



Critics argue that opaque filtering algorithms used by online web services such as Google or Facebook are trying to improve accuracy at the expense of diversity, leading to so-called filter bubbles. As a result, users never see the other side of an argument, viewpoint diversity diminishes and civic discourse is undermined. This thesis first analyzes internal processes and factors in online web services that might cause filter bubbles. Later, it extends the metrics used in viewpoint diversity research and presents the results of an empirical study for Dutch and Turkish Twitter users. The thesis also analyzes software tools and design attempts to combat filter bubbles. It demonstrates that almost all of the tools implement norms required by two popular democracy models. The thesis argues that democracy is essentially a contested concept, and that other less popular democracy models should be included in the design of such tools as well.



---

# BURSTING THE FILTER BUBBLE: DEMOCRACY, DESIGN, AND ETHICS



---

ENGIN BOZDAĞ

---

# **BURSTING THE FILTER BUBBLE: DEMOCRACY , DESIGN, AND ETHICS**

## **Proefschrift**

ter verkrijging van de graad van doctor  
aan de Technische Universiteit Delft,  
op gezag van de Rector Magnificus prof. ir. K. C. A. M. Luyben,  
voorzitter van het College voor Promoties,  
in het openbaar te verdedigen op  
woensdag, 16 September 2015 om 10:00 uur

door

**Engin BOZDAĞ**

Master of Science in Technische Informatica  
geboren te Malatya, Turkije.

Dit proefschrift is goedgekeurd door:

Promotors:

Prof. dr. M.J. van den Hoven

Prof. dr. ir. I.R. van de Poel

Copromotor:

dr. M.E. Warnier

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. M.J. van den Hoven	Technische Universiteit Delft, promotor
Prof. dr. ir. I.R. van de Poel	Technische Universiteit Delft, promotor
dr. M.E. Warnier	Technische Universiteit Delft, copromotor

Independent members:

dr. C. Sandvig	Michigan State University, USA
Prof. dr. M. Binark	Hacettepe University, Turkey
Prof. dr. R. Rogers	Universiteit van Amsterdam
Prof. dr. A. Hanjalic	Technische Universiteit Delft
Prof. dr. ir. M.F.W.H.A. Janssen	Technische Universiteit Delft, reservelid



*Printed by:* CPI Koninklijke Wöhrmann

*Cover Design:* Özgür Taylan Gültekin

*E-mail:* [engin@bozdag.nl](mailto:engin@bozdag.nl)

*WWW:* <http://www.bozdag.nl>

Copyright © 2015 by Engin Bozdağ

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission of the author.

An electronic version of this dissertation is available at

<http://repository.tudelft.nl/>.

# PREFACE

For Philip Serracino Inglott,  
For his passion and dedication to Information Ethics  
Rest in Peace.

*Engin Bozdağ*  
*Delft, August 24, 2015*

# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	2
1.2	Problem Statement . . . . .	8
1.3	Research Methodology . . . . .	9
1.4	Contributions . . . . .	11
<b>2</b>	<b>Bias in Algorithmic Filtering and Personalization</b>	<b>13</b>
2.1	Introduction . . . . .	14
2.2	Information overload and the rise of the filters . . . . .	15
2.3	Personalization – a technical overview . . . . .	17
2.4	A model of Filtering for Online Web Services . . . . .	20
2.4.1	Source Selection Algorithm . . . . .	21
2.4.2	Information Selection and Prioritization Algorithm . . . . .	22
2.4.3	Human operator . . . . .	23
2.4.4	Personalization Algorithm . . . . .	25
2.5	Discussion . . . . .	29
2.5.1	Implications for an ethical analysis . . . . .	29
2.5.2	Implications for design. . . . .	32
2.5.3	Implications for the design of social filtering. . . . .	32
2.5.4	Implications for social network analysis . . . . .	34
2.6	Conclusion . . . . .	35
<b>3</b>	<b>Viewpoint Diversity in Online Social Networks - An Empirical Analysis</b>	<b>37</b>
3.1	Introduction . . . . .	37
3.2	Empirical Studies of Information Diversity in Social Media . . . . .	39
3.3	Theory . . . . .	40
3.3.1	Information Diversity . . . . .	40
3.3.2	Dimension of Information Diversity . . . . .	41
3.3.3	Minorities and Openness . . . . .	42
3.4	Polarization in the Netherlands and Turkey . . . . .	42
3.4.1	the Netherlands . . . . .	43
3.4.2	Turkey . . . . .	43
3.4.3	Conclusion. . . . .	45
3.5	Method . . . . .	45
3.5.1	Data collection. . . . .	45
3.5.2	Research questions . . . . .	45
3.5.3	Entropy . . . . .	46
3.5.4	Translating Research Questions into Metrics. . . . .	46

3.6	Results . . . . .	48
3.6.1	Distribution of seed users and their followers . . . . .	48
3.6.2	Seed User Bias . . . . .	49
3.6.3	Source and Output Diversity . . . . .	49
3.6.4	Minority Access . . . . .	50
3.6.5	Input Output Correlation . . . . .	50
3.7	Limitations . . . . .	52
3.8	Discussion . . . . .	52
<b>4</b>	<b>Democracy, Filter Bubble and Design</b>	<b>55</b>
4.1	Introduction . . . . .	55
4.2	Democracy: different theories, different benchmarks . . . . .	56
4.2.1	Liberal View of Democracy. . . . .	57
4.2.2	Deliberative Democracy . . . . .	58
4.2.3	Republicanism and Contestatory Democracy . . . . .	59
4.2.4	Agonism / Inclusive Political Communication . . . . .	61
4.2.5	Conclusion. . . . .	62
4.3	Software Design to Combat Filter Bubbles . . . . .	65
4.3.1	Liberal / User Autonomy Enhancing . . . . .	65
4.3.2	Deliberative / Enhancing Epistemic Quality of Information . . . . .	68
4.4	Discussion . . . . .	71
4.5	Conclusion . . . . .	77
<b>5</b>	<b>Ethics of Online Social Experiments</b>	<b>79</b>
5.1	The Facebook Emotional Contagion Experiment . . . . .	79
5.2	The Unclear Ethical Terrain of Online Social Experiments. . . . .	80
5.3	Arguments For and Against Online Social Experiments . . . . .	81
5.3.1	Benefits of Online Experiments for the Individual . . . . .	82
5.3.2	Informed Consent and Its Many Interpretations . . . . .	83
5.3.3	The Ubiquity of Online Social Experiments . . . . .	84
5.3.4	Different Perceptions of Risk in Online Experiments . . . . .	85
5.3.5	Benefits of Online Experimentation for the Society . . . . .	85
5.3.6	The Unavoidability of Online Experiments. . . . .	86
5.4	Discussion . . . . .	87
<b>6</b>	<b>Conclusion</b>	<b>91</b>
6.1	Answers to the Research Questions and Revisiting the Objectives of the Study . . . . .	92
6.2	Limitations of the Study. . . . .	94
6.3	Future Research. . . . .	95
6.3.1	Viewpoint Diversity Research . . . . .	95
6.3.2	Transparency . . . . .	97
	<b>Acknowledgements</b>	<b>101</b>
	<b>Summary</b>	<b>103</b>
	<b>Samenvatting</b>	<b>107</b>

---

<b>Curriculum Vitæ</b>	<b>111</b>
6.4 Scientific Publications . . . . .	111
6.4.1 Journal Articles. . . . .	111
6.4.2 Conference Proceedings . . . . .	111
6.5 Professional/Popular Publications . . . . .	112
<b>References</b>	<b>113</b>

# 1

## INTRODUCTION

*The peculiar evil of silencing the expression of an opinion is, that it is robbing the human race; posterity as well as the existing generation; those who dissent from the opinion, still more than those who hold it. If the opinion is right, they are deprived of the opportunity of exchanging error for truth: if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth, produced by its collision with error.*

John Stuart Mill, philosopher, 1859

Should we go to war? How high should the taxes be? What happened at Gezi Park in Istanbul? These are all political questions where different sources would give a different answer. With the advent of social media as an important source of news and opinions [183, 230], some activists and scholars have started worrying that the Internet could lead to online segregation and may increase radicalism and extremism in society, due to receiving biased and one-sided news and opinions. Critics have pointed out the dangers of group forming among like-minded in Internet. Recently, online platforms such as Facebook and Google have been criticized, because with their opaque personalization algorithms they show users viewpoints that they already agree with, hence leading to information silos, or so-called filter bubbles. The reason why filter bubbles have been criticized differs. Some argue that the opaque algorithms used by online platforms make decisions on behalf of the user, coercing him and making him unaware of available choices. Others argue that biases caused by algorithms and human beings themselves might diminish viewpoint diversity, decrease respect toward one another or allow oppressors to prevail due to a lack of information to the citizens, which will prevent them reaching the truth or only one side of the truth. Viewpoint diversity has long been viewed as an essential component of strong democratic societies [82, 150]. Yet others, including Google and Facebook, have argued that the effects of personalization have been exaggerated [26, 249, 352]. The immediate question that comes to mind is whether filter bubbles really exist. However, to answer this question empirically and properly, one must first define what filter bubble is. Do online platforms really have biases that may cause bubbles? Do people themselves construct self-reinforcing filters because they already

have divergent beliefs? This interdisciplinary thesis aims to answer such questions by studying filter bubble phenomenon and the relevant value viewpoint diversity 1) at the conceptual level by using theories from by political philosophy, media and communication and ethics 2) at the empirical level by analyzing filter bubble by extending viewpoint diversity analytics metrics 3) at the technical level, by analyzing tools that are designed to combat filter bubbles.

## 1.1. BACKGROUND

According to a study in the US, nearly half (48%) of the panelists say they accessed news about politics and government on Facebook alone in the past week [230]. More than twice as many Americans are following political candidates on social media in 2014 as was the case during the 2010 midterms, indicating that these platforms are playing an increasingly prominent role in how voters receive political information and follow election news [230]. The results of another study show that among the so-called X generation (born between 1966 and 1980), Internet news audience jumped from 29 percent to 49 percent between 2004 and 2012 [183]. It now matches turning to TV for news, which also declined by 20 percentage points over this period. Similar patterns are apparent among Millennials (born between 1980 and 1995), but they are more extreme. More of those now turn to the Internet for news than to TV (43% versus 35%). A more recent study indicates that 86% of the Millennials usually turn to social media to receive diverse opinions, more than any other media [13]. It is also reported that social media users increasingly express their opinions about the world, national and local news. Between 2010 and 2012, the traffic to news sites from various social media platforms grew 57% [262]. Media companies are starting to get more and more dependent on Facebook for traffic. Vox, for instance, gets 40 percent of visits through Facebook. Other leading news organizations get around a quarter of site visits from the social networking platform [194]. The New York Times and BuzzFeed confirmed plans to begin hosting their work directly on Facebook using a feature called “Instant Articles” [54, 225]. Other studies indicate that online information intermediaries such as Facebook and Twitter are prominent platforms where users follow news about governments and politics [230]. These platforms are thus slowly replacing traditional media channels thereby partly becoming the gatekeepers of our society [45, 83, 294]

These popular platforms are often seen as revolutionary participatory information production and consumption infrastructures that help to overcome barriers to information. This is due to Internet’s high reach [123], diversity and interactivity [286]. According to this optimistic view, news and political information will not be produced and disseminated by only a small number of elites. It will disrupt the power of the elites and will allow a more inclusive discussion, because everyone can voice his or her opinion. The supposition is that there are no or fewer gatekeepers, thus more and diverse viewpoints will be represented online than in offline media [359]. Minorities, the disadvantaged and others who could not utter their voices will be able to do so in new media [53]. Many scholars and journalists describe the online media landscape created by search engines and social media platforms as the “age of plenty”, with “an almost infinite choice and unparalleled pluralization of voices that have access to the public sphere” [66, 126, 174]. “Because of its horizontal, open, and user-friendly nature, the Internet allows for easy

access to, and thus greater participation, in the public sphere” [47, p.144].

These inclusive platforms then promote equal access to diverging preferences and opinions in society and allow people to be exposed to diverse viewpoints. Some further claim that, since the control and power of the elites in traditional media do not work in new media, journalists can communicate virtually anything to anyone [196]. Social network sites such as Twitter are seen as platforms where anyone can affect the news-creation process [159]. In short, in this optimistic view, online intermediaries allow anyone to be able to contribute to the public debate and reach a large public to be able to read anything published by anyone, without any barriers. In the end, Internet and particularly social media will have a positive impact in online participation and democratization due to the personal and interactive nature of the online platforms.

However, such an optimistic view of the Internet as an endless information platform with a diverse set of viewpoints has received many criticisms. Cyber-pessimists argue that online platforms are not that different from traditional media, as the same small set of elites still determine how the political information is obtained, filtered and presented [333]. Redden and Witsche [2009] argue that the use of the Internet for political communication is minor. This is because consumerism, entertainment, non-political networking and online chat dominates the online sphere, not every political item gets attention and political items do not spread widely. This claim is supported with Facebook’s latest study where the researchers found out that only 7% of the content users click are “hard news” [24]. Further, on platforms such as Facebook the degree of involvement might differ per story and even if users do participate in political communication, they only get involved in the last phase of ‘traditional’ news production process by interpreting texts and commenting on them, and not at the decision-making stages of news production [274]. Witschge’s [2008] analysis on Dutch fora shows that even though an alternative or radical voice can be expressed online, it is not successful in opening up a dialogue. Instead, the participants were trying to find ways to exclude it. Others emphasize the important role of opinion leaders and those in higher social hierarchy in shaping others’ perceptions during a conflict and argue that not everyone has the same right to be heard [206]. Papacharissi [2002] argues that online discourse often ends in miscommunication and therefore cannot directly enhance democracy. Hindman [2008] and Sunstein [2007] argue that Internet is not a public forum due to the ease of only reading and linking to like-minded resources and deliberating among like-minded users without hearing contrary views. Due to this homophily, social fragmentation and group polarization will follow which will lead users to even more extreme positions. Others have provided evidence to support this claim [3, 72, 155, 314]. In short, cyber-pessimists argue that social media can lead users to deliberate among their own groups.

While many scholars have argued that users may create their own filters and deliberate among like-minded groups, Pariser [2011] focused on opaque and hidden filters used by the online platforms themselves. In his book “The Filter Bubble: What the Internet Is Hiding from You”, Pariser [2011] mentions that online platforms personalize the information per user, depending on user’s previous interaction history with the system, his preferences and his contacts in social media. The consequence of such social filtering is that results in a search engine for the same query differ per user and two people with the same friends in a social network might see different updates and information. This,

Pariser argues, might create a monoculture, in which users get trapped in their “filter bubble”. Pariser [2011] writes, “personalization filters serve up a kind of invisible auto-propaganda, indoctrinating us with our own ideas, amplifying our desire for things that are familiar and leaving us oblivious to the dangers lurking in the dark territory of the unknown.” [p. 13]

Pariser points out that such “social filtering” may ultimately undermine civic discourse by confirming our existing ideas and limiting our exposure to opposing viewpoints. These algorithms can cause citizens to be ill-informed about current events and the citizens may have increasingly “idiosyncratic perceptions” about the importance of current events and political issues. This might occur because online services are trying to improve accuracy at the expense of serendipity and diversity, often in order to increase user retention. Pariser argues that even if users wanted to diversify their network explicitly, information intermediaries silently filter out what they assume the user does not want to see, hiding information posted by opposite end of the political spectrum. Pariser believes that the algorithmic gatekeepers need to show us things that are not only easy to consume but also things that are challenging, important and uncomfortable and present competing points of view.

After Pariser’s book led to lengthy public discussions and academic debates, researchers wanted to find out whether filter bubbles really exist. Some of the studies support the filter bubble theory. An et al. [2014] found indications of partisan sharing on Facebook. Eslami [2015] found that majority of the studied Facebook users were not aware of algorithmic filters and were very upset when they found out close friends and family were not shown in their feeds. Pentina and Tarafdar [2014] found out that especially those who have a small network on Facebook are vulnerable to the filter bubble effect. Nikolov et al. [2015] studied 4 years of communication data in a university network and found out that social media exposes the community to a narrower range of information sources. Hoffman et al. [2014]’s study shows that users avoid news sources that belong to the opposite end of the political spectrum. Conover et al. [68] found that retweet network in American Twittersphere is highly polarized, while the mention network is not. Kim and Park [2012] provide evidence of polarization in Korean Twittersphere. Liu and Weber [2014] argues that Twitter is not an “idealized space for democratic, rational cross-ideological debate”, because individuals who do not have many followers does not interact much with their ‘ideological-foes’, and tend to attack opinion leaders with many followers if they choose to interact with them. Further, duration of discussions among like-minded users last longer and more than 40% cross-ideological tweets were disagreements [205]. Yom-Tov et al. [2013] studied reading pattern of Bing search engine users and observed that majority of the Republicans and democrats click on sources that belong to their viewpoint, and rarely click on sources that belong to the other side. Gruzd and Roy [2014] found that, in Canadian Twittersphere, people do tend to cluster around shared political views and political polarization exists. Further, 40% of the messages between ideologically distant parties were hostile or negative [145]. Colleoni et al. [67] found out that bubbles exist in the US Twittersphere, however structures of political homophily differ strongly between Democrats and Republicans, and between general users and those who follow the political parties. Barbera and Rivero [29] studied American and Spanish Twittersphere and found out that users participating in the political discussion were mostly men, liv-

ing in urban areas, and with strong ideological preferences. They also found out that users who follow political parties directly are more active in the political discussion than the rest [29]. Freelon et al. [110] studied Syrian Twittersphere and observed high fragmentation among different communities. Smith et al [2014] tracked one hashtag related to America's budget conflict over two days. They found "two big and dense groups that have little connection between them" [299]. Grevet et al. [138] performed a survey with 103 politically engaged Facebook users and found out that weak ties (contacts who are not very close to the user and not like-minded) offer the most diversity, yet users engage less with with them and they can break under pressure from contentious discussions. Gromping [143] analyzed Facebook posts in 2014 Thai general election and found that partisan social media users hardly engaged with opposing viewpoints at all. Hahn [151] studied South Korean Twitter users and revealed polarization and the reinforcement of existing political divisions in society.

However, others, including the online platforms themselves, disagree with Pariser. Google claims that they actually limit personalization and promote variety for certain topics [353]. Facebook's first study argued that people get exposed to novel content through their weak links and therefore bubbles do not exist [26]. More recently Facebook studied filter bubble for self-reported liberals and conservatives and founded out that the News Feed does ideologically filter what we see [24]. In its press outreach, Facebook has emphasized that "individual choice" matters more than algorithms do. Yet, Facebook argues that this effect is modest in comparison to the choices people make that filter information, including who their friends are and what they choose to read given the curation. Barbera [28] studied the ideological positions of users in Germany, Spain and the United States and found out that social media users receive information from a set of diverse sources, thanks to weak ties. Others argue that bubbles already occurred in traditional media as local newspapers reflected a narrow consensus and with the infinite choices on the Web, people will actually choose to consume a more diverse set of content [353]. O'Hara et al. [2015] argue that the evidence for bubbles is not strong enough for regulation and even if bubbles exist, users can escape them. Since users can live in looser and multiple networks (often thanks to social media), they have flexibility, choice and exposure to heterogeneous points of view. Weinberger [352] argues that; 1) the Internet is too young to make conclusions about filter bubbles; 2) the empirical research that exists is very difficult to interpret; 3) fragmentation occurs in traditional media and in offline world; 4) democracy needs bubbles so that people in basic agreement can build relationships and be active in political movements. Beam and Kosicki [35] investigated the impact of personalized news web portals on political bias and found out the average news viewer seems to favor news that does not have bias towards a particular perspective.

Since the term filter bubble has been uttered in 2011 by Pariser, it has received a broader meaning. While personalization bias is a recent phenomenon, algorithmic bias is not. Inclusivity issues of search engines, lack of transparency and other factors that cause bias have been discussed before [135, 168, 334]. As Chapter 2 will show, online platforms suffer from algorithmic and human biases. However, online platforms are not the only source of bias. Psychology literature, particularly "selective exposure theory", studies biases in human information seeking. Issues studied under selective exposure

include cognitive dissonance (psychological discomfort arising from one's partisan attachment and information challenging this attachment)[104], confirmation bias (seeking information for a fast conclusion or stopping seeking new information when a conclusion is reached) [288], avoiding information overload (by reducing sources and ideas, processing information faster or avoiding information seeking all together) [262, 306], avoiding challenging information (because it requires more resources to process[306], and bias in determining quality of information [105]. Further, Sunstein [2007] argued that, due to the availability of manual filters on the Internet and the option to communicate only with like-minded others, group polarization will arise and people will end in more extreme positions. Sunstein used the term "echo chambers" to conceptualize this group polarization [314]. Filter bubble is now also used to describe this behavior as well. [110, 138, 199, 202]. The term now not only encompasses opaque automatic cyberbalkanization imposed on users by the algorithms of the online platforms as emphasized by Pariser, but it also includes other non-automatic voluntary selective exposure and biased information seeking and group polarization. This gives us the situation depicted in Figure 1.2. Biases in design and operation of online intermediaries and biases caused by selective exposure (thanks to manual filters) lead to biased online platforms that we now define as "the filter bubble".

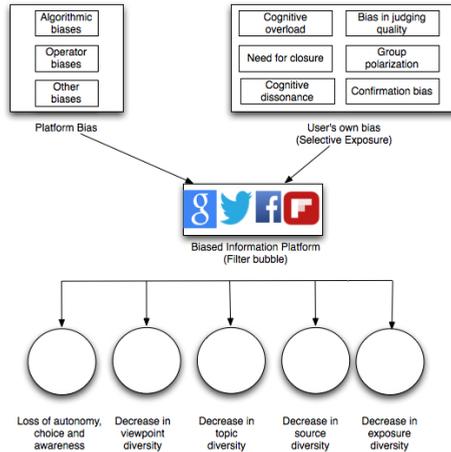
Filter bubbles are mainly criticized due to their two important consequences. First, opaque filters that are imposed upon the user, diminishes their autonomy [242, 256]. Since it is not clear that filters are there at the first place, the user might not know that the information stream is filtered after all. Further, even if the existence of the filters were known, since every user might get a different output of the system, it will be different for the user to compare what he is missing. A non-filtered version often is not available. A daily user of social networking platforms describes this as follows:

Never mind how much data you have about me. No, sorry, systems should not be making that choice for people. At least, people should be given the opportunity to opt-in or opt-out of that model, which in this case it's just not happening. And I am finding that incredibly frustrating and perhaps somewhat disturbing as well for that matter. Why? Mainly, of course, because of that filter bubble. I would want to be the person in charge of what I get exposed to, how I would want to get exposed to, and, most importantly, how I would want to consume that content shared across. And let it be down to me to decide if I would want to mitigate, or not, the fire hose effect of content I get exposed to

Secondly, it has been argued that, due to algorithms being designed for accuracy and short-term user satisfaction information diversity is lost [256]. Following Napoli's framework of media diversity [243], this leads to a decrease in viewpoint (idea) diversity, topic (content) diversity, source diversity and in the end exposure diversity (See Figure 1.1). As we will discuss in Chapter 4, a decrease in diversity of ideas, opinions and perspectives is undesired for almost all models of democracy. Following Rawls' concept of 'primary goods' [272], goods that are supposedly useful (or at least not harmful) to anyone, irrespective of their conception of the good, van den Hoven and Rooksby [2008] argued that information should be accepted as a primary good. Information online is

vital for people to plan their lives rationally and to participate adequately in the common life of their societies [344]. Thus, filter bubbles affect having access to information, which in turn affects the worth of liberty felt by an individual. Contrary to earlier stages of the Internet-era, when the problem information access boiled down to having access to hardware, nowadays the problem of access to information concerns the ability of intentionally finding the right information, or unintentionally stumbling on upon relevant information [46].

