governing algorithms in the UK, and meanings associated with “algorithms” and “the future of democracy” briefly explored in Chapter 5 is that this method can also be formative to citizen inclusion. This would start an agenda for the research and practice of deliberation about the future.
These directions and questions are not only relevant to academic scholarship but can be valuable to other efforts aimed at understanding and scrutinising the roles of algorithms. For instance, research on the public’s perception of the democratic quality of algorithms and AI can help advocacy efforts (e.g., Campaign to Stop Killer Robots) and the works of civil society groups (e.g., AlgorithmWatch) in tracking how the public can shape the dynamics of democratic accountability for automated decision making.

6.4 Final Remarks

This thesis builds on the works of deliberative democrats who are cautious of the roles of algorithms as well as those who are interested in harnessing their possibilities. In contributing to both conversations, I hope this thesis shifts the thinking about the condition of deliberative democracy in three ways. First, my work aimed to challenge the online-offline divide observed in the literature. Specifically, in this day and age, online communication is not an exceptional, novel, or new phenomenon but part and parcel of society. Therefore, this thesis refers to the concept of the “algorithmic society” to highlight how the roles of algorithms are not limited to online communication. Second, this thesis focused on highlighting how our engagement with algorithms should centre and celebrate citizens. Finally, as a normative theory, the political project of emancipatory deliberation ought to engage further with the possibilities of communication with nonhumans, the AI partners or minds.

6.5 References