Figure 1.1: Filter bubble



The phenomenon of filter bubble has received much attention in news, academia and in the industry. For instance, due to the critique on social media, Facebook performed empirical studies to investigate whether bubbles occur [24, 26]. However, as we will discuss in Chapter 2 and 4, these studies use various and different notions of democracy (if defined at all) to analyze filter bubble. Filter bubble is a complex phenomenon that requires a good analysis of the underlying concepts, such as equal access, inclusiveness and autonomy. Existing theoretical work, for instance from computer ethics or media communication often lack scientific empirical backing. Further, the existing empirical work testing filter bubble in social media define filter bubble and the relevant value diversity rather implicitly. This leads to the conceptualization of the filter bubble as not “hearing the other side” [40, 228, 240] with two dominant sides or not being “exposed to the mainstream” [249]. However, according to media diversity literature, viewpoint diversity is not only about two main sides hearing each other or avoiding extremism by being exposed to the mainstream, but also about the radical voices and non-mainstream opinions to reach a larger public effectively. Therefore, different operationalization of viewpoint diversity can lead to different results in empirical work. Further, all work on filter bubble criticize the algorithms that the platforms operate on, however they do not delve into details of factors that causes the bubbles. Do the factors that affect traditional media also affect the new media? How do humans influence the online platforms next to algorithms? Are there other influences? What is the cause of the filter bubble? In this

thesis, we will aim to answer such questions.

## 1.2. PROBLEM STATEMENT

The main aim of this thesis is to reduce the knowledge gap on filter bubble between different disciplines and between the theory and practice. The main research question of this thesis can be formulated as follows:

**RQ:** How can we design for viewpoint diversity, so that filter bubbles are avoided?

To answer this research question, the following subquestions are defined:

**RQ1:** What is filter bubble and which factors lead to it?

While many works, including Pariser's (2011) book point out the dangers of online filters, they do not delve into details that may cause these issues. Further, Pariser's work does not refer to a rich literature on gatekeeping and search engine bias. Chapter 2 aims to contribute to the existing literature on gatekeeping theory and search engine bias by extending the traditional gatekeeping theory with algorithmic gatekeeping performed by online services including personalization. It shows that factors affecting traditional media, such as advertiser bias, personal judgments, or organizational bias are also present in online platforms. Further, it shows that bias is not only invoked by algorithms, but also by human operators of those platforms. It shows that switching from human editing to algorithmic gatekeeping does not remove all human biases.

**RQ2:** Can we conceptualize and operationalize viewpoint diversity?

There are many empirical work that studied source diversity [244], polarization [15, 24, 67, 68, 110, 145, 178, 205], media bias [14, 280, 351], user bias in search engines [368], URL novelty [26], and weak-links [24, 138]. These studies present conflicting results. Some claim that bubbles do not exist [24, 28], while others claim that they do [143, 151]. However, according to media and communication literature, viewpoint diversity is not only measured by the number of available viewpoints or the interaction between major political fractions, but also by the possibility of the minorities and other disadvantaged to reach the larger public. Chapter 3 of this thesis first introduces different dimensions of the highly complex value viewpoint diversity using political theory and provides an overview of the metrics used in the literature of viewpoint diversity analysis. Later, it also proposes new metrics using media ethics theory and provide a framework to analyze viewpoint diversity in Twitter for different political cultures. Finally, it presents the results for a case study on minorities that is performed for Turkish and Dutch Twitter users and shows that minorities cannot reach a large percentage of Turkish Twitter users. With the last of these contributions, using theory from communication scholars and philosophers, this chapter shows how minority access is missing from the typical dimensions of viewpoint diversity studied by computer scientists and the impact it has on viewpoint diversity analysis.

**RQ3:** Which issues does filter bubble cause for different models of democracy and what can design do to solve them?

Chapter 4 tries to answer these questions by studying different democracy models from political philosophy and shows how each one of them criticizes a different consequence of the filter bubble. This chapter summarizes most important norms of different democracy models in relation to viewpoint diversity. It analyzes various software tools and designs that are developed to fight filter bubbles. It shows that the aims and goals of these tools can differ radically, depending on the designer's understanding what filter bubble is and why it is an undesirable development.

**RQ4:** What are the ethical issues associated with those solutions?

Online platforms are conducting large scale experiments in order to combat filter bubbles [368], while some argue that those experiments themselves may cause bubbles [220]. In 2014, a controversy has occurred surrounding the so-called Facebook emotional contagion study [185]. In this study, Facebook wanted to test the assumption that basic emotions, positive and negative, are contagious, that is, that they spread from person to person by exposure. To do this, they have adjusted the News Feed of hundreds of thousands of users to randomly filter out specific posts with positive and negative emotion words to which they would normally have been exposed. The article provoked some very strong reactions both in the international news media and among scholars. Chapter 5 analyzes the arguments of two parties (data science advocates and data science critics) through a qualitative discourse analysis. Since similar studies are performed to fight the filter bubble by certain platforms, it is very important to discuss the relevant ethical values, including informed consent, transparency and autonomy.

### 1.3. RESEARCH METHODOLOGY

To answer the research questions in the previous section, we use Value Sensitive Design (VSD) methodology. VSD is "a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process" [116]. The word "value" here is defined rather broadly: "what a person or group of people consider important in life" [116]. A key feature of VSD is a tripartite methodology, consisting of iteratively applied conceptual, empirical, and technical investigations. Conceptual investigations include discovering the relevant values that can inspire or inform the design of a tool by using a list of important human values as a starting point (e.g. privacy, autonomy) [108]. This first step also includes using theories from other disciplines (e.g. philosophy, media studies, etc.) to conceptualize specific values and clarify any issues that can arise due to different interpretations of the same value by different disciplines or due to value conflicts (e.g., privacy vs security). Empirical investigations include the entire range of qualitative and quantitative methods used in social science research. This can include observations, interviews, surveys, or measurement of user behavior. Technical investigations focus on how existing technological properties can support or hinder human values. Technical investigations also involve translating the identified values by operationalizing, embodying, implementing and expressing them in system design [108].

Since years, a number of critiques and suggestions towards VSD have also been published. For instance VSD takes the position that certain values are universally held and

how they exactly play out can differ per culture, which is seen as problematic by many [10, 43, 73]. Others critique VSD's ethically neutral stance, leaving unclear what values and which theories it includes, thus making value conflicts unresolvable [9, 214]. This has the danger of committing naturalistic fallacy by taking empirical findings as given, not distinguishing mere preferences from moral values and implementing them into design directly without performing a normative analysis [214]. Borning and Mueller [2012] argue that a VSD researcher's own culture and assumptions may limit the qualities of authority and impartiality, as he might have biases due to his background, experiences and relation to the participants in the study. Similarly, others argue that VSD fails to address the use of deliberative methods and tools to promote joint reflection and stakeholder participation during the design process [43, 267]. Further, it has also been pointed out that the values listed by the founders of VSD might be too simplistic, as they stem from a particular culture and viewpoint (Western, upper class academics) and hence they should be contextualized and explicitly stated [43, 76]. Le Dantec et al [2009] argue that having a list of values may blind the researcher to values that fall outside that list.

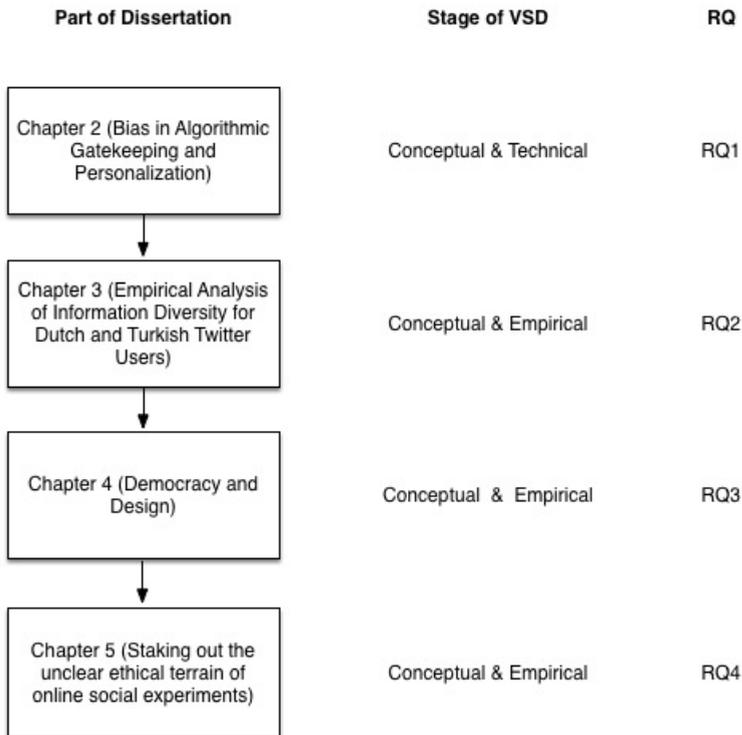
While VSD has its weaknesses, it has been an important part of computer human interaction field under computer science and lead to many influential findings and analyses [73, 75, 112, 115, 117, 118, 227, 229, 293, 362]. In this thesis, VSD is used, because the complex issues arising from the filter bubble require conceptual, empirical and technical investigations. First of all, there is a conceptual unclarity about the problem. Many people use the same word (filter bubble) to address the issue or devise tools to get rid of it, while they actually talk about different issues and even if they talk about the same issue (information diversity) they have different conceptions of this value. Second, some empirical studies were performed to find out if bubbles exist. However, since those studies did not have a conceptual investigation of this complex problem, the findings represent one particular interpretation of the filter bubble. This conceptual confusion has manifested itself in technical artifacts that all claim to "combat filter bubbles", yet do very different things, or solutions that are seen as unacceptable by some.

As critics have argued, starting with the list of values VSD has provided and then see which of these values are implicated in filter bubble would be the wrong approach. Instead, we have studied the publications, news and opinions on this topic and then curated a list of values [46]. From this list we have the following values identified: viewpoint diversity, freedom from bias, privacy, informed consent, autonomy, transparency and identity. In this thesis, we will focus on viewpoint diversity and freedom from bias, while also discussing implications for transparency.

Applying VSD gives us the steps depicted in Figure 1.2. Chapter 2 belongs to conceptual and empirical studies, because it analyzes the concept filtering, bias and gatekeeping using theories from media and communication studies. It also uses observations to identify factors that cause bias. Chapter 3 belongs to conceptual and empirical investigations as it uses theories from media and communication studies on viewpoint diversity and employs quantitative methods to measure user behavior in social media. Chapter 4 belongs to conceptual and technical studies, as it analyzes norms of democracy using theories from political philosophy and it studies different technical solutions that are developed to combat filter bubbles. Chapter 5 uses theories from various fields to address the value transparency and informed consent and analyzes arguments from

different users. Therefore it belongs to the conceptual investigations.

Figure 1.2: How Value Sensitive Design is applied in this thesis



## 1.4. CONTRIBUTIONS

This thesis contributes to the literature by doing the following:

Chapter 2 is, to our knowledge, the first work that captures various factors that can lead to algorithmic and operational biases in online services, including personalization.

Chapter 3 is, to our knowledge, one of the first works that studies filter bubbles empirically and the first work that operationalizes the concept “equal access” and the first work that compares input and output diversity. It shows the results of an empirical study to show bubbles may also happen due to user’s own grouping habits. It compares different political cultures and defines different metrics. It shows that minorities may become invisible in Twitter for certain political cultures.

Chapter 4, is, to our knowledge, the first work that analyzes the concept using different democracy models. It analyzes how the design attempts to fight filter bubbles differ fundamentally and how these designs can cause other ethical issues.

Chapter 5, is, to our knowledge, the first work that studies various arguments that are used in discussing ethical issues of experiments carried by online services. Filter bubbles

1

are either caused by those experiments, or are used to fight them.

# 2

## BIAS IN ALGORITHMIC FILTERING AND PERSONALIZATION

*A squirrel dying in your front yard may be more relevant to your interests right now than  
people dying in Africa.*

Mark Zuckerberg, CEO of Facebook

## 2.1. INTRODUCTION

Information load is a growing problem in today's digitalized world. As the networked media environment increasingly permeates private and public life, users create their own enormous trails of data by for instance communicating, buying, sharing or searching. The rapid and extensive travelling of news, information and commentary makes it very difficult for an average user to select the relevant information. This creates serious risk to everything from personal and financial health to vital information that is needed for fundamental democratic processes. In order to deal with the increasing amounts of (social) information produced on the web, information intermediaries such as Facebook and Google started to introduce personalization features: algorithms that tailor information based on what the user needs, wants and who he knows on the social web. The consequence of such personalization is that results in a search engine differ per user and two people with the same friends in a social network might see different updates and information, based on their past interaction with the system. This might create a monoculture, in which users get trapped in their "filter bubble" or "echo chambers" [256, 311, 313]. Social media platforms, search and recommendation engines affect what a daily user sees and does not see. As knowledge, commerce, politics and communication move online, these information intermediaries are becoming emergent gatekeepers of our society, a role which once was limited to the journalists of the traditional media.

The gatekeeping process is studied extensively by multiple disciplines, including media studies, sociology and management. Gatekeeping theory addresses traditional media bias: how certain events are being treated more newsworthy than others and how institutions or influential individuals determine which information passes to the receivers [298]. Gatekeeping theory does address the rising power of online information intermediaries, but it focuses on two things: a) the increasing role of the audience in which users can determine what is newsworthy through social networks b) the changing role of the journalist, from a gatekeeper to a gatewatcher [48, 294]. The existing theory often considers the online information intermediaries themselves as neutral or treats a web service only as an algorithm, operating without human bias [49, 159, 197]. Because these information intermediaries automate their core operations, often, mistakenly, they are treated as objective and credible. Machines, not humans, appear to make the crucial decisions, creating the impression the algorithms avoid selection and description biases inherent in any human-edited media.

Several authors have shown that computer systems can also contain biases. Friedman and Nissenbaum [1996] show that software can systematically and unfairly discriminate against certain individuals or groups of individuals in favor of others. Bias can manifest itself in a computer system in different ways; pre-existing bias in society can affect the system design, technical bias can occur due to technical limitations, emergent bias can arise sometime after software implementation is completed and released [113]. Several authors have shown how search engines can contain technical biases, especially in coverage, indexing and ranking [86, 234, 334, 348, 360]. However, these works are only focusing on the popularity bias. As we will show, many other factors can cause bias in online services.

In this paper we show that online services that process (social) data are not merely algorithms; they are complex systems composed of human operators and technology.

Contrary to popular belief, humans do not only take part in developing them, but they also affect the way they work once implemented. Most of the factors that cause human bias in traditional media still play a role in online social media. Finally, even though personalization is seen as a solution by some to prevent technical biases that exist in non-personalized online services [129], we show that personalization not only introduces new biases, but it also does not eliminate all of the existing ones. Others have already pointed to the dangers of implicit and explicit personalization in online services and traditional media [176, 256, 311, 346]. However, they do not identify the potential sources of bias, processes and factors that might cause particular biases. They also do not connect this debate to existing literature in gatekeeping and search engine bias. Our descriptive model of algorithmic gatekeeping aims to achieve this. As Goldman [2011] has recently written about search engine bias: “competitive jostling has overtaken much of the discussion. It has become almost impossible to distinguish legitimate discourse from economic rent-seeking”. This overview of bias will hopefully serve as a reference point and contribute to further rational discussion.

Friedman and Nissenbaum [1996] argue that technical bias places the demand on a designer to look beyond the features internal to a system and envision it in a context of use. Minimizing bias asks designers to envision not only a system’s intended situation of use, but to account for increasingly diverse social contexts of use. Designers should then reasonably anticipate probable contexts of use and design for these. If it is not possible to design for extended contexts of use, designers should attempt to articulate constraints on the appropriate contexts of a system’s use. We believe that our detailed model will help designers and policy makers to anticipate these probable contexts of use and formulate scenarios where bias can occur. The paper is structured as follows: In Section 2, we give background information to the problem. In Section 3, we give a summary of personalization and how it poses unique problems. In Section 4, we introduce a model of algorithmic and human filtering for online web services including personalization. In Section 5, we discuss implications for ethical analysis, social network analysis and design. Section 6 concludes this paper and lists several questions for future research.

## 2.2. INFORMATION OVERLOAD AND THE RISE OF THE FILTERS

According to Cisco, in 2015, the amount of consumer generated data on the Internet will be four times as large as it was in 2010 [62]. McKinkey’s research shows that “big data” is a growing torrent. In 2010, 30 billion pieces of content were shared every month with 5 billion mobile phones contributing to it [217]. An IBM study reports that every two days we create as much digital data as all the data (digital or non-digital) that was created before 2003 and 90% of the information in the world today has been created in the last two years alone [166]. In online (social) services, users actively contribute explicit data such as information about themselves, their friends, or about the items they purchased. These data go far beyond the click-and-search data that characterized the first decade of the web. Today, thanks to the advent of cloud computing, users can outsource their computing needs to third parties and online services can offer software as a service by storing and processing data cheaply. This shifts the online world to a model of collaboration and continuous data creation, creating so-called “big data”, data which cannot be processed and stored in traditional computing models [217].

Even though the amount of generated data on the social web has increased exponentially, our capabilities for absorbing of this information have not increased. Because the mind's information processing capacity is biologically limited (for example, we possess neither infinite nor photographic memory), we get the feeling of being overwhelmed by the number of choices and end up with “bounded rationality” [160]. Researchers across various disciplines have found that the performance (i.e., the quality of decisions or reasoning in general) of an individual correlates positively with the amount of information he or she receives, up to a certain point. If further information is provided beyond this point, the performance of the individual will rapidly decline [99].

One means of managing information overload is through accessing value-added information—information that has been collected, processed, filtered, and personalized for each individual user in some way [210]. Lu argues that people rely on social networks for a sense of belonging and interpersonal sources are recognized as more credible and reliable, more applicable, and can add value through intermediate processing and evaluation to reduce information overload. The general public prefers personal contacts for information acquisition [210]. As most of the data is produced and stored in the cloud, users delegate the filtering authority to cloud services. Cloud services are trying to extract value and insight from the vast amount of data available, and fine-tune it in order to show what is relevant to their users, often using the users' interpersonal contacts and social networks.

For instance, a search engine returns a list of resources depending on the submitted user query. When the same query was submitted by different users, traditional search engines used to return the same results regardless of who submitted the query. In general, each user has different information needs for their query. The user then had to browse through the results in order to find what is relevant for him. In order to decrease this “cognitive overstimulation” on the user side, many cloud services are exploring the use of personalized applications that tailor the information presented to individual users based upon their needs, desires, and recently on who they know in online social networks. Personalized systems address the overstimulation problem by building, managing, and representing information customized for individual users. Online services achieve this by building a user model that captures the beliefs and knowledge that the system has about the user [122]. In this way the system can predict what will be relevant for the user, filtering out the irrelevant information, increasing relevance and importance to an individual user.

Google uses various “signals” in order to personalize searches including location, previous search keywords and recently contacts in a user's social network [134]. As Figure 1 shows, different users receive different results based on the same keyword search. Facebook on the other hand registers the user's interactions with other users, the so-called “social gestures”. These gestures include like, share, subscribe and comment [330]. When the user interacts with the system by consuming a set of information, the system registers this user interaction history. Later, on the basis of this interaction history, certain information is filtered out. For instance content produced by certain friends might be hidden from the user, because the user did not interact with those friends over a period of time. Further, photos and videos receive a higher ranking than regular status posts and some posts receive a higher ranking than others [320]. Personalization algo-

gorithms thus control the incoming information (user does not see everything available), but also determine the outgoing information and who the user can reach (not everything shared by the user will be visible to others).

Personalization is a kind of information filtering. However, filtering is not a new concept. During our daily lives we filter information ourselves or delegate the filtering authority to experts, who are called gatekeepers [269]. This is because it would require an unreasonable effort and time for any individual to audit all the available information. The gatekeeper controls whether information passes through the channel and what its final outcome is, which in turn determines the way we define our lives and the world around us, affecting the social reality of every person. Traditional media is used to perform this “gatekeeping” role for news, determining what is newsworthy and important for its audience. However, as information technology and cloud computing are gaining importance, online web services that we use every day are slowly taking over the gatekeeping process that used to be performed by the traditional media.

According to van den Hoven and Rooksby [2008], information is a Rawlsian “primary good”, a good that everybody requires as a condition for well-being. Information objects are means to the acquisition of knowledge and in order to be an autonomous person to plan a rational life, we need information [256]. The more (relevant) data individuals can access in their planning, the more rational their life plan will be. Access to information is, then, a value because it may be instrumental in adding alternatives to one's choice set, or in ruling out alternatives as unavailable. As a requirement of justice, in high-technology information societies, people should be educated in the use of information technologies, and have affordable access to information media sufficient for them to be able to participate in their society's common life. Bagdikian [2004] similarly argues that media power is political power and the power to control the flow of information is a major factor in the control of society. Giving citizens a choice in ideas and information is as important as giving them choice in politics.

In 2005, the Pew Internet and American Life Project reported on the rise of search engines, and surveyed users' knowledge of how they worked. It concluded that “search engines are attaining the status of other institutions—legal, medical, educational, governmental, journalistic—whose performance the public judges by unusually high standards, because the public is unusually reliant on them for principled performance” [101]. Personalization and other forms of algorithmic filtering are thus “replacing the traditional repositories that individuals and organizations turn to for the information needed to solve problems and make decisions” [234]. The services that employ such algorithms are gateways that act as intermediaries between information sources and information seekers. They play a vital role in how people plan and live their lives. Since access to information is a value, and online filters allow or block access to information, building these algorithms is not only a technical matter, but a political one as well. Before discussing how bias can manifest itself in personalization, it is important to first understand how personalization works.

### 2.3. PERSONALIZATION – A TECHNICAL OVERVIEW

Most personalization systems are based on some type of user profile, a data instance of a user model that is applied to adaptive interactive systems. User profiles may include

**Ajax**

[www.ajax.nl/](http://www.ajax.nl/) - Translate this page 

logo\_aegon\_splash. logo\_adidas\_splash. naar de officiële site · english · spanish · chinese · greek.

In English - Aftellen richting Ajax - Olympique Lyon

**Ajax - Home**

[www.ajax.nl/Home.htm](http://www.ajax.nl/Home.htm) - Translate this page 

De officiële **Ajax**-website met nieuws en informatie over de spelers, wedstrijden, kaartverkoop en alle andere zaken die met **Ajax** hebben te maken.

Programma - Wedstrijden - Official Fanshop - Tickets & Travel

**english.ajax.nl - Home**

[english.ajax.nl/](http://english.ajax.nl/) 

**Ajax** goes Brazil during winter break. **Ajax** goes ... Next round still within reach for **Ajax**. Next round still ... Book now your matchpackage for a homematch of **Ajax**. ...

 Show stock quote for AJAX

About 388,000,000 results (0.10 seconds)

**Ajax (programming) - Wikipedia, the free encyclopedia**

[en.wikipedia.org/wiki/Ajax\\_\(programming\)](http://en.wikipedia.org/wiki/Ajax_(programming))

**Ajax** (also **AJAX**; pronounced */ˈeɪdʒæks/*; an acronym for **Asynchronous JavaScript and XML**) is a group of interrelated web development methods used on the ...

[Ajax framework](#) - [JSON](#) - [XHR](#) - [List of Ajax frameworks](#)

**AFC Ajax - Wikipedia, the free encyclopedia**

[en.wikipedia.org/wiki/AFC\\_Ajax](http://en.wikipedia.org/wiki/AFC_Ajax)

Amsterdamsche Football Club **Ajax** also referred to as **AFC Ajax**, **Ajax** Amsterdam or simply **Ajax** (after the legendary Greek hero), is a professional football club ...

**AJAX Tutorial**

[www.w3schools.com/ajax/default.asp](http://www.w3schools.com/ajax/default.asp)

Are you using **AJAX** to create interactive web applications? Learn how Altova's suite of application management and web data management tools can optimize ...

[AJAX Example](#) - [AJAX Introduction](#) - [AJAX Database](#) - [AJAX ASP/PHP](#)

About 385,000,000 results (0.12 seconds)

**Ajax (programming) - Wikipedia, the free encyclopedia**

[en.wikipedia.org/wiki/Ajax\\_\(programming\)](http://en.wikipedia.org/wiki/Ajax_(programming)) 

**Ajax** (also **AJAX**; pronounced */ˈeɪdʒæks/*; an acronym for **Asynchronous JavaScript and XML**) is a group of interrelated web development methods used on the ...

[Ajax framework](#) - [JSON](#) - [XHR](#) - [List of Ajax frameworks](#)

**AFC Ajax - Wikipedia, the free encyclopedia**

[en.wikipedia.org/wiki/AFC\\_Ajax](http://en.wikipedia.org/wiki/AFC_Ajax) 

Amsterdamsche Football Club **Ajax** also referred to as **AFC Ajax**, **Ajax** Amsterdam or simply **Ajax** (after the legendary Greek hero), is a professional football club ...

**Ajax (mythology) - Wikipedia, the free encyclopedia**

[en.wikipedia.org/wiki/Ajax\\_\(mythology\)](http://en.wikipedia.org/wiki/Ajax_(mythology)) 

**Ajax** or **Aias** (*/ˈeɪdʒæks/* or */ˈaɪ.əs/*; Greek: Αἴας, gen. Αἴαντος) was a mythological Greek hero, the son of Telamon and Periboea and king of Salamis. ...

Figure 2.1: Effects of personalization on Google. First screenshot is with a logged in user from the Netherlands. Second screenshot is from an anonymous user from the Netherlands. Last screenshot is from a logged in user from the US.

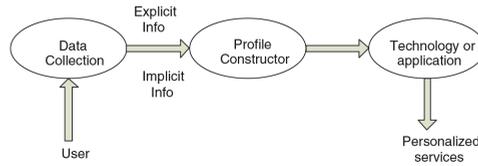


Figure 2.2: User profile construction for personalization (Adapted from [122])

demographic information, (e.g., name, age, country, education level), and may also represent the interests or preferences of either a group of users or a single person. In general, the goal of user profiling is to collect information about the subjects in which a user is interested, and the length of time over which they have exhibited this interest, in order to improve the quality of information access and infer the user's intentions. As shown in Figure 2.2, the user profiling process generally consists of three main phases. First, an information collection process is used to gather raw information about the user. Depending on the information collection process selected, different types of user data can be extracted. The second phase focuses on the construction of a user profile on basis of the user data. Here the collected and stored data are analyzed and processed. In the final phase, the compiled user profile is used in the actual web service, for instance a customized newsfeed in a social networking site, personalized results in a search engine query, or recommended products in an e-commerce site.

A system can build a user profile in two ways:

- **Explicitly:** the user customizes the information source himself. The user can register his interests or demographic information before the personalization starts. The user can also rate topics of interest.
- **Implicitly:** the system determines what the user is interested in through various factors, including web usage mining (i.e., previous interaction with the system such as clickthroughs, browsing history, previous queries, time spend reading information about a product), IP address, cookies, session id's, etc.

Explicit user information collection will allow the user to know that the personalization is taking place and he can tailor it to his needs. However, one problem with explicit feedback is that it places an additional burden on the user. Because of this, or because of privacy concerns, the user may not choose to participate. It is also known that users may not accurately report their own interests or demographic data, or, since the profile remains static whereas the user's interests may change over time [122]. Implicit user information collection, on the other hand, does not require any additional intervention by the user during the process of constructing profiles. It also automatically updates as the user interacts with the system. One drawback of implicit feedback techniques is that they can typically only capture positive feedback. When a user clicks on an item or views a page, it seems reasonable to assume that this indicates some user interest in the item. However, it is not clear that when a user fails to examine some data item it is an indication of disinterest [122].

Different techniques can be used to make suggestions to users on which information is relevant for them. Recommendation systems try to analyze how a user values certain products or services and then predict what the user will be interested in next. A recommendation mechanism typically does not use an explicit query but rather analyses the user context (e.g., what the user has recently purchased or read, and, if available, a user profile (e.g., the user likes mystery novels). Then the recommendation mechanism presents to the user one or more descriptions of objects (e.g., books, people, movies) that may be of interest [4, 120].

If this recommendation is done solely by analyzing the associations between the user's past choices and the descriptions of new objects, then it is called "content-based filtering". Due to increasing user collaboration and user-generated content, personalization can also be done socially. The so-called social information filtering [292] or collaborative filtering [120] automates the process of "word-of-mouth" recommendations: items are recommended to a user based upon values assigned by other people with similar taste. The system determines which users have similar taste via standard formulas for computing statistical correlations [292]. For instance, Facebook uses a collaborative filtering called Edgerank, which adds a weight to produced user stories (i.e. links, images, comments) and relationships between people [320]. Depending on interaction among people, the site determines whether or not the produced story is displayed in a particular user's newsfeed. This way, a produced story by a user will not be seen by everyone in that user's contact list. All stories produced by user X can be completely hidden in user Y's newsfeed, without the knowledge of both users.

According to Chatman [1987] and Lu [2007], people's information needs are highly diversified and individualized, making applicable and value-laden information most desirable, and yet the hardest to obtain. Interpersonal sources can, to a great extent, minimize these difficulties and maximize the utility of information. Even though personalization technologies such as Grouplens [275] have existed for a while, the rise of social networks and the exponential increase in produced and shared information in online services are changing the impact this technology has. According to Garcia-Molina et al. [2011], information providing mechanisms (e.g. search engines) and personalization systems have developed separately from each other. Personalization systems like recommendation engines were restricted to a single homogenous domain that allowed no keyword search. Search engines on the other hand were geared toward satisfying keyword search with little or no emphasis on personalization or identification of intent. These two systems were separated partly due to a lack of infrastructure. Today, due to a combination of a powerful and cheap back-end infrastructure such as cloud computing and better algorithms, search engines return results extremely fast, and there is now the potential for a further improvement in the relevancy of search results. So, we now see a trend where personalization and information providing mechanisms are blending

## 2.4. A MODEL OF FILTERING FOR ONLINE WEB SERVICES

Existing work on gatekeeping theory often points out the changing role of the journalist from a gatekeeper to a gatewatcher [48, 294]. With the increasing popularity of the online media and social networks, every user can share information depending on what he thinks is important. Scholars thus argue that by using online services, the audience can

exert a greater control over news selection and can focus on issues that they consider more relevant, which in turn empowers audiences and erodes the degree of editorial influence over the public's issue agenda [11]. Some even argue that the gatekeeping role performed by the traditional media becomes irrelevant; gates are disappearing [201]. Information may diffuse through social networks next to mass media channels; therefore any audience member can be a gatekeeper for others. Journalists now become a "gatewatcher", providing a critical analysis of existing topics that are chosen by the community [48].

Some also claim that the platforms the new "gatewatchers" operate are neutral. According to Bruns [2011], tools such as Twitter are neutral spaces for collaborative news coverage and curation operated by third parties outside the journalism industry. As a result, the information curated through collaborative action on such social media platforms should be expected to be drawn from a diverse, multiperspectival range of sources. Also Lasorsa et al. [2012] claim that platforms such as Twitter are neutral communication spaces, and offer a unique environment in which journalists are free to communicate virtually anything to anyone, beyond many of the natural constraints posed by organizational norms that are existing in traditional media.

However, as we shall show, the gatekeeping process in online information services is more than a simple transition from editor selection to audience selection or from biased human decisions to neutral computerized selections. We argue that human factors play a role not only in the development of algorithms, but in their use as well. We show that factors that caused bias in mass media news selection still play a role in information selection in online web services. Online information intermediaries, similar to the traditional media, can control the diffusion of information for millions of people, a fact that gives them extraordinary political and social power. They do not provide equal channels for every user and they are prone to biases. Just as any computer system, they can unfairly discriminate against certain individuals or groups of individuals in favor of others [113].

#### 2.4.1. SOURCE SELECTION ALGORITHM

At the stage of "Collection and Selection" (Figure 2.3), the online service starts to collect its information from various sources. For instance a search engine will automatically crawl the web, while the social network site will collect information produced by its users. However, similar to the traditional media where gatekeeping starts with journalists [58, 295], algorithmic gatekeeping already starts at source selection. First of all, not all information is digital, thus all non-digital information will be absent from online information intermediaries. Further, not all digitally available information will be available to each service, for instance search engines do not index all the data available on the Internet, leading to coverage bias [129, 348]. Google admits that the company does not index every one of the trillion pages on the web, because they are similar to each other or because Google considers some of them not useful to the searcher [131]. Technical reasons can also prevent a search engine to crawl a site. The design of the website might make the source collection and indexing process difficult or the site itself might be explicitly blocking the crawling process [31]. Further, if a resource has a bad reputation, for instance if it is suspected as an illegal site, it might be left out of the whole collection

process. It is also possible that the source does not want to be included in the index due to various reasons. For instance not every page in Facebook or Twitter is indexable by Google [308].

## 2

#### 2.4.2. INFORMATION SELECTION AND PRIORITIZATION ALGORITHM

In traditional media, newspaper editors select some of the messages produced by journalist to make news [32]. Algorithms used in web services (such as ranking algorithm in a search engine, or news feed algorithm in a social network) make similar decisions. The design of these algorithms is affected by choices made by designers, i.e., which factors to include in the algorithm, and how to weigh them <sup>1</sup>. To serve majority interests, information intermediaries often include popularity metric in their ranking algorithm. A search algorithm for instance can give more weight to information coming from popular websites, to support majority interests and values. As a result, seekers will have trouble finding the less popular and smaller sites [245].

Because the information filtering is automated, it might be manipulated by activities from third parties. This happens with the so-called “black-hat” Search Engine Optimization (SEO) techniques. This is a method of raising the profile of a Web site with methods that Google considers tantamount to cheating [290]. Another factor is own product/service prioritization. The EU recently received a complaint from a shopping search site that claimed it and other similar sites saw their traffic drop after Google began promoting its own services above conventional search results [8, 93, 94, 363]. Google also integrates content from its social networking platform Google Plus into Google search results, causing protest by the social networking platform Twitter [289]. Studies also showed that Google and Bing search engines both reference their own content in its first results position when no other engine does [93, 364]. Facebook is criticized for favoring the products of its partners [109]. The algorithm can also prioritize certain types of information over others. For instance, it is claimed that Facebook treats video and pictures as more important than links and status updates [318]. Similarly, comments on an item are four times more valuable than “likes” [361].

In traditional media, regardless of the size of an event such as a public protest, the likelihood that the event will be reported in the media will depend on the current agenda. This is because both print and electronic media regularly focus upon selected issues over a sequence of days, creating the phenomena of “issue attention cycles” [298]. We can observe similar behavior in social media. Twitter has a feature called trending topic (TT), in which most popular topics Twitter users are talking about in a particular location are highlighted. However Twitter does not solely check popularity of an item while determining TT’s, it favors novelty over popularity. Twitter checks if the user updates on a specific topic is increasing quickly enough. Even if a topic is large volume wise, if the increase rate is small or if it is not novel, it won’t make it to the “trending topics” [326]. This means that it is much easier for a term never seen before to become a Twitter trend and the longer a term stays in the trending topic list, the higher velocity required to keep

<sup>1</sup>For instance, Facebook uses an algorithm called Edgerank to determine how a newsfeed of a user is constructed. It is believed that several factors are used to select/prioritize user updates, such as affinity between the receiver and sender, and the date of the published update. However, the exact formula is unknown. See [320]

it there [209]. This novelty factor caused the hashtag “IcantRespectYourlf” to be a TT in the US, while #OccupyWallStreet not making it to the list. This is because when #OccupyWallStreet was a TT throughout the world, it had previously trended in the U.S., and now there were no more new people in the U.S. talking about it.

According to Gillespie [2012], this choice fosters a public more attuned to the “new” than to the discussion of persistent problems, to viral memes more than to slow-building political movements. The exact algorithm that determines the trending topics is unknown and this opacity makes the TT, and their criteria, deeply and fundamentally open to interpretation and suspicion [125].

TT differs in important ways from those employed in personalization, as it presents itself as a measure of popularity. However, since algorithms such as TT can differ per country, region or city, they might be used to customize content, as an important signal. Popularity can thus be an input to customize items for a group of users. This is still tailored content, but not for an individual, but for a group of individuals.

Finally, the age of an information source or the age of the information item can also matter. In Google search engine, the number of years a domain name is registered has an impact on search ranking; domain names that exist for a period of time are preferred over newly registered ones [169]. In Facebook, the longer a status update has been out there, the less weight it carries. A news item is prioritized over an old item [320]. This might for instance lead companies to post updates when their audience is most likely to be online and using Facebook.

### 2.4.3. HUMAN OPERATOR

In traditional media, individual factors such as personal judgment can play a role during the selection of news items for a newspaper. An editor's decisions can be highly subjective and can be based on the gatekeeper's own set of experiences, attitudes and expectations, leading to a selection bias [119]. Online web services such as search engines frequently claim that such human bias do not exist in their systems. They claim that their core operations are completely automated, but this is false. Humans in online services also make editorial judgments about what data to collect delete or disregard. According to Goldman, online services manually inspect their index and make adjustments [129]. For instance search engines make manual adjustments of a web publisher's overall rating or modify search results presented in response to particular keyword searches [129]. The Dutch newspaper Trouw's entire domain name and all hosted pages were removed from Google index because of a violation of the company policy [78, 144]. Google itself has admitted that the company manually demotes websites [223]. Similar to blacklisting, search engines can also perform whitelisting. For instance Google recently mentioned that it uses whitelists to manually override its search algorithms [222].

Information deletion or withholding is not specific to search engines. Facebook a photo of two men kissing from a user's Wall due to a violation of the site's terms of service [374]. There are also claims that Facebook denies and removes advertisements designed for gay audience with no nudity or sexual content, labeling it “inappropriate” [2]. Others claimed that Facebook labeled their posts containing links to a political activism site as spam and prevented the users disseminating this information [22]. Facebook has also removed pages because of offensive content, but later reinstated them [167, 179]. Face-

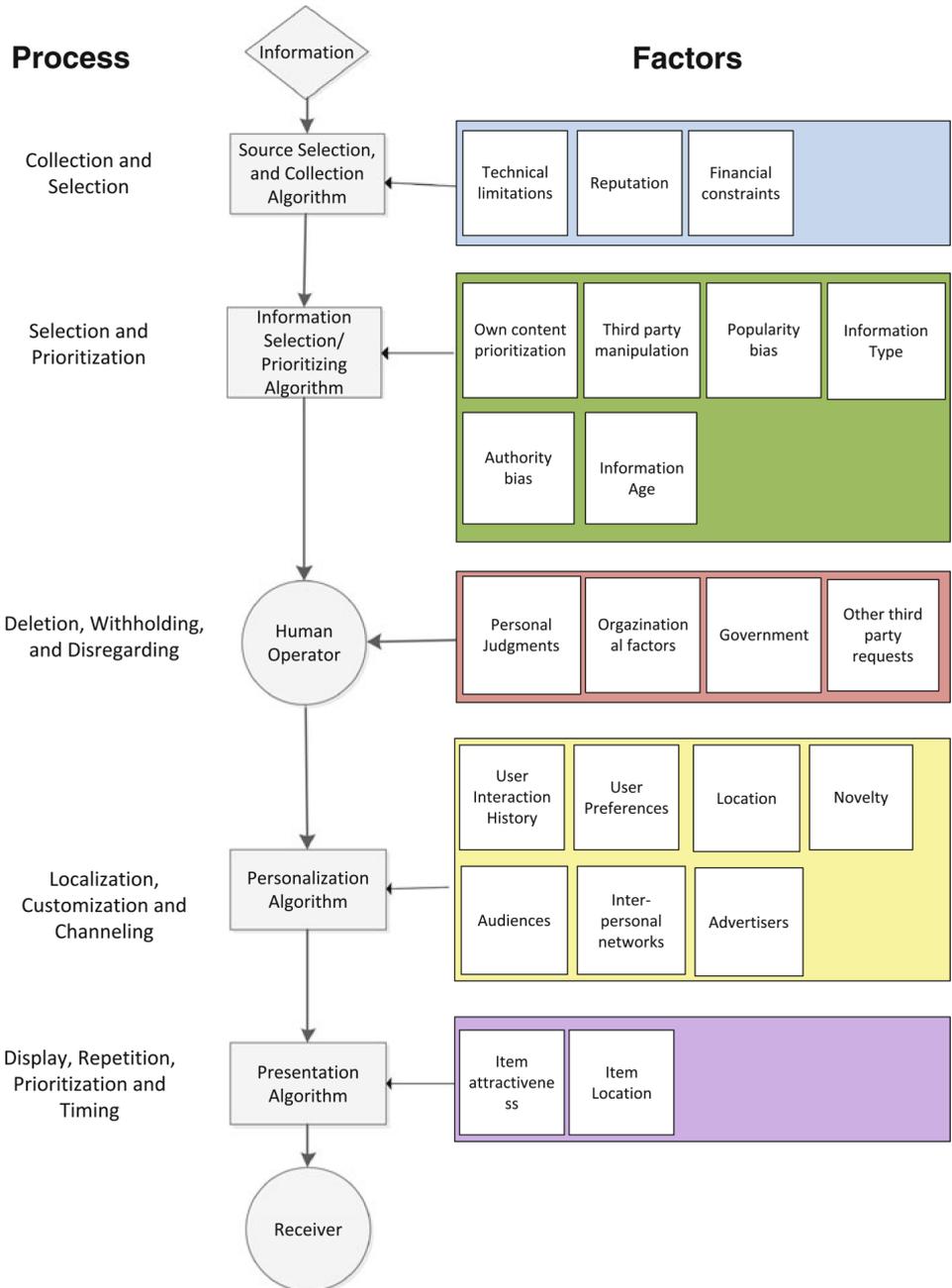


Figure 2.3: A model of filtering for online web services including personalization

book spokesman blamed the human reviewer in some of the cases, but did not reveal the criteria the company uses on what makes content offensive or in violation with the company's terms of use. Twitter similarly removes certain 'trending topics' if it considers it as "offensive" [70].

Scholars in media studies argued that organizational factors in traditional media play a more important role than individual judgments. In the uncertainty of what tomorrow's news will be, journalists use so-called routines, patterned, repeated practices and forms, to view and judge in order to define news as predictable events [107]. Similarly, online web services employ operators to delete, withhold or disregard information, to enforce company guidelines. Even though these operators have to obey a set of rules to apply, they have, just like journalists, their own values and can pass personal judgments. This might give the image that the operator is bound to strict rules, and acts merely as an enforcer. However people do not always execute rules in the same way and individual-level characteristics are still important [294].

Human operators of online services have to evaluate removal requests coming from governments. For instance, recently, A Delhi Court ordered 22 social networking sites (including Facebook, Google, Yahoo and Microsoft) to remove all "anti-religious" or "anti-social content and file compliance reports. Google has a list of content removal requests from governments all around the world [132]. Operators also have to deal with requests coming from third parties. For example, Google regularly removes content due to copyright claims coming under the Digital Millennium Copyright Act, Section 512(c). This act gives providers immunity from liability for their users' copyright infringement, if they remove material when a complaint is received [59].

#### 2.4.4. PERSONALIZATION ALGORITHM

According to Goldman [2005], personalized ranking algorithms reduce the effects of technical bias introduced by algorithms in online intermediaries. Goldman argues that personalization algorithms increase relevancy and produce a different output per individual user. This in turn diminishes the weight given to popularity-based metrics and reduces the structural biases due to popularity. Personalization might increase relevance, however as we show in this subsection, designing only for this value will introduce problems.

#### USER INTERACTION HISTORY AND USER PREFERENCES

As we have argued in Section 3, users could personalize the information they receive by giving their preferences explicitly. In this way they can receive personalized information on the criteria they know. However, if the user's interests change over the time and if the user does not update their filter, they might miss some information that might be of interest to her. Lavie et al. [2009] found that people might be interested in things that they did not know they were interested in, due to the formulation of the topic. Some users have asserted that they were not interested in politics, but later it was shown that their perception of "politics" was limited to local politics. They later have shown interest in international politics [198]. Lavie et al. argue that, overall, users cannot accurately assess their interests in news topics. Similarly Tewksbury [2003] reports that user's declared and actual interests may differ.

In his book Republic.com, Sunstein [2002] developed his concern that explicit per-

sonalization will assist us to avoid facts and opinions with which we disagree, leading people to join online groups that conform with their existing beliefs. Since democracy is most effective when citizens have accurate beliefs and to form such beliefs, individuals must encounter information that will sometimes contradict their preexisting views. Sunstein argues that explicit personalization will undermine deliberative democracy by limiting contradictory information.

Implicit personalization using user interaction history has its own concerns. Pariser [2011] argues that online services can cause citizens to be ill-informed about current events and may have increasingly idiosyncratic perceptions about the importance of current events and political issues. This might occur because online services are trying to improve accuracy at the expense of serendipity, leading to what Pariser calls “filter bubble”. Even if users wanted to diversify their network explicitly, information intermediaries silently filter out what they assume the user does not want to see, hiding information posted by opposite end of political spectrum. For Sunstein, explicit excessive personalization leads to never seeing the other side of an argument and thus fostering an ill-informed political discourse. For Pariser, excessive implicit personalization leads to an unhealthy distaste for the unfamiliar. The problem is thus an automatic cyberbalkanization, not an “opt-in” one. It happens behind the scenes and we do not know what we are not seeing. We may miss the views and voices that challenge our own thinking

Pariser argues that online personalization algorithms are designed to amplify confirmation bias, Consuming information that conforms to our beliefs is easy and pleasurable; consuming information that challenges us to think differently or question our assumptions is difficult. Pariser notes that we all have internal battles between our aspirational selves (who want greater diversity) and our current selves (who often want something easy to consume). Pariser argues that the filter bubbles edit out our aspirational selves when we need a mix of both. Pariser believes that the algorithmic gatekeepers need to show us things that are not only easy to consume but also things that are challenging, important and uncomfortable and present competing points of view. Pariser states that filter bubbles disconnect us from our “ideal selves”, that version of ourselves that we want to be in the long-run, but that we struggle to act on quickly when making impulse decisions.

#### LOCATION

As we have shown in Section 3, content can also be personalized based on location. Large Web-search engines have been “personalizing” search to some extent for years. Users in the U.K. will get different results searching for certain terms, especially commercial ones, than users in the U.S. Results can change between different cities as well [120]. The idea is that the user will be more interested in local content. However, this will depend on context of information. For instance, if I am looking for a restaurant, I would want my search engine to personalize results based on location, the system should show me pizzerias in Rotterdam, but not in New York. However, if I am looking for some technical information in a forum to solve a PC problem, then I do not necessarily care about the location (if I can speak multiple languages). Currently, most personalization systems filter information based on location without taking the context into the account. This might always favor local content, even if the quality or the relevance of the local content is inferior to a non-local content.

## AUDIENCES

While traditional news media outlets want to satisfy their readers and viewers, it is much more difficult for them to modify their selection criteria in real time, than it is for online gatekeepers. Online gatekeepers have immediate feedback about what queries are issued, what content is selected and what sites are accessed. For instance online services can observe user behavior through entered queries or clicked links to modify its algorithms accordingly. However, online services can also capture user's intent by using social gestures. Examples of these social gestures include the "like" and "subscribe" buttons in Facebook and the "+1" button in Google search. By clicking on these buttons users express their interests and see what item is popular. Google currently does not use these (anonymous) votes to personalize search results, but such approaches are well known in computer science literature. Search behavior of communities of like-minded users can be harnessed and shared to adapt the results of a conventional search engine according to the needs and preferences of a particular community [300]. Because similarities will exist among community members' search patterns and web search is a repetitive and regular activity, a collaborative search engine can be devised. This human PageRank or "social-graph", using +1 results to give context to the popularity of a page, can be a supplement (or alternative) to the link graph Google is currently using.

Some claim that the community is wiser than the individual. However, community driven filtering has its own problems. For example, in social news aggregator Reddit, where anonymous users submit links to items, comment on them, vote on the submitted items and comments, the community determines what is newsworthy, for every topic. Users can personalize their news feed by explicitly subscribing to certain subtopics, but the popularity metric is used in every subtopic. In Reddit, the timing of the story submission is important. If a good news item is submitted outside of Internet prime-times, it will not receive enough votes to make it to the front page. The result is that most submissions that originate in the US end up being dominated by US comments, since new comments posted several hours after the first will go straight to the middle of the pile, which most viewers will never get to. Submission time has a big impact on the ranking and the algorithm will rank newer stories higher than older. In Reddit, first votes also score higher than the rest. The first 10 upvotes count as high as the next 100, e.g. a story that has 10 upvotes and a story that has 50 upvotes will have a similar ranking. Controversial stories that get similar amounts of upvotes and downvotes will get a low ranking compared to stories that mainly get upvotes [281]. Further, the user will receive positive or negative points on the story he submitted. The individual might remove the story due to decreasing points in his reputation.

It is also known that in such vote-based social news sites, the amount of contacts or followers one has can also determine whether his story will make it to the front page. Having a large number of contacts will make it easier to reach the front page (more friends, more votes). Also, some social news aggregators divide the stories into topics. If a topic has a small number of subscribers, the chance that it will make it to front page is small [181]. Even the items that do not make it to the front page will bring traffic to the submitted site. Therefore social news aggregators like Reddit are being used and manipulated by online marketing professionals, in order to draw more traffic to their products or services. Similarly, Facebook's like button can also be gamed. Digital marketing com-

panies can create fake users and buy “friends” and “likes” [327]. These companies use software to automate clicking the “Like” button for a certain page. Such software can bypass Facebook’s security system. If popularity is devised by only the number of likes and used as an input for users in a certain region, it can also cause bias in personalization.

## 2

### INTERPERSONAL NETWORKS

According to Chen and Herson [1982], the general population tends to obtain information through interpersonal networks, rather than formal means [56]. Durrance [1984] found that more than 64% of her research participants used interpersonal sources [92]. Sturges maintains that there is a “fundamental preference for information mediated by human interaction” and that “there is evidence of this from all parts of the world and from most important aspects of human life” [307]. Katz and Lazarsfeld [2005] argue that we live in communities and we are inherently tied to different social connections. We interact in formal or informal social groupings, in so-called “primary groups” such as families, friends, work teams, clubs or organizations. These primary groups delineate major life boundaries for each one of us in society, our routine activities mainly occur in these primary groups.

Since our lives are mainly contained in primary groups, our attitudes and opinions tend to derive from them as well as our sources of information. Primary groups provide us with “social reality” to validate our actions. As we encounter unknown situations and difficult decisions, we turn to and consult our social contacts, including both strong (e.g., family and friends) and weak ties (e.g., colleagues, acquaintances) to help us form opinions and find solutions [136]. Lu [2007] argues that, through interactions concerning a particular issue, a primary group tends to develop a common view and collective approach, hence, provides a social reality that helps and validates decision making by its members. Because members of a primary group share the community language and background information, their communication is made effortless. Information so transmitted becomes easily accessible and digestible [210].

Because of these reasons, instead of relying on user’s explicit preferences, or using an anonymous popularity metric, personalization services started to use interpersonal relationships to filter information. For instance Facebook launched a program called “instant personalization” with an exclusive set of partners, including the restaurant aggregator site Yelp, Microsoft online document management site docs.com, customizable Internet radio sites Pandora and Spotify. These partners have been given access to public information on Facebook (e.g., names, friend lists, and interests and other information users have shared on their Facebook profiles) to personalize a user’s experience on the partner’s site. As an example, online music service Spotify requires a Facebook account, and using the friends list in Facebook, it shows the user what her friends have listened to. The idea here is, since these contacts are part of our primary group, we can trust their judgment on which information is newsworthy. If our primary groups are available in every web service we use, then our experience using that web service can be customized.

Similarly Google introduced social search in 2009, personalizing search results based on people you know in Facebook and Twitter, rather than your personal behavior. As a latest move, in 2012, Google introduced a feature called “Search plus your world”. This feature personalizes the results using user connections in Google Plus, Google’s social

networking platform. This means you might see a picture of a friend's car when you search for a new automobile, or a restaurant recommended by a friend when you search for a place to eat. Even if you aren't a Google+ user, Google search results will show content posted publicly on the social network that it judges to be relevant—profile pages and pages dedicated to particular topics [182].

### ADVERTISERS

Traditional mass media is primarily supported by commercial sponsorship. This can cause the newspapers to delete, change or prioritize news items due to advertising pressure [301]. Same pressure applies to online services; the majority of online service revenues come from advertising [247, 287, 331]. Personalization is a very attractive tool for advertisers, as user data collected for information filtering can be used for behavioral targeting. This sort of online targeting provides more relevant online advertising to potential upcoming purchases. Using the built up user profile in online services, advertising networks can closely match advertising to potential customers. According to Guha et al. [2010], Facebook uses various profile elements to display targeted advertisement including age, gender, marital status, and education. A Facebook advertiser can target users who live within 50 miles of San Francisco, are male, between 24-30 years old, single, interested in women, like skiing, have graduated from Harvard and work at Apple [184]. Google allows advertisers to target ads based not just on keywords and demographics, but on user interests as well [252]. Companies have recognized that providing advertisements along with their recommendations (suitably distinguished from the recommendation results) can be extremely profitable. For instance, the auction site Ebay provides a “deal of the day” for all visitors to the site, in addition to “buy it now”, special items directly sold from a provider for a fixed price—both of these are essentially advertisements [120].

### PRESENTATION ALGORITHM

Once information is chosen through the information selection algorithm and personalized for the user, it does not mean that it will be seen and consumed. The placement of the information might determine if it makes it out of the filter. Joachims and Radlinski [2007] show that the way a search engine presents results to the user has a strong influence on how users act. In their study, for all results below the third rank, users did not even look at the result for more than half of the queries. Bar-Ilan et al. [2009] report similar findings. Yue et al. [371] report that the attractiveness of information can also cause presentation bias if the title and abstract of a resource is bolded, it generates more clicks. They also show that people tend to click on the top and bottom results. These findings show that what the user will consume can be affected by the algorithm, even after source selection and personalization.

## 2.5. DISCUSSION

### 2.5.1. IMPLICATIONS FOR AN ETHICAL ANALYSIS

Personalization is the latest step in this algorithmic filtering process. As we have argued, even though personalization algorithms have existed since the 1990's, information providing services such as search engines did not contain such algorithms until recently.

This is mainly due to the recent availability of cheap and powerful backend infrastructure and the increasing popularity of social networking sites. Today information seeking services can use interpersonal contacts of users in order to tailor information and to increase relevancy. This not only introduces bias as our model shows, but it also has serious implications for other human values, including user autonomy, transparency, objectivity, serendipity, privacy and trust. These values introduce ethical questions. Do private companies that are offering information services have a social responsibility, and should they be regulated? Should they aim to promote values that the traditional media was adhering to, such as transparency, accountability and answerability? How can a value such as transparency be promoted in an algorithm? How should we balance between autonomy and serendipity and between explicit and implicit personalization? How should we define serendipity? Should relevancy be defined as what is popular in a given location or by what our primary groups find interesting? Can algorithms truly replace human filterers?

A relevant value to bias is information diversity. For instance if a search engine is exercising bias toward an advertiser, it will be limiting the diversity and democracy inherent to the information [135]. Information diversity is a rich and complex value that can be conceptualized in many different ways, and its interpretation differs significantly per discipline. In media studies, it might be translated as “minority voices having equal access in the media ” or “the degree which the media relates to the society in such a way to reflect the distribution of opinion as it appears in the population” [335]. In Computer Science literature, it can be defined as “variety in the products offered by the system”, “helping user find items he cannot easily find himself” [373] or “identifying a list of items that are dissimilar with each other, but nonetheless relevant to the user’s interests” [370]. While media studies are analysing this ethical value in detail, almost all scholars of search engine diversity seem to be limiting their understanding of “bias” and “diversity” to popularity bias [135]. As our model shows, popularity is only one of the many factors that cause bias. We need a normative conceptualization of the value information diversity that borrows notions from media studies, such as media ownership, content diversity, viewpoint diversity, reflection and open-access [335]. Only then can we translate this complex value into design requirements of information intermediaries and move towards a solution.

We believe that normative arguments based on our model will be stronger, more concrete and constructive. As an example, take the value user autonomy. Autonomy is centrally concerned with self-determination, making one’s own decisions, even if those decisions are sometimes wrong [114]. Autonomy is thus the individual’s ability to govern herself, be one’s own person, to be directed by considerations, desires, conditions, and characteristics that are not simply imposed externally upon one, but are part of what can somehow be considered one’s authentic self [61]. It is this aspect of decision-making that allows us to be responsible for the consequences of our actions. While designing technology, one can thus assume that designers should maximize user autonomy by following the simple dictum that more control leads to more user autonomy. After all, if autonomous individuals need to have freedom to choose ends and means, then it could be said that wherever possible and at all levels, designers should provide users the greatest possible control over computing power. Considering this notion of autonomy, one could

argue that personalization algorithms should always be fully customized and should be based on explicit personalization. However, as the model shows, explicit personalization based on user preferences is also prone to bias. People might be interested in things that they did not know they were interested in, due to the formulation of the topic. Further, users might not accurately assess their interests in certain information items. As we have mentioned, user's declared and actual interests may differ.

This seems to suggest that autonomy in this context should not be understood as "full user control". User autonomy seems to have less to do with simply the degree of control and more to do with what aspects of the algorithm are controllable, and the user's conception and knowledge of the algorithm. As Friedman and Nissenbaum note, achieving higher order desires and goals will enhance autonomy, whereas excessive control may actually interfere with user autonomy by obstructing a user's ability to achieve desired goals [114]. This means that, implicit personalization must be combined with explicit personalization to decrease excessive control. For instance a personalized search engine might be implemented in such a way that, the system enters a dialogue with the user, explicitly stating that a certain query is personalized, explaining why and due to which reasons it is personalized. The system can thus make assumptions to predict what the user might like, but it should refine itself by asking simple questions to the user to confirm if those assumptions were correct. While the user might not control the full algorithm, the system might receive feedbacks and show the user under which conditions it is making certain recommendations.

As we have argued, information should be accepted as a primary good, a vital good for people to plan their lives rationally and to participate adequately in the common life of their societies [344]. Thus, having access to information affects the value of liberty perceived by an individual. We therefore argue that personalizing algorithms affect the moral value of information as they facilitate an individual's access to information. Contrary to earlier stages of the Internet-era, when the problem of information access boiled down to having access to hardware, today the problem of access to information concerns the ability to intentionally find the right information, or the likeliness of unintentionally stumbling upon the relevant information.

Some argue that users should sabotage the personalization system by deliberately clicking on links that make it hard for the personalization engines, erasing cookies, unlocking everyone on a social network, posting something and then ask the Facebook friends to click the "Like" button and comment, or simply switch to a service that does not use personalization [96, 256]. However, these tactics are tedious, not always possible to perform and their effect depends on the implementation of the current system. Further, personalization might actually have a positive effect on the ecology of the cyberspace: the incentives to game the system and invest in practices like "search engine optimization" can become weaker [129, 231]. We should come with design suggestions to minimize the bad effects and improve the good effects of this technology instead of trying to get rid of it all together.

The question is then not whether to have personalization or not, but how to design morally good personalization technology. 'Having too much information with no real way of separating the wheat from the chaff' is what Benkler [2006] calls the Babel objection: 'individuals must have access to some mechanism that sifts through the universe of

information, knowledge, and cultural moves in order to whittle them down into manageable and usable scope'. The question then arises whether the service providers currently active on the Internet are able to fulfill the 'human need for filtration'. Although the fulfillment does not hinge on proprietary services alone as there are cooperative peer-production alternatives that operate as filters as well, the filtering market is dominated by commercial services such as Google and Facebook [163]. Having an option to turn it on or off is not really a choice for the users, as they will be too dependent on it in the existence of information overload.

### 2.5.2. IMPLICATIONS FOR DESIGN

In order to anticipate different contexts of use in personalization, a value based study such as Value Sensitive Design [108, 116] seems to be the right direction. Value Sensitive Design (VSD) consists of an empirical investigation accompanied by a philosophical analysis and a technical study. Friedman and Nissenbaum [1996] argue that designers should not only envision a system's intended situation of use, but to account for increasingly diverse social contexts of use. Designers should then reasonably anticipate probable contexts of use and design for these. If it is not possible to design for extended contexts of use, designers should attempt to articulate constraints on the appropriate contexts of a system's use. Bias can manifest itself when the system is used by a population with different values than those assumed in the design. This is especially true for the design of most online information intermediaries, where users from the whole world will be served instead of only local ones.

Another issue that is relevant to the design of personalization algorithms and other filtering mechanisms is exposure diversity. Even if an information intermediary provides a balanced information diet, this does not guarantee that the user will actually consume this information [157, 238, 243]. Content diversity is not equal to exposure diversity. We need to devise methods to increase the consumption of challenging content by users. Munson and Resnick [2010] distinguished two types of users: challenge averse (those who ignore diverse content) and diversity seeking. They tried to show more diverse content to those who were challenge averse, for instance by highlighting agreeable items or showing agreeable items first. However, this did not increase users' consumption habits, they still ignored challenging items. This requires us to research further how challenging items can be made attractive to users so that they actually consume the incoming information.

### 2.5.3. IMPLICATIONS FOR THE DESIGN OF SOCIAL FILTERING

Media scholars often argue our interpersonal contacts have become our gatekeepers [294]. However, if this approach becomes ubiquitous in design, it can lead to problems. First, this obviously raises concerns for privacy. An item a user has consumed can be shared with others without their notice. The Electronic Privacy Information Center, American Civil Liberties Union and American Library Association claim the changes have made sharing information on Facebook a passive rather than active activity. In this way, users might reveal more than they intend [241]. Even if sharing process was more active, it can still cause issues. For instance, an item a user has shared in a social network in certain context and has forgotten can reappear in a Google search result in a different

context. Further, an implicit user profile built for personalization leads to epistemological problems. Does the knowledge about the user (gathered by user's interaction with the system) represent the reality? Does the user interact with its primary group the same way he interacts in the offline world? How much does a user have a say in this built profile and to what degree can he control the dissemination of this representation of himself?

Second, not everyone in our online social networks will be part of our primary group; not every online "friend" is our real friend and we might share different things with our online friends. We sometimes add people to our network because of courtesy, as it otherwise might cause relationship problems in the offline world ("Why did you not answer my friend request?"). To remedy this, we can arrange the level of our relationship with others in a social network; we can divide them into lists or groups. We can then choose what we want to share with which group. However, our contact list in a social network can be connected with a different service, for personalization. When we use our social network in another service, lists we have created can suddenly disappear. For instance, Spotify uses Facebook contact list to provide recommendations per individual user. However, it ignores all the lists that have been created and shows what all friends have listened to regardless of the relationship between the user and the friend. The categorization the user has set in the Facebook platform in order to define and control his relationships are gone when the Facebook data is used elsewhere. Next to increasing information overload, this can also cause privacy issues. Even if I choose to share things with some people in Facebook context, everything I listen to in Spotify will be shown to all my Facebook users. This context loss will be more common as more services integrate with each other.

Third, not everyone has competence on every subject. Scholars in various disciplines have found that there are strategic points for the transmission of information in every group [5, 55, 210]. Even though it is possible that people can interact randomly with anyone who has available information, information transmission is never a simple aggregation [177, 296]. Some individuals, who are more information-savvy, will automatically occupy strategic positions to facilitate access to information to others. Depending on the subject matter, not everyone in a group is equally important or qualified in providing information. Those who have more knowledge will act as gatekeepers. I might trust John's competence in football, and use him as my gatekeeper in this subject, but not in the area of international politics. However, in most online services, we get to see everything published by a user, or nothing at all. We need mechanisms to assess the competency of the information sharer and determine the needed gatekeeper for a given context.

Fourth, online services are trying to capture user's intent by using social gestures. Examples of these social gestures include the "like" and "subscribe" buttons in Facebook and the "+1" button in Google search. By clicking on these buttons users express their interest and communicate to their peers. However, this sort of expression seems somehow limiting [256]. The reason of the expression and the emotion behind the expression is not captured by the button. There is a difference between liking a film, liking a director, liking a genre or liking films of a certain period. I might like a film for various reasons: to recommend to friends, to express my identity, to receive further film recommendations or to add it into my collection for later use. Such buttons are simplifying complex

human actions and emotions into a single dimension. As Friedman and Nissenbaum [1996] have argued, attempting to formalize human constructs such as discourse, judgments, or intuitions and trying to quantify the qualitative, discretizing the continuous will lead to biases.

Fifth, online services assume that users want to have an online experience where consuming any sort of information is done socially and collaboratively. This is why Google is making social search the default type of search and Facebook persuades users to share more information or leave a trace of a completed activity, by its “frictionless sharing”. These approaches aim to make sharing an effortless activity, in which everything is shared and hopefully some things will be found interesting by the users. However by promoting ease, they are undermining not only privacy, but also autonomy. In a frictionless sharing environment, user now cannot actively reflect on things he consumes and choose on what to share.

Finally, if we know the information we consume is being shared and read by our primary groups, we might change our behavior on what to share, and even choose what to consume if this is shared automatically. According to Sunstein [2008], group members may fail to disclose what they know out of respect for the information publicly announced by others. That is, even if we have big doubts about claims made by the majority of a group, we might think they are not errors at all; not so many people can be wrong. Individuals can also silence themselves to avoid the disapproval of peers and supervisors. As a result of these two forces, information cascades might occur; individual errors might amplify instead of being corrected, leading to widespread mistakes. Information held by all or most will be prioritized over held by a few or one.

#### 2.5.4. IMPLICATIONS FOR SOCIAL NETWORK ANALYSIS

While bias might manifest itself in the social platform, users themselves might be biased in information sharing. Therefore we need to determine whether bias occurs naturally in social networks, as personalization algorithms use more and more social data. Do users tend to follow like-minded users? Do they do this intentionally? Do they only share things that they agree with? Do they receive diverse information directly or indirectly? Do they only want to follow popular items coming from major news sources as the current services, or does the minority receive a chance to contribute to the debate? Is the sharing behaviour of the user changing with what he is receiving? Does culture have an affect in diverse information seeking behaviour? To answer such questions, we need to perform more empirical studies.

Facebook performed one of the few studies that actually studies bias in social networks [26]. The empirical study suggests that online social networks may actually increase the spread of novel information and diverse viewpoints. According to Bakshy [2012], even though people are more likely to consume and share information that comes from close contacts that they interact with frequently (like discussing a photo from last night's party), the vast majority of information comes from contacts that they interact with infrequently. These so-called “weak-ties” [136] are also more likely to share novel information.

Even though this is one of the first empirical studies that aims to measure information diffusion, there are some concerns with it: First of all, the study is not repeatable

and the results are not reproducible. Facebook scientists simply manipulated newsfeed of 253 million users, which only Facebook can perform. Second, our weak ties give us access to new stories that we wouldn't otherwise have seen, but these stories might not be different ideologically from our own general worldview. They might be new information, but not particularly diverse. The research does not indicate whether we encounter and engage with news that opposes our own beliefs through links sent by "weak links". It could very well be that we comment on and re-share links to cat videos sent by our previous neighbour, or read a cooking recipe posted by our vegetarian friend, ignore anything political or challenging/contradictory to our world view. The study measures the amount of different information one gets, not different world-views. Third, the users might refrain from novel information if they consider it to be offensive or distasteful to their (strong or weak) ties. Fourth, even if users are shown novel information, this does not mean they will be exposed to it. They might simply choose to ignore challenging items. Fifth, the information intermediary might filter out the novel content provided by our weak ties. If, for instance, Facebook decides which updates you see on your wall based on the frequency of an interaction, weak ties might as well disappear, as the user will not interact very often with a weak tie. At the moment the only way to prevent this is to manually click on each and every user and choose "show me all updates from this user". Otherwise Facebook will make a decision on what is important based on some unknown criteria.

## 2.6. CONCLUSION

Gatekeeping theory acknowledges the increasing popularity of social networking, on-line information seeking and information sharing services. It is often claimed that since users can select and share information online, they can be gatekeepers for each other. This then diminishes the power of media professionals. However, in this paper we have shown that even though the traditional gatekeepers might become less important, users are not becoming the sole gatekeepers. The gates are certainly not disappearing. Platforms on which users operate have an influence; they are one of the new gatekeepers. Online gatekeeping services are not just algorithms running on machines; they are a mix of human editors and machine code designed by humans. People affect the design of the algorithms, but they also can also manually influence the filtering process after the algorithm has been designed. Therefore, switching from human editing to algorithmic gatekeeping does not remove all human biases. Technical biases such as third party manipulation or popularity will exist due to the computerized form of gatekeeping. Also, individual factors such as personal judgments, organizational factors such as company policies, external factors such as government or advertiser requests will still be present due to the role of humans in providing these services.

In this paper, we introduced a model of algorithmic gatekeeping based on traditional gatekeeping model and focused on particular filtering processes including personalization. We show that factors that caused bias in mass media news selection still play a role in information selection in online web services. We have shown that search results in Google can differ, but an extensive empirical research is needed to determine the extent of so-called "echo chambers" in social networks. What percentage of information do users miss or feel like they are missing if they turn on a personal filter or an inter-personal

filter? Is there enough variety in their choice of friends? Are users aware of these algorithms? Do they modify their filter periodically or switch to other forms of information sources? Are there routines that are used in the design of personalization algorithms, just like routines used in traditional gatekeeping? How does the introduction of implicit and explicit filtering algorithms affect user trust in systems and user autonomy? More research is needed in order to answer these questions.

# 3

## VIEWPOINT DIVERSITY IN ONLINE SOCIAL NETWORKS - AN EMPIRICAL ANALYSIS

*One of the things that amazes me about Twitter is the way it utterly eradicates artificial barriers to communication. Things like status, geopolitics and so on keep people from talking to one another. Those go away in Twitter.*

Dick Costolo, CEO of Twitter

### 3.1. INTRODUCTION

It is well known that traditional media have a bias in selecting what to report and in choosing a perspective on a particular topic. Individual factors such as personal judgment can play a role during the selection of news for a newspaper. Selection bias, organizational factors, advertiser and government influences can all affect which items will become news [45]. About 37% of Americans see a great deal of political bias in news coverage and 68% percent prefer to get political news from sources that have no particular point of view [265]. Similarly, in a survey performed before general elections in the UK, 96% of the population said they believe they have seen clear bias within the UK media [351]. Evidence of bias ranges from the topic choice of the New York Times to the choice of think-tanks that the media refer to [79].

Many democracy theorists claim that modern deliberative democracy requires citizens to have socially validated and justifiable preferences. Institutional designs must show particular attention to procedures of preference formation and learning within politics and civil society. Citizens must be exposed to opposed preferences and viewpoints and be able to defend their views [90, 158, 248]. Exposure to biased news information can foster intolerance to opposing viewpoints, lead to ideological segregation

---

This chapter has been published in *Computers in Human Behavior* [98]

and antagonisms in major political and social issues [14, 127, 280]. Being aware and overcoming bias in news reporting is essential for a fair society, as media has the power to shape voting behavior and a democratic society [280].

Social information streams, i.e., status updates from social networking sites, have emerged as a popular means of information awareness. Political discussions on these platforms are becoming an increasingly relevant source of political information, often used as a source of quotes for media outlets [171]. Traditional media are declining in their gatekeeping role to determine the agenda and select which issues and viewpoints should reach their audiences [49]. Internet users have moved from scanning traditional mediums such as newspapers and television to using the Internet, in particular social networking sites [14]. Social networking sites are thus acting as gatekeepers [45]. A small number of users, who are critically positioned in the structure of Twitter, can determine political communication with their own political perspectives [171]. These new gatekeepers exert strong and selective influence on the information passed within Twitter.

It is often argued that the Internet, by promoting equal access to diverging preferences and opinions in society, actually increases information diversity. Many scholars characterize the online media landscape as the "age of plenty", with an almost infinite choice and unparalleled pluralization of voices that have access to the public sphere [174]. Some argue that social media will disrupt the traditional elite control of media and amplify the political voice of non-elites and minorities [53]. The expansion of choice and participatory nature of the Internet not only ends the "scarcity" and "concentration" problems faced by the traditional media, but it also diminishes the power of gatekeepers.

Others claim that tools such as Twitter are neutral spaces for collaborative news coverage and curation operated by third parties outside the journalism industry. As a result, the information curated through collaborative action on such social media platforms should be expected to be drawn from a diverse, multi-perspectival range of sources [49]. Some further claim that platforms such as Twitter are neutral communication spaces, and offer a unique environment in which journalists are free to communicate virtually anything to anyone, beyond many of the natural constraints posed by organizational norms that are existing in traditional media [196]. Some argue that digital tools such as social media will inevitably lead to the pluralization of the public sphere and non-mainstream political actors can influence the political agenda thanks to the "multiacity" power of those platforms [358].

On the other hand, there are skeptical voices that argue that the Internet has not fundamentally changed the concentrated structure typical of mass media, but reflects the previously recognized inequalities [174]. It is also argued that it has brought about new forms of exclusion and hierarchy [316]. While it has increased some sort of political participation, it has empowered a small set of elites and they still strongly shape how political material is presented and accessed [162]. Others have pointed out the danger of "cyberbalkanization" caused by the Internet [256, 312]. They argue that the filters we choose on the internet, or the filters that are imposed upon us will weaken the democratic process. This is because it will allow citizens to join into groups that share their own views and values, and cut themselves off from any information that might challenge their beliefs.

Group deliberation among like-minded people can create polarization; individuals

may lead each other in the direction of error and falsehood, simply because of the limited argument pool and the operation of social influences. Increased polarization makes it more difficult for society to find common ground on important issues [312]. Research shows that ‘confirmation bias’ occurs when like-minded individuals form a group in order to make a decision [288]. When participants receive new information in a decision case after they have reached a preliminary conclusion, a clear preference was demonstrated for information supporting the preliminary group decision. Finally, the ability of online intermediaries such as recommender systems and social networks to customize their items to the taste of individuals, together with users’ preference to reading opinions which reinforce their own viewpoints, raises the phenomenon referred to as “filter bubble” [256].

In short, there is little consensus on whether the technological and socio-cultural changes in online media have actually increased diversity and plurality [174]. The interesting question is whether cyberbalkanization indeed occurs on online social networks. There are empirical studies that have observed a high level of information diversity in Twitter and Facebook, mainly due to retweets and weak-links [15, 26, 310]. While being very valuable contributions to the literature, these studies often focus on American users and they define information diversity either as “novelty”, or “source diversity”. However, as we will show later, novel information does not necessarily contribute to information diversity and highly competitive media markets with many sources may still result in excessive sameness of media contents. As we will argue, marginalized members of segregated groups, structurally underprivileged actors and minorities must receive special attention and just measuring number of available sources will not guarantee viewpoint diversity.

In this paper, in order to understand the impact of political culture, we analyze data from hundreds of political information sources and their followers in Twitter for two different countries. Naturally, we do not expect that the concept information diversity can be reduced to a single quantity or metric. Therefore, we first introduce different definitions of information diversity using the theory from communication studies and political philosophy. We provide a set of metrics that are based on this theory. Finally we present the result of an empirical study we performed using these metrics. Our main contributions are the following:

- We perform a conceptual analysis of the value “information diversity”.
- We demonstrate multiple metrics that capture different biases in a large sample of Twitter users from the Netherlands and Turkey.
- We show that, if information diversity is defined as “minority access”, we do observe “bubbles” in users’ newsfeeds

### 3.2. EMPIRICAL STUDIES OF INFORMATION DIVERSITY IN SOCIAL MEDIA

An empirical study performed by Facebook suggests that online social networks may increase the spread of novel information and of diverse viewpoints. According to Bakshy (2012), even though people are more likely to consume and share information that

comes from close contacts that they interact with frequently, the vast majority of information comes from contacts that they interact with infrequently. These so-called "weak-ties" [136] are also more likely to share novel information. However, there are some concerns with this study. First, Facebook does not provide open access to everyone, thus we may not repeat or reproduce the results using Facebook data. Second, our weak ties give us access to new stories that we would not otherwise have seen, but these stories might not be different ideologically from our own general worldview. They might be novel information, but not particularly diverse. The concepts serendipity, diversity and novelty are different from each other [310]. The Facebook research does not indicate whether we encounter and engage with news that opposes our own beliefs through links sent by "weak links".

Twitter, with its API, provides an excellent environment for information diversity research. An et al [2012] observe extreme polarization among media sources in Twitter. In another study, they found that, when direct subscription is considered alone, most Twitter users receive only biased political views they agree with [15]. However, they note that the news media landscape changes dramatically under the influence of retweets, broadening the opportunity for users to receive updates from politically diverse media outlets. Sun et al [2013] performed an empirical study using statistical models to identify serendipity in Twitter and Weibo. Using likelihood ratio test and by measuring unexpectedness and relevance, they observe that serendipity has a strong presence in information diffusion in microblogging communities. Saez-Trumper et al. [2013] found that political bias is evident in social media, in terms of the distribution of tweets that different stories receive. Further, statement bias is evident in social media; a more opinionated and negative language is used than the one used in traditional media. Twitter users are more interested in what is happening around them and what is happening to those around them. While communities talk about a broad range of news, Twitter users dedicate most of their time to a few of them [280]. Wei et al. [2013] found out that individual journalist have the strongest influence on Twitter for UK users. Further, they observed that all influential British Twitter users (mainstream media, journalists and celebrities) display some kind of bias towards a particular political party in their tweets.

### 3.3. THEORY

In this section, we first give a short overview "information diversity" and explain why it is a vital value for a democratic society. Later, we show different dimensions of this value and show how it can be defined.

#### 3.3.1. INFORMATION DIVERSITY

A cyberbalkanized Internet or "filter bubble" is not acceptable in a deliberative, liberal democracy. Whereas aggregative versions of democracy hold that legitimacy lies in the fair counting of votes for and against a decision, deliberative democrats hold that a decision is only legitimate if it is determined by a fair, informed discussion [106]. Because no set of values or preferences can claim to be correct by themselves, they must be justified and tested through social encounters which take the point of view of others into account [158]. In addition to the normative value of discussion, information-sharing is required

for many of the practical benefits that proponents of deliberation hope deliberative institutions will provide, such as higher quality policy, greater appreciation of the views of the opposing side, cultural pluralism and citizen welfare [243]. According to deliberative democrats, we must focus on why and how we come to adopt our views, and whether they can be defended in a complex social setting with people of opposed preferences. This will complement voting, the necessary mode of participation, by a “conscious confrontation of one’s own point of view with an opposing point of view, or of the multiplicity of diverse viewpoints that the citizen, upon reflection, is likely to discover within his or her own self” [248]. Under conditions of ideal deliberation, “no force except that of the better argument is exercised” [149].

Information diversity is also an important concept in communication studies. The freedom of media, a multiplicity of opinions and the good of society are inextricably mixed [243]. *Free Press theory*, a theory of media diversity, states that we establish and preserve conditions that provide many alternative voices, regardless of intrinsic merit or truth, with the condition that they emerge from those whom society is supposed to benefit its individual members and constituent groups [338]. What is good for the members of the society can only be discovered by the free expression of alternative goals and solutions to problems, often disseminated through media [243].

While many scholars from different disciplines agree that information diversity is an important value that we should include in the design of institutions, policies and online services, this value is often reduced to a single definition, such as “source diversity”, or “hearing the opinion of the other side”. In the next subsections, we explain that just having a deliberation is not enough, and a bias against arguments made by deliberators who are in the minority in terms of their interests in the decision being made can exist.

### 3.3.2. DIMENSION OF INFORMATION DIVERSITY

Following Napoli [1999], we may distinguish three different dimensions of diversity. The first dimension is *source diversity*, which is diversity in terms of outlets (cables and channel owners) or program producers (content owners). It is assumed that if source diversity increases, the second dimension *content diversity* will also increase. Content diversity consists of diversity in format (program-type), demographic (in terms of racial, ethnic, and gender), and idea-viewpoint (of social, political and cultural perspectives). The third dimension *exposure diversity* deals with audience reach and whether users have actually consumed the offered items diversely.

In the US, with the “free marketplace of ideas” theory, it is assumed that increasing source diversity will increase content diversity and exposure diversity will follow these two. American media policy consequently focuses on source diversity by way of competition and antitrust regulation [336]. However, whether more media competition (more sources) really brings about more media variety is a highly debated question and research addressing this relationship has not provided definitive evidence of a systematic relationship [175, 219, 243, 335]. Highly competitive media markets may still have low content diversity and media monopolies can produce highly diverse supply of media content [336]. It has also been argued that to fulfill the objectives of the marketplace of ideas metaphor, policymakers need to focus on exposure diversity. So, one should not look at availability of different sources or content, but whether the public consumes

them diversely [243].

### 3.3.3. MINORITIES AND OPENNESS

According to van Cuilenburg [1999], media diversity has to be externally benchmarked in some way and should always be compared with relevant variations in society and social reality. Computer Science researchers often use popularity (i.e., trending topics) or locality (items posted by friends and friend of friends) to determine the importance of news items [57]. Some use number of available sources to measure diversity [15], mirroring the “free marketplace” approach of diversity used in American media policy, which is based on the idea of competition and freedom of choice [336].

However, Karppinen [2009] argues that the aim of media diversity should not be the multiplication of genre, sources or markets, but giving voice to different members of the society. We should not see diversity as something that can be measured through the number of organizations or channels or just “having two parties reach all citizens”. Karppinen holds that we should focus on democratic distribution of communicative power in the public sphere and whether everyone has the chance and resources to get their voices heard. Karppinen argues: “the key task for media policy from the radical pluralist perspective is to support and enlarge the opportunities for structurally underprivileged actors and to create space for the critical voices and social perspectives excluded from the systematic structures of the market or state bureaucracy” [174]. If democratic processes and public policies exclude and marginalize members of segregated groups from political influence to the extent that privileged groups often dominate the public policy process, they will magnify the harms of segregation. These “minorities” must be politically mobilized and included as equals in a process of discussing issues [369].

McQuail and van Cuilenburg [1983] propose to assess media diversity by introducing two normative frameworks. The norm of *reflection* checks whether “media content proportionally reflects differences in politics, religion, culture and social conditions in a more or less proportional way”. The norm of *openness* checks whether media “provide perfectly *equal access* to their channels for all people and all ideas in society”. If the population preferences were uniformly distributed over society, then satisfying the first condition (reflection) would also satisfy the second condition (equal access). However, this is seldom the case [335]. Often population preferences tend toward the middle and mainstreams. In such cases, the media will not satisfy the openness norm, and the preferences of the minorities will not reach a larger public. This is undesired, because “social change usually begins with minority views and movements (...) asymmetric media provision of content may challenge majority preferences and eventually may open up majority preferences for cultural change in one direction or another”. [335]. Van Cuilenburg [1999] argues that the Internet has to be assessed in terms of its ability to give open access to new and creative ideas, opinions and knowledge that the old media do not cover yet. Otherwise it will only be “more of the same”.

### 3.4. POLARIZATION IN THE NETHERLANDS AND TURKEY

Before discussing methods and the results of our empirical study that focused on Dutch and Turkish users, we give a short overview of political diversity for two countries and

explain why they are interesting for a case study of information diversity.

### 3.4.1. THE NETHERLANDS

Pillarization (Dutch: “verzuiling”) is a process that occurred in the Netherlands and reached its highest point in 1950’s. During this period, several ideological groups making up the Dutch society are systematically organized as parallel complexes that are mutually segregated and polarized [268, 347]. As part of this social apartheid dividing the population into subcultures, political parties were used for political mobilization of the ideologically and religiously defined groups and social activities were concentrated within the particular categorical group [303]. Few contact have existed between different groups and internally the groups were tightly organized [204]. Elites at the ‘top’ level communicated, while the ones at the ‘bottom’ did not. Pillarization had an effect on parental choice of an elementary school for children, the voting for political parties and the choice on which daily newspaper to read [191]. People belonging to a pillar retreated into their own organizations and entered into a ‘voluntary’ isolation, because they perceive that values important to them are threatened [218].

Depillarization (Dutch: “ontzuiling”) started in mid 1960’s as a democratization process and pillarization has lost much of its significance since the 1960s as a result of secularization and individualization. Even though depillarization has started, many institutional legacies in present-day Netherlands still reflect its pillarized past, for example in its public broadcasting system or in the school system [349]. The Netherlands continues to be a country of minorities, which may be a main reason that consensus seems so ingrained in the Dutch political culture [263]. The Dutch parliament has 12 political parties. Due to the very low chance of any party gaining power alone, parties often form coalitions.

Netherlands has created several media policies set afterwards to implement diversity in the media. The Media Monitor, an independent institution, measures ownership concentration, editorial concentration and audience preferences [19]. It also measures diversity of television programming on the basis of a content classification system, by categorizing program output in categories like news and information, education, drama, sports, etc. [221, 336].

### 3.4.2. TURKEY

Turkey has regularly held free and competitive elections since 1946. The country has alternated between a two-party political system and a multi-party system. Electoral politics has often been dominated by highly ideological rival parties and military interventions changed the political landscape several times [321]. Elections in 2002 led to a two-party parliament, partially due to a ten per cent threshold. The Justice and Development Party (AKP) won the elections and still is the ruling party, having an absolute majority. The parliament is currently formed by 4 political parties. While AKP has 59% of the MP’s, secular CHP has 24%.

AKP’s dominance and the despair and sense of marginalization felt by its opponents threaten to create a political polarization along with fierce institutional clashes between the AKP government and the secular elites that retain a foothold in the military, the judicial system, and other parts of the bureaucracy. Muftuler-Bas and Keyman [2012] argue

that "many other polarizing social and political struggles remain unresolved in Turkey, and mutually antagonistic groups remain unreconciled. This social and political polarization remains potentially explosive and reduces the capacity for social consensus and political compromise". Similarly Unver [2011] claims that "the society is pushed towards two extremes that are independent of party politics. (...) Competing narratives and "realities" clash with each other so intensely, that the resultant effect is one of alienation and "other-ness" within the society."

Some scholars argue that, the top-down imposition of concepts such as democracy, political parties and parliament as part of westernization efforts is causing the socio-political polarization in Turkey [12]. Agirdir [2010] argues that "the system does not breed from the diverse interests and demands of the society, but around the values and interests of a party leader and the narrow crew around her". Economic voting behavior, religiosity, and modern versus traditional orientation seem to be the strongest drivers of polarization [367]. Some argue that, after 2011 polarization has increased and reached its highest points in Turkish history [253]. Report of research group KONDA indicate that the polarization mainly occurs between three groups: religious conservatives, traditionalist conservatives and moderns [6]. The difference of opinion between different clusters about secularity, tolerance and political change issues in total contradiction of each other, therefore a danger of absolute social polarization is imminent [6]. Kiris [2011] observes an identity-based polarization, between secularists and islamists, between Turkish nationalists and Kurdish Ethnic Nationalists, and between Alevis and Sunnis (different sects of Islam).

Turkish Radio Television Supreme Council (RTUK) was established in order to regulate the private broadcasting and to control the compliance of the broadcasts with the legal framework. RTUK is granted with the authority of giving penalties (for breaching the legal framework) to the broadcasters, which may range from warning to the suspension of the TV and radio channels. RTUK, which is responsible for supervision of TV and radio programs in Turkey, consists of total twelve members. Three out of twelve are chosen and appointed by the President. RTUK makes sure that the constitutional language, which is Turkish, is used in programs and the elements of Turkish-Islamic ethics and Turkish-Islamic world view will be given a significant place. Further, it control whether Turkish language, Turkish history, historical values, Turkish way of life, thoughts and feelings are given a significant place in broadcasting programs [1].

RTUK is sometimes referred as "the Censure Board" [236] and its decisions of penalizing the broadcasters so as to implement the Radio and Television law have been criticized domestically and internationally [30, 80]. RTUK played a crucial role in linking legislation to implementation with respect to control of Kurd nationalist and Muslim religious broadcasts. It singled out Kurd nationalist and Muslim religious propaganda and enforced the most coercive penalties for this type of infraction. Almost all high intervals of cease-broadcast days were applied to broadcasts that aired separatist and religious propaganda [80]. RTUK does not have a diversity policy and the lack of diversity in programme-making is said to undermine the quality of the audio-visual media [30].

### 3.4.3. CONCLUSION

In short, the Netherlands and Turkey are two different countries if we consider the political landscape and diversity policy. The Dutch society is less polarized than it was half a century ago, while the Turkish society is thought to be heavily polarized. The Dutch Parliament contains many political parties, no party has absolute power to govern alone. Turkey, on the other hand has few political parties represented in the government and the ruling party has almost 60% of all the seats. Further, the Dutch media is regulated with a diversity policy. While Turkey has a similar institution, it acts more as a censor board and does not employ an active diversity policy. If the social networking platforms mirror the society, then we can expect the Dutch users to receive more diverse content, while the Turkish users to be more polarized and have a less diverse newsfeed.

## 3.5. METHOD

In this section we provide our method of data collection, present our research model and the metrics we have devised to measure information diversity.

### 3.5.1. DATA COLLECTION

In January 2013, over a period of more than one month we crawled microblogging data via the Twitter REST API<sup>1</sup>. We started from a *seed* set of Dutch and Turkish Twitter users  $U_s$ , who mainly publish news-related tweets. We have selected different types of users including mainstream news media, journalists, individual bloggers and politicians. The list of these “influential” users were picked up from different ranking sites. For the Dutch ranking, we used Peerreach<sup>2</sup>, Twittergids<sup>3</sup> and Haagse Twitter-stoIp<sup>4</sup>. For Turkish ranking, we used TwitterTurk<sup>5</sup> and TwitterTakip<sup>6</sup>.

By monitoring the Twitter streams of  $U_s$ , we were able to add another set of users  $U_n$ , who followed and retweeted at least 5 items from users in  $U_s$ . After removing users who were involved in spam, we had 1981 Dutch users and 1746 Turkish users. We mapped the political leaning of Dutch seed users into five groups and the political leaning of Turkish seed users into nine groups. We did this using a number of public data [52, 189, 323, 345]. The political stance in the landscape is determined by [30, 52, 190].

### 3.5.2. RESEARCH QUESTIONS

The main question in this research is the following: “Does political culture affect information diversity in Twitter?”. To answer this question, we have provided some sub-questions.

1. **Q1: Seed User Interaction** Do seed users from one end of the political spectrum ever tweet links from another category? Do they reply to each other? The results

<sup>1</sup><https://dev.twitter.com/docs/api/1.1>

<sup>2</sup><http://peerreach.com/lists/politics/nl>

<sup>3</sup><http://twittergids.nl/>

<sup>4</sup><http://alleplanten.net/twitter/site/de-resultaten/belangrijke-personen/>

<sup>5</sup><http://twitturk.com/twituser/users/turk>

<sup>6</sup><http://www.twittertakip.com/>

of this question is relevant to the previously conducted studies that studied media bias on Twitter, such as Wei et al. [2013]

2. **Q2: Source Diversity** Is the newsfeed of social media users diverse? Are they receiving updates from a diverse set of users? Does indirect exposure (e.g., via retweets or weak-links) increase diversity marginally? Result of these questions are relevant to the previously conducted studies, such as An et al's [2011].
3. **Q3: Exposure Diversity** Do users share items from a diverse set of users or mainly from the same political category? This question is relevant to the framework provided by Napoli, which we have mentioned in Section 3.3.2.
4. **Q4: Openness** Can minorities reach the social media users, so that "equal access" principle is satisfied? This question is relevant to the normative theory of [219] and [175], which we discussed in Section 3.3.3.
5. **Q5: Input-Output Correlation** Do users post political messages whose political position reflects the political position of those messages that the users receive? Or do the messages they chose to retweet show a political position significantly skewed from the political position of the messages which they receive? Result of this question is relevant to the previously conducted studies such as Jurgens et al.'s [2011].

### 3.5.3. ENTROPY

While translating the concepts introduced in the previous subsection into metrics, we apply the following entropy formula used by van Cuilenburg [2007] to measure traditional media diversity, which is based on the work of Shannon [1948] :

$$-(\sum p_i \log p_i) / -(\log(1/n)) \quad (3.1)$$

In [339], " $p_i$ " represents the proportion of items of content type category  $i$ .  $n$  represents number of content type categories. We use this formula for calculating source diversity and exposure diversity in our Twitter study. For instance in source diversity, " $p_i$ " represents incoming tweets from seed users with a specific political stance, while " $n$ " represents all possible categories. As a result of this formula, the user will have a diversity between 0 and 1, where 0 represents minimum diversity and 1 represents maximum diversity. Figure 3.1 shows a user that receives equal amount of tweets from all political categories and has an incoming diversity of 1. He only retweets from one political category, therefore he has an outgoing diversity of 0.

### 3.5.4. TRANSLATING RESEARCH QUESTIONS INTO METRICS

**Source Diversity** For each user we used Equation 3.1 to compare the tweets published by her direct followees (people he follows) from different groups of which the political leanings have been categorized as discussed above (See Figure 3.2). We then also added the tweets a user gets through retweets and investigated if the user receives more diverse information through indirect media exposure (See Figure 3.3).

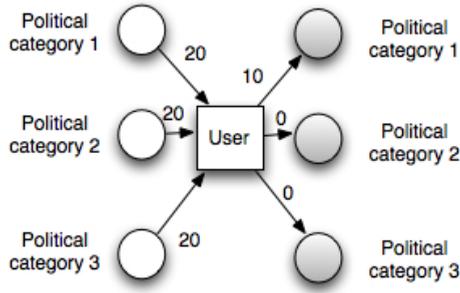


Figure 3.1: Applying entropy

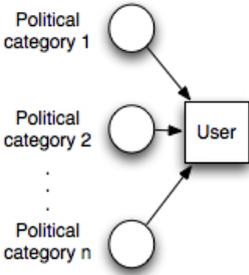


Figure 3.2: Direct source diversity

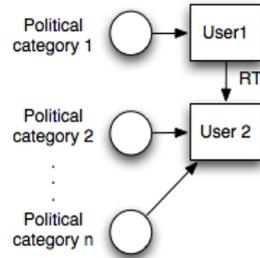


Figure 3.3: Indirect source diversity

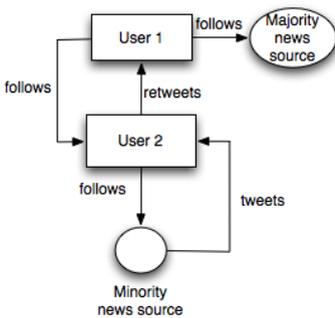


Figure 3.4: Minority access

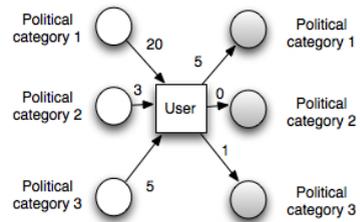


Figure 3.5: Input-Output Correlation

**Outgoing Diversity** To measure what the user is sharing after he was exposed to different incoming information, we used Equation 3.1 to compare the retweets he makes for each political category.

**Openness** For this definition of diversity, we first defined all seed users who belong to a political party that is either not represented in the parliament, or is represented with few MP's. We also included MP's of a large political party who belong to an ethnic minority. That makes for instance the Kurdish Party BDP and its MP's a minority in Turkey, while we consider the Greens as a minority in the Netherlands. See Appendix A for a list of minorities. Both users defined as minorities create about 15% of the all observed tweets for both countries.

We then looked whether the user is receiving minority tweets directly or indirectly (See Figure 3.4). We defined two metrics to measure minority access. We first look at the ratio of minority tweets a user gets out of all minority tweets:

$$\frac{\# \text{ received minority tweets}}{\# \text{ all published minority tweets}} \quad (3.2)$$

We later calculate the ratio of minority tweets in a users' timeline

$$\frac{\# \text{ received minority tweets}}{\# \text{ received tweets from seeds}} \quad (3.3)$$

**Input-Output Correlation** For each user in our sample we look whether the maximum number of the political position of the messages retweeted by a user is significantly skewed from the political position of the messages that she receives.

$$\max(\text{incoming political category}) = \max(\text{outgoing political category}) \quad (3.4)$$

For instance, Figure 3.5 shows a biased user which receives most items from category 1, and also retweets mainly from category 1.

## 3.6. RESULTS

This section shows the results for the defined metrics. We tested statistical significance of our results with a two-tailed  $t$ -Test where the significance level was set to  $\alpha = 0.01$  unless otherwise noted.

### 3.6.1. DISTRIBUTION OF SEED USERS AND THEIR FOLLOWERS

Figures 3.6 and 3.7 show the distribution of the seed users for both countries. Figures 3.8 and 3.9 show the distribution of regular users. We see that our selection of popular users covers the political spectrum and it is not concentrated on a single political category. We have used several sources to do the seed user categorization [30, 189, 189, 190, 323, 345]. We used the retweet behavior of the users to assign them to a political category to identify their political stance.

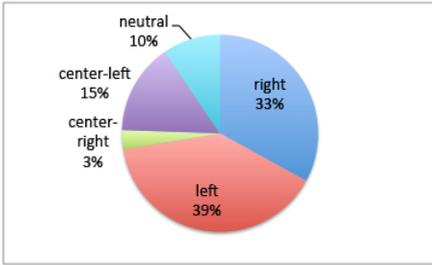


Figure 3.6: Dutch seed user distribution

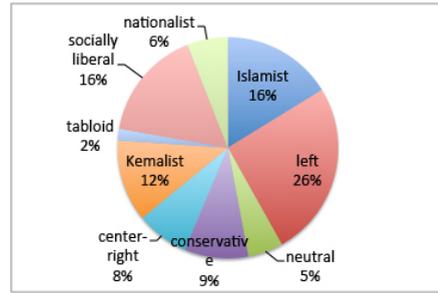


Figure 3.7: Turkish seed user distribution

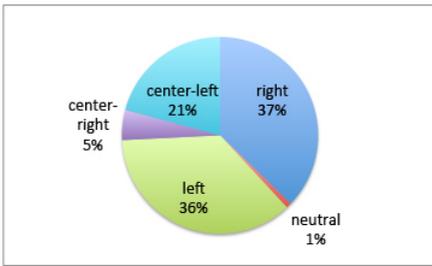


Figure 3.8: Dutch user distribution

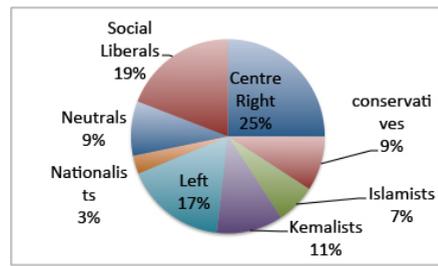


Figure 3.9: Turkish user distribution

**3.6.2. SEED USER BIAS**

Tables 3.1a and 3.1b show the retweet and reply behavior of seed users. Each row shows the category of users who retweet an item or reply to another user. The columns show the source of their retweet or the user they interact with. We observe that 73% of the left seed users retweet left items and reply to left users, while 72% of the right users do the same. The situation is more extreme for Turkish seed users: 93% of left seed users retweet from and reply to left, while 94% of the right seed users show the same behavior.

**3.6.3. SOURCE AND OUTPUT DIVERSITY**

Table 3.2a shows the results for research questions Q2 and Q3. Here we see that on a scale of 0 to 1, the diversity of the incoming tweets for an average user is approximately 0.6 and the results are not very different for both countries. While diversity is not perfect, we cannot really observe a true cyberbalkanization and we cannot observe a significant difference between two countries. We observe that indirect communication (retweets) does increase diversity, but not dramatically. Figure 3.10 and Figure 3.11 show the distribution of source diversity among users. We observe that, indirect communication decreases the number of polarized users who have a diversity approaching 0 for both countries. Approximately 27% of the Dutch and 29% of the Turkish users have an indirect diversity under 0.5. However, if we look at the diversity of an average user's output, we see much lower numbers. It is approximately 0.4 for both countries. Figure 3.12 shows the distribution of output diversity among the population. About 52% of the Dutch and 66% of the Turkish users have an output diversity lower than 0.5. We do not observe a big difference

Table 3.1: Seed user bias

(a) Netherlands

User / Source	Left	Right
Left	73%	27%
Right	72%	28%

(b) Turkey

User / Source	Left	Right
Left	93%	7%
Right	94%	6%

Table 3.2: Different dimensions of diversity

(a) Source Diversity (on a scale of 0 to 1)

	NL	TR
Direct	0.63	0.58
Indirect	0.68	0.62
Outgoing	0.43	0.40

(b) Input-Output Correlation

	NL	TR
# users	657	828
% users	33%	47%

(c) Minority Access

	NL	TR
minority reach	15%	2%
minority exposure	23%	2%
% users under <0.05 reach	14%	57%
% users under <0.05 exposure	23%	55%

between different countries.

### 3.6.4. MINORITY ACCESS

Table 3.2c shows the results for the research question Q4. First row, which we call “minority reach” shows the result for Equation 3.2 and the second row, which we call “minority exposure” shows the result for Equation 3.3. We observe that an average Dutch Twitter user will receive 15% of the produced minority tweets, whereas an average Turkish user will only receive 2% of them. Later, we observe that minority tweets make up 23% of an average Dutch users’ incoming tweets from seed users, while it only makes up 2% for a Turkish user. Figures 3.13 and 3.14 show the distribution of users for this metric. Here we observe a significant difference between two countries. About 55% of the Turkish users have a minority exposure under 0.05 and 57% of them have a minority reach under 0.05. The percentages are much lower for the Dutch users: 14% and 23% respectively.

### 3.6.5. INPUT OUTPUT CORRELATION

Table 3.2b shows the results for the research question Q5. The first row shows the number of “biased” users. These are users whose output correlates with their input. Such users make up 33% of the Dutch and 47% of the Turkish userbase. Further, if we only consider a bias towards a certain political category that is higher than 15% (for both input and output), 26% of the Dutch and 36% of the Turkish users show this behavior.

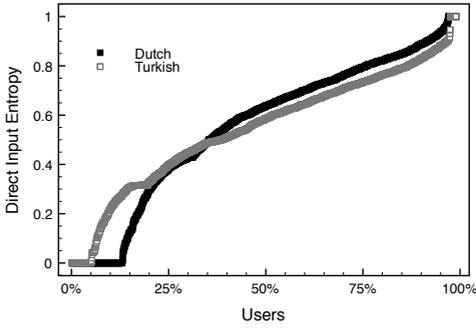


Figure 3.10: Direct source diversity

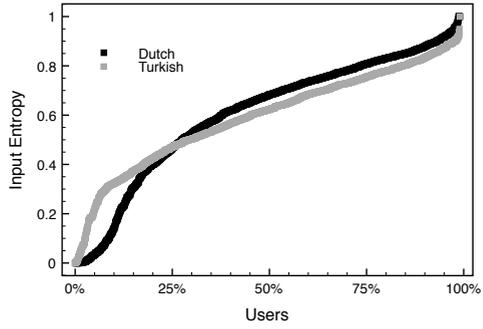


Figure 3.11: Indirect source diversity

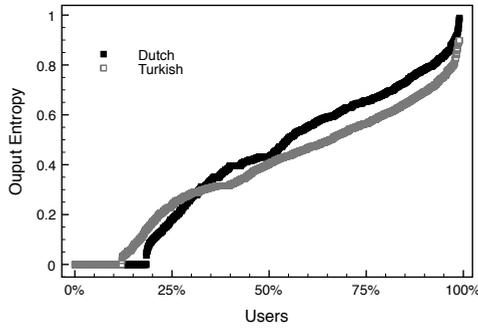


Figure 3.12: Output diversity

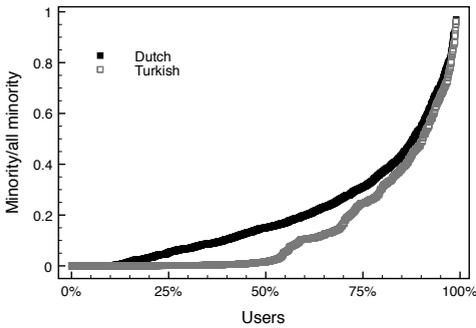


Figure 3.13: Minority Reach

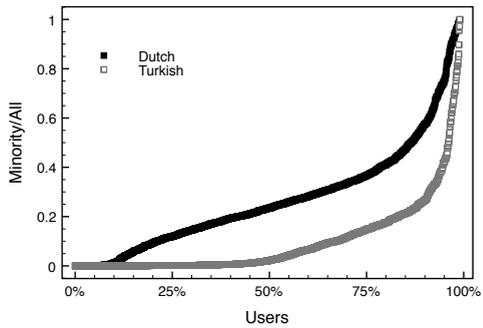


Figure 3.14: Minority Exposure

### 3.7. LIMITATIONS

This study has several limitations. First of all, next to the accounts of traditional media outlets on Twitter, we also selected politicians and bloggers. While they mainly tweet political matters, it is possible that they have shared personal and non political matters as well.

Second, while the results give us an idea on the political landscape of the studied countries, Twitter does not represent 'all people'. As boyd and Crawford [2011] have stated, "many journalists and researchers refer to 'people' and 'Twitter users' as synonymous (...) Some users have multiple accounts. Some accounts are used by multiple people. Some people never establish an account, and simply access Twitter via the web". Therefore we cannot conclude that our sample represent the real population of the studied countries.

Third, input-output correlation does not always implicate that the volume of the content affects the items users share. Users might already be biased before they select their sources and can therefore follow more from certain sources and share from certain categories.

Fourth, users will make different uses of Twitter. Some might use it as its primary news source, therefore following mainstream items, while others will use it to be informed of the opposing political view or to find items missing in the traditional media. Therefore, we do not know why some users only follow sources from a specific political category. More qualitative studies are needed.

### 3.8. DISCUSSION

In this study we have shown different dimensions of diversity and discussed another dimension, namely minority access. This dimension is often missing in the research performed by computer scientists. However, as many communication scholars and philosophers have argued, while the media should reflect the preferences present in the society, it should also allow equal access to everyone, including those whose common social location tends to exclude them from political participation. Public life needs to include differently situated voices to be able to articulate their concerns and interests, not just the ones who are in majority.

We have shown that different definitions of diversity can introduce different metrics and the question whether "the filter bubble exists" will have different answers depending on the metric and culture. For instance, according to the results of our study, source diversity does not differ much for Turkish and Dutch users and we certainly cannot observe a bubble. However, if we consider output, then we see that the diversity is much lower. Further, if we consider the minority access as a diversity metric, we see that minorities cannot reach a large percentage of the Turkish population.

In the abundance of digital information and filters to deal with information overload, ideas and opinions of minorities should not be lost. Design choices in software codes and other forms of information politics still largely determine the way information is made available and who can speak to whom under what condition [174]. According to Karpinen [2009], it is important to make decisions about standards, because those "can have lasting influence on media pluralism, even if they are not necessarily recognized as

sites of media policy as such". However, making minority voices reach a wider public is no easy matter. While identifying minorities and their valuable tweets is no easy task, showing these items to "challenge averse" users is a real challenge [238]. For instance Munson et al. [2013] provided people with feedback about the political lean of their reading behaviors and found that such feedback had only a small effect on nudging people to read more diversely. More research is needed to understand how users' reading behavior change and to determine the conditions that would allow such a change.

In the recent months, Turkey experienced several political protests that spontaneously erupted against the destruction of trees and the building of a shopping mall at Gezi Park in Taksim Square and large scale corruptions within the government. Twitter and Facebook played a vital role during these movements and became the only communication medium when traditional media performed self-censorship [89, 152, 251]. It would be very useful to see whether the political stance of our observed users have changed. It is also challenging to identify the opinion leaders during these movements and find whether they communicate with each other or form their own "bubbles". It is further valuable to see if minorities were able to reach a wider public during those protests. A hashtag based political communication and diversity analysis could bring new insights.

Our study was focused on Twitter and studied whether users have put themselves in bubbles by following individuals from only one end of the political spectrum and showed a biased sharing behavior. Twitter itself does not employ a personalization algorithm in a user's timeline. However other social networking platforms, such as Facebook, do use a personalization algorithm and filter certain information on user's behalf [45]. Future studies can perform black-box testing techniques to determine whether filters used by these platforms lead to bubbles (See [172]). Creating multiple profiles while modifying certain factors, such as political affiliation, age, location, etc. can help us detect bubbles, if they exist.

## APPENDIX A: LIST OF MINORITIES

Dutch minorities: Keklik Yucel, SGP, Khadija Arib, Vera Bergkamp, Sadet Karabulut, Farshad Bashir, Tanja Jadnanansing, Piratenpartij NLD, Partij van de Dieren, Fatma Koser Kaya, ChristenUnie, Groenlinks, Marianne Thieme, Femke Halsema.

Turkish minorities: Ayca Soylemez, Evrensel, Aydinlik, Ozgur Gundem, Pinar Ogunc, Bianet, Sebahat Tuncel, Sol Haber Portali, Halkin Gazetesi Birgun Yildirim Turker, Ufuk Uras, Selahattin Demirtas, Sirri Sureyya Onder, Sinan Ogan, Hasip Kaplan.

Note that both minorities create about 15% of all tweets produced by seed users.



# 4

## DEMOCRACY, FILTER BUBBLE AND DESIGN

*The overriding question, 'What might we build tomorrow?' blinds us to questions of our ongoing responsibilities for what we built yesterday.*

Paul Dourish and Scott Mainwaring

### 4.1. INTRODUCTION

Cyberbalkanization refers to the idea of segregation of the Internet into small political groups with similar perspectives to a degree that they show a narrow-minded approach to those with contradictory views. For instance Sunstein [2007] argued that thanks to the Internet, people could join into groups that share their own views and values, and cut themselves off from any information that might challenge their beliefs. This, according to Sunstein, will have a negative effect on the democratic dialogue. Recently others have argued that personalization algorithms used by online services such as Facebook and Google display users similar perspectives and ideas and remove opposing viewpoints on behalf of the users without their consent [256]. According to Pariser [2011], users might get different search results for the same keyword and those with the same friend lists can receive different updates. This is because information can be prioritized, filtered and hidden depending on a user's previous interaction with the system and other factors [45, 83]. This might lead to the situation in which the user receives biased information. In case of political information, it might lead to the situation that the user never sees contrasting viewpoints on a political or moral issue. Users will be placed in a "filter bubble" and they will not even know what they are missing [256]. As a consequence, the epistemic quality of information and diversity of perspectives will suffer and the civic discourse will be eroded.

---

This chapter has been accepted for publication in Ethics and Information Technology as "Bozdog E. and van den Hoven J. Breaking the Filter Bubble: Democracy and Design"

After Pariser's book has been published, the danger of filter bubbles received wide attention in the media, in academia and in industry. Empirical studies have been conducted to confirm or to debunk its existence. While algorithms and online platforms in general have been criticized because they cause filter bubbles, some designers have developed algorithms and tools to actually combat those bubbles. However, as we will show in this paper, the methods and goals of these tools differ fundamentally. Some try to give users full control and allow them to even increase their bubble. Some modify users' search results for viewpoint diversity without notifying the user. This is because the filter bubble has become a term that encompasses various criticisms. These criticisms differ because democracy is essentially a contested concept and different democracy models require different norms. As this paper will show, some will criticize the filter bubble due to its negative effect on user autonomy and choice, while others emphasize the diminishing quality of information and deliberation. In this paper we will show that while there are many different democracy theories, only the diversity related norms of a few of them are implemented in the tools that are designed to fight filter bubbles. We will also show that some norms (e.g., the inclusion of minorities in the public debate) are completely missing. We will argue that if we want to fully use the potential of the Internet to support democracy, all these diversity related norms should be discussed and designed, and not just the popular or most dominant ones.

In this paper, we first provide different models of democracy and discuss why the filter bubble poses a problem for these different models. Next, we provide a list of tools and algorithms that designers have developed in order to fight filter bubbles. We will do this by discussing the benchmarks these tools use and the democracy model the tools exemplify. We will show that not all relevant democracy models are represented in the overview of available diversity enhancing tools. Finally, we discuss our findings and provide some recommendations for future work.

## 4.2. DEMOCRACY: DIFFERENT THEORIES, DIFFERENT BENCHMARKS

Democracy refers very roughly to "a method of group decision making characterized by equality among the participants at an essential stage of the collective decision making" [60]. While some models of democracy emphasize the autonomy and individual preferences of those who take part in this collective decision making, others highlight the inclusion of free and equal citizens in the political community and the independence of a public sphere that operates as a middle layer between state and society [150]. Some emphasize the need of an informed (online) debate and the epistemic quality of information before decisions are made [153]. Others point out the need to increase the reach of minorities and other marginalized groups in the public debate [369].

While the filter bubble has been a concern for many, there are different answers to the question as to why filter bubbles are a problem for our democracy. The answer one gives to the question depends on one's understanding of the nature and value of democracy, on one's conception of democracy. Different democracy theories exist and they have different normative implications and informational requirements. A tool that implements one particular norm will be quite different in its form and goals than another

tool which implements a different norm. Before we provide examples of different tools, we will provide a framework of some basic conceptions of democracy and the relevant norms for each model.

#### 4.2.1. LIBERAL VIEW OF DEMOCRACY

The classical liberal view of democracy attempts to uphold the values of freedom of choice, reason, and freedom from tyranny, absolutism and religious intolerance [91, 158]. Liberalism started as a way to challenge the powers of “despotic monarchs” and the church. Once liberalism achieved victory over these old “absolute powers”, many liberal thinkers, began to express fear about “the rising power of the demos” [158, 213, 226]. They were concerned by the new dangers to liberty posed by majority rule against minorities and the risk of the majority tyrannizing over itself, leading to a need for people to ‘limit their power over themselves’.

Bentham [1780] argues that, since those who govern will not act the same way as the governed, government must always be accountable to an electorate called upon frequently and that electorate should be able to decide whether their objectives have been met. Next to voting, ‘competition’ between potential political representatives, ‘separation of powers’, ‘freedom of the media’, ‘speech and public association’ should be ensured to sustain ‘the interest of the community in general’ [37]. Individuals must be able to pursue their interests and goals without the risk of arbitrary political interference from the governing bodies, to participate freely in economic transactions, to exchange labor and goods on the market and to appropriate resources privately [158].

The liberal view of democracy is often criticized, because it construes democracy as an aggregation of individual preferences through a contest (in the form of voting), so that the preferences of the majority win the policy battle. However, this model has no way of distinguishing normatively legitimate outcomes from the preferences and the desires of the powerful, and makes no distinction between purely subjective preferences and legitimate and shared (quasi objective) judgments [64, 65, 369].

Filter bubbles are a problem according to the liberal view, because the non-transparent filters employed by online algorithms limit the *freedom of choice*. In addition, the liberal view states that citizens must be aware of different opinions and options, in order to make a *reasonable* decision. A filter imposed on a user –unbeknownst to them - will violate their autonomy, as it will interfere with their ability to choose freely, and to be the judge of their own interests. Further, the principle of *separation of powers* and the *freedom of the media* can also be in danger, if the algorithms are designed in such a manner as to serve the interests of certain individuals or groups. Finally, filters might damage the “liberty of thought”. Liberty of thought, discussion and action are the necessary conditions for the development of independence of mind and autonomous judgment. Liberty of thought creates reason and rationality, and in turn the cultivation of reason stimulates and sustains liberty. If one is ‘coerced’ by the filters, reason will also diminish. While some thinkers such as Mill [1859] also emphasize the diversity of opinion, most liberal thinkers do not mention this as a requirement. Liberal citizens must be ‘potentially’ informed so that the elected act accountably, but deliberation according to the liberal view is not necessary. Loss of autonomy caused by filters seems to be the main issue, according to the liberal view, while diversity of opinions and perspectives is not a concern.

### 4.2.2. DELIBERATIVE DEMOCRACY

Elster [1997] characterizes deliberative democracy as “decision making by discussion among free and equal citizens”. Deliberative democrats propose that citizens address societal problems and matters of public concern by reasoning together about how to best solve them. This can be made possible by deliberative procedures, which help to reach a moral consensus that satisfies both rationality (defense of liberal rights) and legitimacy (as represented by popular sovereignty) [148]. Individuals participating in the democratic process can change their minds and preferences as a result of reflection. According to Cohen [2009], deliberative democracy can be seen (1) as a matter of forming a public opinion through open public discussion and translating that opinion into legitimate law; (2) as a way to ensure elections are themselves infused with information and reasoning; (3) as a way to bring reasoning by citizens directly to bear on addressing regulatory issues. In all cases the goal is to use the common reason of equal citizens who are affected by decisions, policies or laws, instead of having them enter into bargaining processes or represent them by means of the aggregation of their individual preferences. Democracy, no matter how fair, no matter how informed, no matter how participatory, does not qualify as deliberative unless reasoning is central to the process of collective decision making.

There are different versions of deliberative democracy [332]. Rawls’ [1971, 1997] conception of deliberation is based on the idea of public reason, which is defined as “the basic moral and political values that are to determine a constitutional democratic government’s relation to its citizens and their relation to one another”. By means of public deliberation, people settle their disputes with respect and mutual recognition towards each other. Habermas [1998] provides similar conditions in his “ideal speech situation”. The Rawlsian approach aims at ‘accommodation’ of differences in a pluralistic society without criticizing people’s fundamental views of life, their so-called ‘comprehensive doctrines’ or ‘bringing them into deliberative discussion’. Habermas’ approach does the opposite, by also making moral or philosophical ideas and ideals part of the deliberative challenge. Both Rawls and Habermas advocate a ‘rational consensus’ rather than ‘mere agreement’ in political deliberation. For this purpose, Rawls uses the term ‘reasonable’, and Habermas introduces the notion of ‘communicative rationality’.

Deliberative democrats argue that deliberation 1) enlarges the pools of ideas and information [65]. 2) helps us discover truths [216, 317]. 3) can lead us to a better grasp of facts [153]. 4) can lead us to discover diverse perspectives, practical stances towards the social world that are informed by experiences that agents have [41]. 5) can help us discover the seriousness of our disagreements and discover that there is a disagreement after all [63]. 6) can lead to a consensus on the “better or more reasonable” solution [195]. 7) promotes justice, as it requires full information and equal standing 8) leads to better epistemic justification and legitimacy than simply voting [153]. This is because political decisions based on deliberation are not simply a product of power and interest. It involves public reasons to justify decisions, policies or laws. 9) leads to better arguments, since a citizen has to defend his proposals with reasons that are capable of being acknowledged as such by others [65]. 10) allows citizens to reflect on their own arguments, that will lead to self-discovery and refined arguments [63]. 11) promotes respect, as it requires people to consider the opinions of others, despite fundamental differences

of outlook [153].

Critics of deliberative democracy argue that full fledged deliberation is difficult to attain because 1) there is inequality in deliberative capabilities of citizens, which gives advantages to the rhetorically gifted and those who possess cultural capital and argumentative confidence in leading the discussions [7]. 2) there is widespread incompetence and political ignorance among masses [7]. 3) voters are not interested in the common good, but only in self-interests [51]. 4) people are biased and may hold beliefs without investigation. Majority rule will amplify these mistakes and make democratic decisions worse [51]. 5) While participation of citizens is possible in small nations, vast numbers of people will inevitably entail deterioration of participation [158]. Past a certain threshold, deliberation turns into a chaotic mess [195]. 6) Most citizens cannot spend the time to master the issues well enough to take meaningful stands on major issues. The information processing cost and transaction cost is too high [343]. 7) Deliberation among like-minded users can cause polarization. When people deliberate on a relatively homogenous argument pool, they consolidate fairly easily, which is bad for outsiders. Evidence from social psychology suggests that it is the viewpoints of the majority, not of the informed minorities, that can be expected to drive the relevant group judgments [7]. The informed minorities may refrain from disclosing what they know due to social pressure and be reluctant to dissent, thus not submitting the information to deliberation [314]. 8) Forcing participants to deliberation with limiting their arguments due to commonly shared rational premises, public reason or common good will prevent dissenting voices to share their perspectives and identities on their own terms [369].

Filter bubbles are a problem for deliberative democrats, mainly because of the low quality of information and the diminishing of information diversity. If bubbles exist, the pool of available information and ideas will be less diverse and discovering new perspectives, ideas or facts will be more difficult. If we only get to see the things we already agree with on the Internet, discovering disagreement and the unknown will be quite difficult, considering the increasing popularity of the Internet and social media as a source of political information and news [230]. Our arguments will not be refined, as they are not challenged by opposing viewpoints. We will not contest our own ideas and viewpoints and as a result, only receive confirming information. This will lead us not to be aware of disagreements. As a consequence, the quality of arguments and information and respect toward one other will suffer.

### 4.2.3. REPUBLICANISM AND CONTESTATORY DEMOCRACY

In contemporary political theory and philosophy, republicanism focuses on political liberty, understood as non-domination or independence from arbitrary power. The republican conception of political liberty defines freedom as a “sort of structural independence, the condition of not being subject to the arbitrary or uncontrolled power” [264]. Pettit [1999] argues that people are free to the extent that no other group has “the capacity to interfere in their affairs on an arbitrary basis”. To ensure that, according to Pettit [1999], there must be an “active, concerned citizenry who invigilate the exercise of government power, challenge its abuses and seek office where necessary”. In this theory, freedom as non-domination supports a conception of democracy where contestability takes the place usually given to consent. The most important implication is not that the

government does what the people want, but that people can always contest whatever decision the government has taken. While the republican tradition does not overlook the importance of democratic participation, the primary focus is clearly on avoiding the evils associated with interference and oppression.

Pettit [1999] argues that the media has a major role in forming the public opinion, ensuring non-domination and the possibility of effective contestation. However, Pettit argues, the media often fail badly in performing these roles. According to Pettit, at every site of decision-making (legislative, administrative and judicial), there must be procedures in place to “identify and display the considerations relevant to the decision”. The citizens should be able to contest these decisions if they find that the considerations did not actually determine the outcome. The decisions must be made “under transparency, under threat of scrutiny, and under freedom of information”. A group, even if they are a minority, should be able to voice contestation and must be able to speak out in a way that is liable to affect the proposed legislation. They must be able to contest in an effective manner, and they must be able to make themselves heard in decision-making quarters. To provide this, there must be reliable channels of publicity and information in place, so that the performance of the governing parties is systematically brought to attention.

If we apply these norms to the design of online platforms, we can argue that online information platforms 1) must make the right information available to the citizens and should allow them to track when something important or relevant happens. In this way, citizens can become aware of possible oppression and can become active when they feel there is a need to. This can for instance be achieved by human curation that aims at including important event that might affect the whole of society, in the information diet of everyone. It can also be achieved by means of personalization, so that, an event that is particularly important for a user can be highlighted for that user. 2) provide effective methods of contestation, so that citizens can make themselves heard with their contestations and affect the proposed legislation or policy. This means that people should not only be able to contest, but also that the contestation should reach a large public so that it can result in an effective and inclusive discussion.

Filter bubbles are a problem for advocates of contestatory democracy, because they interfere with realization of both conditions mentioned above. Bubbles both block the incoming and outgoing information channels. In order to raise critical questions, one must be aware of something that is a candidate for contestation. Someone cannot protest if they do not know that things relevant to them are happening. A filter bubble can block the reliable channels of publicity and information and may increase the risk that citizens are unaware of important news. Filter bubbles prevent awareness of both the items that people could disagree with and the information on the basis of which they could justify their reasons for disagreeing. Furthermore it may turn out to be much more difficult to communicate and share ideas with *potentially* like minded others outside your filter bubble. For not every post or comment on Facebook will reach your followers and a website with key information might never make it to the top of one's Google's search results.

#### 4.2.4. AGONISM / INCLUSIVE POLITICAL COMMUNICATION

While most deliberative democracy models aim for consensus concerning a ‘common interest’, agonists see politics as a realm of conflict and competition and argue that disagreement is inevitable even in a well-structured deliberative democratic setting, and even if the ideal of consensus regulates meaningful dialogues [233]. According to these critics, different and irreconcilable views will coexist and such an overlapping final consensus can never be achieved. Having consensus as the main goal and the refusal of a vibrant clash of democratic but opposing political positions will lead to “apathy and disaffection with political participation” [232, 369]. According to Mouffe [2009], the aim of democratic politics according to advocates of this agonistic conception of democracy should not be seen as overcoming conflict and reaching consensus, because such a consensus would actually be a consensus of the hegemony.

The aim of ‘agonistic pluralism’ then, is to construct the opposing viewpoints in such a way that it is no longer perceived as an enemy to be destroyed, but as an ‘adversary’. Thus, conflict must be in center stage in politics and it must only be contained by democratic limits. An adversary is “somebody whose ideas we combat but whose right to defend those ideas we do not put into question” [233]. The difference with “deliberative democracy” is that ‘agonistic pluralism’ does not eliminate passions from the sphere of the public, in order to reach a consensus, but mobilizes those passions towards democratic designs. Democracy should then be designed so that conflict is accommodated and unequal power relations and hegemony in the society is revealed [87].

Mouffe [232] argues that although the advocates of deliberative democracy claim to address pluralism and the complexity of the society, their reference to reason and rationality tends to exclude certain groups from the political arena; therefore, they are essentially not pluralistic. Similarly, Young [369] argues that if consensus becomes the ultimate goal, some difficult issues or issues that only concern a minority might be removed from discussion for the sake of agreement and preservation of the common good [369]. The idea of a generalized and impartial public interest that transcends all difference, diversity and division is problematic, because the participants in a political discussion most likely differ in social position or culture. Our democracies contain structural inequalities (e.g., wealth, social and economic power, access to knowledge, status). Some groups have greater material privilege than others, or there might be socially or economically weak minorities. Therefore in such settings “the common good” is likely to express the interests and perspectives of the dominant groups [369]. The perspectives and demands of the less privileged may be asked to be put aside for the sake of a common good whose definition is biased against them.

Young [2002] argues that when there are structural conflicts of interest which generate deep conflicts of interest, processes of political communication are more about struggle than about agreement. However, according to Young, the field of struggle is not equal; some groups and sectors are often at a disadvantage. Fair, open, and inclusive democratic processes should then attend to such disadvantages and institutionalize compensatory measures for exclusion. Democratic institutions and practices must take measures explicitly to include the representation of social groups, relatively small minorities, or socially or economically disadvantaged ones. Disorderly, disruptive, annoying, or distracting means of communication are often necessary or effective elements in

such efforts to engage others in debate over issues and outcomes. Cristiano [2006] argues that due to cultural differences in society, deep cognitive biases make individuals fallible in understanding their own and other's interests and compare the importance of others' interest with their own. By default, people will fail to realize equal advancement of interests in society. Thus, special measures must be taken to make sure that equality is satisfied.

Filter bubbles are a problem for agonists and supporters of inclusive political communication, because they hide or remove channels through which opposing viewpoints can clash vibrantly. Minorities, and those who are disadvantaged due to structural inequalities need special exposure to be able to reach out with their voice to larger publics. However, filters that show us what we already agree with usually do not include such minority voices. If filters only show us what they consider "relevant" for us, then, the only way to reach a large public will be through advertisements or by gaming the filters. This will violate the inclusion norm of modern democracies, as only the wealthy who can afford such advertisements, or technologically advanced minds who can use algorithms to their own advantage or can game the algorithms of others and override them, will be able to express themselves.

4

#### 4.2.5. CONCLUSION

Table 4.1 summarizes the democracy models we have introduced, the benchmarks they require, the points of critique they imply concerning the phenomenon of Filter Bubble. Liberal democrats stress the importance of self-determination, awareness, being able to make choices and respect for individuals. Filter bubbles are a problem for the liberal democrats especially due to restrictions on individual liberty, restrictions on choice and the increase in unawareness. Deliberative democracy attempts to increase information quality, discover the truth, discover facts, discover perspectives and discover disagreements. This in the end leads to better epistemic justifications, better arguments and it increases legitimacy and respect towards one other. The filter bubble, according to deliberative democrats, hurts the civic discourse, mutual understanding and sensemaking. Contestatory democracy on the other hand focuses on channels that allow citizens to be able to contest effectively, if there is a need. It does not aim for deliberation, but it requires citizens to have key information on important issues, and be aware of the oppressors. In contestatory democracy, the media should thus provide reliable channels of publicity, so that the performance of the governing parties is systematically brought to attention and can be contested. The filter bubble is a problem for contestatory democracy, because it removes the reliable channels so that key information on both topics and grounds of contestation cannot be sent and received. Agonists criticize the consensus goal of deliberative democrats and argue that other norms such as inclusion should also be the goal of democracy. They argue that special attention must be paid to the voice of minorities and other disadvantaged members of society and by making sure that dissent is continuously present. The filter bubble is a problem for agonists, because it will silence radical voices, will only reflect the viewpoints and perspectives of the mainstream and it will change agonism to antagonism.

Model of Democracy	Norms	Criticism of the Filter Bubble
<b>Liberal</b>	<p>Awareness of available preferences</p> <p>Self-determination</p> <p>Autonomy</p> <p>Free media</p> <p>Adaptive preferences</p> <p>Respect human dignity</p>	<p>User is unaware of the availability of options</p> <p>User is restrained and individual liberty is curtailed.</p> <p>The media is not free, it serves the interests of certain parties.</p> <p>Powers are not separated (advertiser and the information provider are the same)</p>
<b>Deliberative</b>	<p>Discover facts, perspectives and disagreements</p> <p>Determine common interests</p> <p>Construct identity by self-discovery</p> <p>Refine arguments and provide better epistemic justifications Consensus</p> <p>Respect towards each other's opinions</p> <p>A collective spirit</p> <p>Free and equal participants</p> <p>Rationality</p>	<p>Epistemic quality of information suffers</p> <p>Civic discourse is undermined</p> <p>No need to have better epistemic justifications.</p> <p>Respect for other opinions is decreased.</p> <p>Legitimacy is more difficult to achieve. There is a loss of a sense of an informational commons</p> <p>Communication suffers as gaining mutual understanding and sense-making is undermined</p>

<b>Republican and Contestatory</b>	<p>Freedom from domination by oppressors</p> <p>Contest matters effectively</p> <p>Be aware of the oppressors</p>	<p>Diminishes one's ability to contest.</p> <p>Diminishes one's awareness of the oppressors and their potentially manipulative interventions</p>
<b>Agonistic / Inclusive Political Communication</b>	<p>Conflict rather than consensus</p> <p>Passions rather than rationality</p> <p>Struggle rather than agreement</p> <p>Inclusion</p>	<p>The adversary becomes the enemy</p> <p>The minorities are excluded from the democratic process, their voices are lost</p>

Table 4.1: Models of Democracy and Design criteria

### 4.3. SOFTWARE DESIGN TO COMBAT FILTER BUBBLES

Many activists, including Eli Pariser [2011] have suggested to users that they should sabotage personalization systems by erasing web history, deleting cookies, using incognito option, trying other search engines and fooling the personalization system either by entering fake queries or liking everything ever produced by your friends. However, these options are not only tedious, but they are bad for the user as well. As we will show in this section, personalization algorithms and other tools can actually also be designed and used to broaden a user's worldview.

As we have seen in Section 2, while filter bubbles should be seen as worrying developments in the digital world from the point of view of democracy, different conceptions and models of democracy point to different undesired consequences of such bubbles, ranging from loss of autonomy to the diminishing epistemic quality of information. In recent years, various tools have been developed by computer scientists either in the industry or in academia to fight filter bubbles. However, as designers hold different values and are assuming different models of democracy model either implicitly or explicitly, the tools they develop will reflect those values and democracy models. As Friedman [2006] argues, technology is not neutral and the values and biases that designers hold will manifest themselves in the end product.

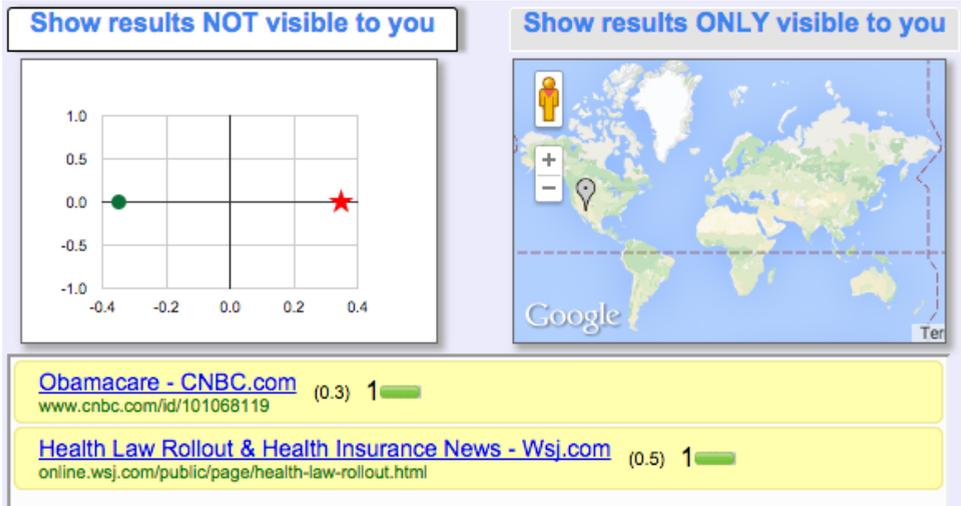
In order to identify the state of the art tools and designs and analyze which criteria and methods they employ, we have created a carefully curated list. To come up with this list, between January 2014 and June 2014, we have performed the following inquiries: 1) we have checked the academic articles that cite Munson and Resnick [2010], one of the first papers that designed an experiment and created a tool to fight the filter bubble, in the HCI community. 2) we have frequently followed HCI researchers on Twitter and included the tools/experiments they have mentioned on filter bubble. 3) We have used Google search engine with specific keywords to find non-academic tools, including filter bubble, "design", "selective exposure". This gave us in total 15 tools/designs.

In this section, we will show that, the different interpretations of the phenomenon filter bubble have led to different designs, tools and empirical studies. These tools differ in their goals ranging from personal fulfillment and development of cultural taste to promotion of tolerance and intercultural understanding. We will show that, some of the tools even allow the user to increase filter bubbles. The tools also differ in their methods, ranging from modifying users' newsfeeds/search results without their notice to visualizing bubbles to increase user awareness. We will show that, while their methods differ, the benchmarks they use to break the filter bubble can be the same. We will also show that, a design can include criteria from multiple democracy conceptions that we discussed in the previous section.

#### 4.3.1. LIBERAL / USER AUTONOMY ENHANCING

As we have stated in Section 2.1, in the liberal view of democracy, filter bubbles can be seen as a form of market failure that diminishes user control and hence autonomy, hide available options and coerce people in such a way that they cannot get what they want. Users will not get the search results they were looking for, or do not receive the updates from friends they want to in a social networking platform. Designers that take this view will develop tools that aim to promote awareness of filter bubbles and attempt to give





(a)



(b)

Figure 4.3: Bobble [365] displays a user Google search results that only he received (in yellow) and results that he missed but others have received (in red)

Xing et al [2014] developed a browser add-on called Bobble that allows the user to compare his Google search results with other profiles worldwide. The tool (See Figure 4.3) uses hundreds of nodes to distribute a user's Google search queries worldwide each time the user performs a Google search. For example, when a user performs a Google search with keyword "Obamacare", this search keyword is distributed to 40+ worldwide Bobble clients that perform the same Google search and return corresponding search returns. The user can then see which results are displayed on his browser, but not on others, and vice versa. It is a tool for users to get an idea of the extent of personalization taking place. The tool to increase user's awareness of Google's filters. However, it does not aim to increase deliberation or provide challenging information by its design.

Nagulendra and Vassileva [2014] developed a visualization design to display to users their filter bubbles (Figure 4.4). The tool helps users understand how information filtering works in an online peer-to-peer social network. The tool shows the user which categories and friends are in their bubble and which ones are not. Further, it allows

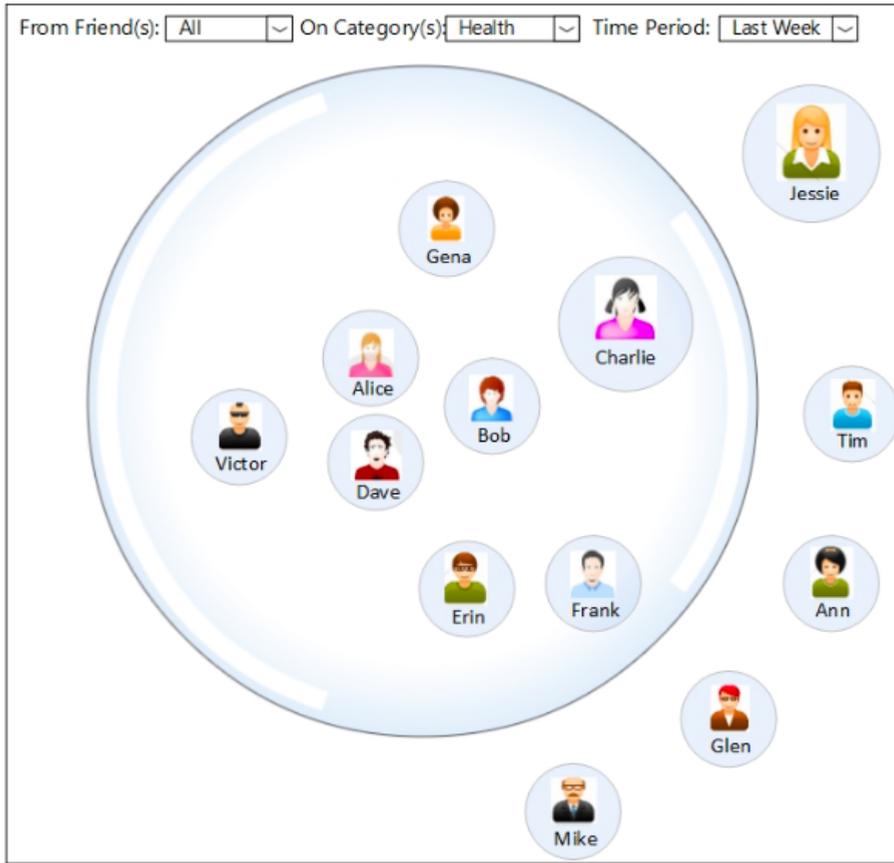


Figure 4.4: Nagulendra and Vassileva's software [2014] allows users to control their filter bubbles.

them to control the algorithm by manipulating the visualization to “escape” the bubble, namely adding/removing friends on a certain topic to the filters. The tool aims to maximize user control over his filter bubble, increase awareness of the filter bubble, promote understandability of the filtering mechanism and ultimately increase user satisfaction. It, however, does not make an attempt to expose users into challenging information. If the user wants to remain in a bubble, the tool will allow him to do that. Also in this case, a liberal notion of democracy with an emphasis on user autonomy is at the background of the development of this tool.

#### 4.3.2. DELIBERATIVE / ENHANCING EPISTEMIC QUALITY OF INFORMATION

As we have mentioned in Section 2.2, filter bubbles can be seen as a problem, not because they prevent users getting what they want, but because they diminish the quality of the public discussion. Deliberative democracy assumes that users are, or should be, exposed to diverse viewpoints, so that they can discover disagreements, truths, perspec-

tives and finally make better decisions. Polarized users or users exposed to low quality (but agreeable and relevant) information will have bad consequences. In order to increase epistemic quality of information, the wide range of opinions and perspectives on a particular topic may be made more visible and users can compare their opinions with others, even if they are opposing their own views. In the end, respect, legitimacy and consensus can be reached. In this subsection, we will list some of the tools that allow users to discover different viewpoints by visualization, showing pro/con arguments for a controversial topic, nudging them to listen to others, or by diversifying search results by modifying them for political search queries.

Microsoft search engine Bing studied the effect of used language for nudging Bing Search engine users [368]. In this study, a sample of 179,195 people who used news related queries were selected and then their political behavior and their link click pattern were observed. Researchers found that, while 81% (76%) of Republicans (Democrats) click on items from one of the most polarized outlets of their own view, they rarely clicked on polarized outlets of the other side (4% and 6% respectively), suggesting a filter bubble in action. The researchers then modified the Bing search engine's results page. They matched Democratic to Republican-leaning queries on the same topic manually (e.g., obamacare and affordable health care). They then modified the results for the queries for a subset of people who issued them (treatment group), resulting in a diversified set of results: the results contained items from both republican and liberal sources, regardless of what the user has searched for. This did not increase the number of clicks on items from the opposing political news outlets. However, when the authors chose websites that use a language similar to the user's own language, they observed a change of 25% toward the center. The authors thus conclude that when the language model of a document is closer to an individual's language model, it has a higher chance of being read despite it describing an opposite viewpoint. The researchers aimed for "increasing exposure to varied political opinions with a goal of improving (and enhancing) civil discourse" [368].

Considerit [111, 188] is a deliberation (pro/con) tool that is developed with the aims of (1) helping people learn about political topics and possible tradeoffs between different opinions (2) nudging them toward reflective consideration of other voters' thoughts (3) enable users to see how others consider tradeoffs. ConsiderIt (Figure 4.5) provides an interface where users can create pro/con lists by including existing arguments others have contributed, to contribute new points themselves, and to use the results of these personal deliberations to expose salient points by summarizing their stance rather than a yes/no vote. Users can see ranked lists of items that were popular full opposers, firm opposers, slight opposers, neutrals, slight supporters, firm supporters and full supporters. In a pilot study called "The Living Voters Guide" (LVG), the system was put into testing during 2010 Washington state elections that had certain proposals on areas of tax, sale of alcohol, candy or bottled water, state debt, bail and other political topics. In LVG, 8823 unique visitors browsed the site and 468 people submitted a position on at least one item. In a small survey of 7 users, 46.3% of them have reported that they have actually changed their stances on at least one measure and 56% of them saying they switched from support to oppose or vice versa. 32% of them have reported that they moderated their stances and 12% saying they strengthened them [188].

OpinionSpace [102] plots on a two-dimensional map the individual comments in a



Figure 4.5: Considerit [111, 188] helps people learn about political topics and possible tradeoffs between different opinions.

web forum, based on the commenters' responses to a short value-based questionnaire. By navigating this space, readers are better able to seek out a diversity of comments as well as prime themselves for engaging the perspective of someone with different values (Figure 4.6). When users interrogate an individual comment, they are prompted to rate comments for how much they agree with and respect it. The size of the comment's dot on the map then grows when people with different values than the speaker respect and/or agree with it, facilitating users in seeking out comments that resonate widely.

Reflect [187] modifies the comments of webpages in order to encourage listening and perspective taking. It adds a listening box next to every comment, where other users are encouraged to succinctly restate the points that the commenter is making, even if there is disagreement (Figure 4.7). This is a nudge to listen to other users. Other readers can afterwards read the original comment and other listeners' interpretations of what was said, supporting broader understanding of the discussion. This way, those who not have to "like" or "recommend" the comment to recognize or appreciate the speaker. By nudging towards listening and reflecting, an empathetic and constructive normative environment is formed, where not only those who speak and reflect are positively affected, but those who read as well. In mid-September 2011, popular online discussion platform Slashdot enabled Reflect on four stories. During the trial, 734 reflections were written by 247 discussants, an average of 1.0 reflection per comment. While flaming and pure replies were present (31%), the majority of the reflections were neutral, different neutral interpretations or meta observations. The tool also allowed the community to rate reflections, making certain reflections under a threshold invisible. After users downvoted flaming or cheeky replies on those reflections, almost 80% of all the visible reflections were neutral reflections.

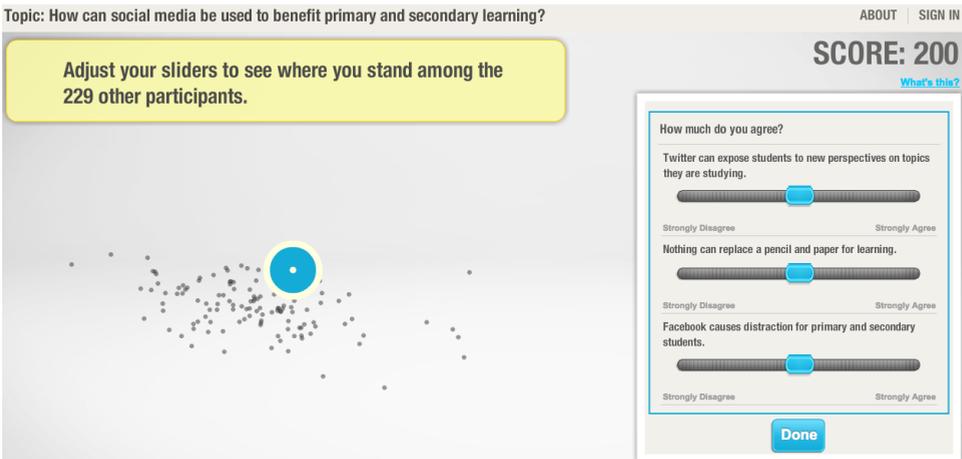


Figure 4.6: Opinionspace [102] allows users to browse a diverse set of ideas, see responses from like-minded participants or responses from participants who differ in opinion.

Rbutr<sup>2</sup> is a community driven Chrome add-on, that informs a user when the webpage they are viewing has been disputed, rebutted or contradicted elsewhere on the Internet (Figure 4.8). Users can add opposing viewpoints for an item, so that future users will see that an opposing viewpoint exists for the item they are reading. Rbutr aims to increase information quality and informed opinions by promoting fact and logic-checking.

There are other tools and studies that aim to increase epistemic quality of information. Liao and Fu [2013, 2014] studied the effect of perceived threat, the level of topic involvement, and the effect of expertise and position indicators. Munson and Resnick [2010] studied the effect of nudging by sorting or highlighting agreeable news items and experimenting with the ratio of challenging and agreeable news items. Newscube [258, 259] is a tool that detects different aspects of a news using keyword analysis, and displays users news items with different perspectives in order to decrease media bias. Hypothes.is<sup>3</sup> is a community peer-review tool that allows the users to highlight text and add comments and sentence-level critic. Political Blend [88] is a mobile application that matches people with different political views and nudges them to have a cup of coffee face to face and discuss politics.

Table 5.1 below summarizes our analysis of the studied tools.

## 4.4. DISCUSSION

One of the key finding of our analysis is that the benchmarks specified by agonistic and contestatory models of democracy are completely missing in all of the tools that aim to fight the filter bubble. While it is possible to come across critical voices, disadvantaged views or contestation using tools such as OpinionSpace or ConsiderIt, it is also highly likely that these voices and views get lost among the “popular” items, which are

<sup>2</sup><http://rbutr.com/>

<sup>3</sup><https://hypothes.is/>

Model	Examples	Design Criteria (Benchmarks)
Liberal	Balancer  Scoopinon  Bobble  Nagulendra and Vas-sileva's tool	Allow users to be aware of their own (and the platform's) biases  Allow users to understand biases  Allow the user to control incoming information and filters
Deliberative	Bing Study  Considerit  OpinionSpace  Rbutr, Newscube  Political Blend	Discover diverse facts, perspectives and disagreements  Reflection on own (and others') arguments  Aim for informed debate with epistemic justifications  Increase the epistemic quality of information

Table 4.2: Tools that are developed to combat filter bubbles, the benchmarks they use and the models they belong to

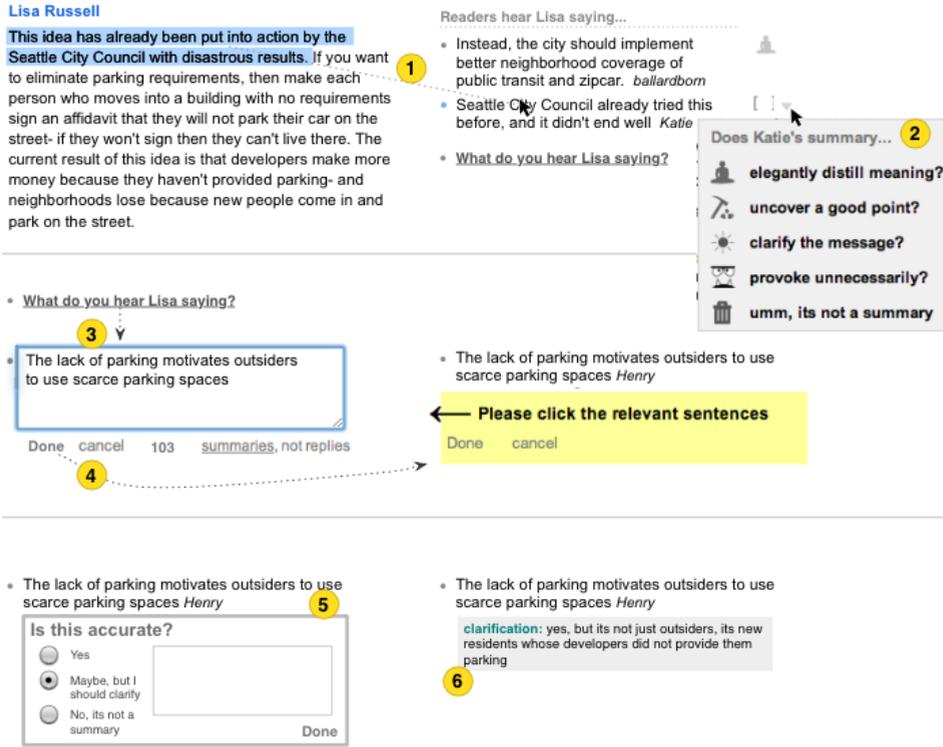


Figure 4.7: Reflect [187] nudges users to listen to each other by making them restate the points that the commenter is making, even if there is disagreement.

of interest for the majority of the audience. However, as McQuail and van Cuilenburg [1983] have argued, media should not only proportionally reflect differences in politics, religion, culture and social conditions, but provide equal access to their channels for all people and all ideas in society. If the population preferences were uniformly distributed over society, then satisfying the first condition (reflection) would also satisfy the second condition (equal access). However, this is seldom the case [335]. Often population preferences tend toward the middle and to the mainstream. In such cases, the media will not satisfy the openness norm, and the view of minorities will not reach a larger public. This is undesirable, because social change usually begins with minority views and movements [335].

In modern democracies, some citizens are able to buy sufficient media time to dominate public discussion, while others are excluded. If the political outcomes result from an exclusive process, where those with greater power or wealth are able to dominate the process, then from the point of view of democratic norms that outcome is illegitimate. However, even if people are formally included in the democratic process, inclusion issues arise if they are not taken seriously or treated with respect. The dominant party may find their arguments not worthy enough for consideration. Then, people, while

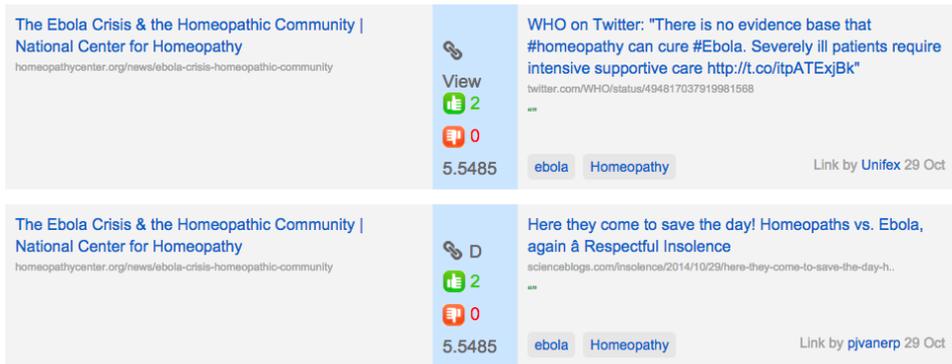


Figure 4.8: Rbutr is a Chrome add-on that informs a user when the webpage they are visiting has been disputed.

4

they formally have a chance to express their ideas, actually lack an effective opportunity to influence the thinking of others. Van Cuilenburg [1999] argues that the Internet has to be assessed in terms of its ability to give open access to new and creative ideas, opinions and knowledge that the old media do not cover yet. Otherwise it will only be more of the same. Recent research shows that equal access might be a problem on the Internet as well. Bozdag et al. [2014] studied the diversity of political information for Dutch and Turkish Twitter users, by analyzing about 2000 users for each country and studying around 10 million tweets. According to Bozdag et al. [2014], while minorities in both countries produce roughly the same amount of tweets, they cannot reach a very significant amount of Turkish users, while they can in the Dutch Twittersphere.

Several agonistic design attempts have been developed in the industry throughout the years to reveal hegemony (one of the requirements of agonistic design). Most of these tools perform social network analysis to identify actors and their connections (networks of force) and represent the multifaceted nature of hegemony. For instance the project *Mulksuzlestirme* (dispossession in Turkish) compiles data collectively and then uses mapping and visualization techniques to show the relations between the capital and power within urban transformations in Turkey. The interactive map (See Figure 8) displays the established partnerships between the government and private developers and shows to which investors collected taxes have been transmitted through the redevelopment/privatization of public spaces. For instance, it shows that one corporation that is involved in many government projects also owns major news organizations in the country, including the Turkish version of the CNN. By means of visualization, the designer allows users to browse and discover interesting relationships between the media and corporations to reveal hegemony.<sup>4</sup>

While tools such as *Mulksuzlestirme* might reveal key information for political debates and elections, many of these tools are not widely known. Tools like these can spread in unfiltered platforms such as Twitter, if powerful actors and opinion leaders can spread them through their followers [171]. However, Twitter has stated that it plans to deploy a personalized algorithmic timeline in the future [254]. If one wants their mes-

<sup>4</sup>See [87] for other examples

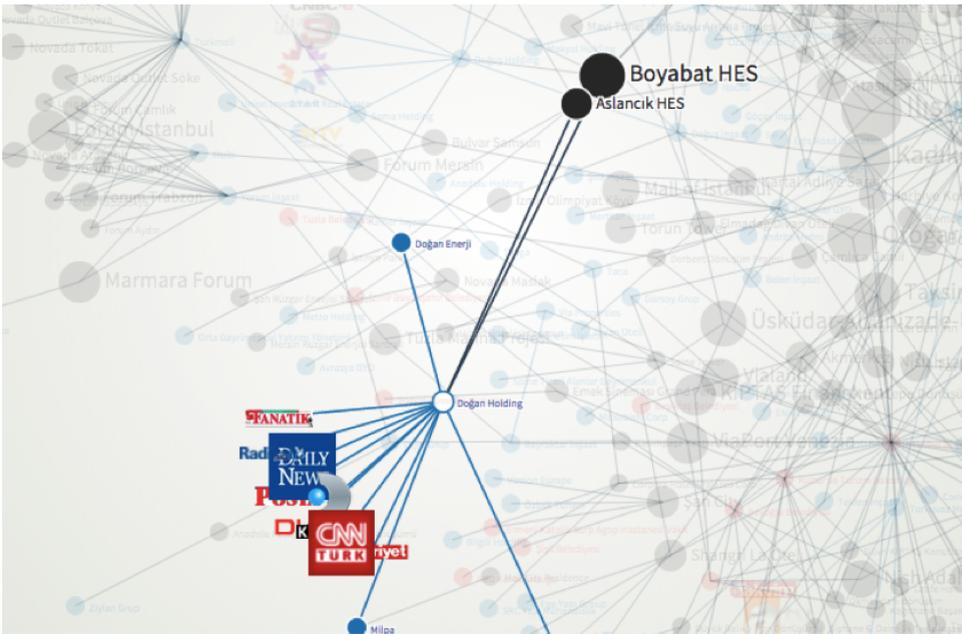


Figure 4.9: Screenshot from Mulksuzlestirme (Dispossession) project. The map shows the connections between a corporation, several media outlets that it owns and urban transformation projects that it has received.

sage to spread in a filtered/personalized platform, it has to bypass the filters or perhaps trick them. In order to accomplish this, one either has to pay for advertisements (and must hence possess the necessary financial means) or one must have the technical skills (such as search engine optimization). Many people do not have either of these means, but yet, they might have key information that is vital for contestation. Further, we could not find designs/tools that implement other benchmarks of agonism, such as special attention to minority voices.

We do not know why only norms of liberal and deliberative democracy models are represented in the tools that are developed to break filter bubbles. This might be due to a lack of designers' exposure to different democracy theories for the designers. It can also be the case that the designers are aware of all the models and implied norms, but choose to implement only certain ones in design. We have no evidence of reasoned choices to this effect on the part of the designers. Future work, such as interviewing the designers could shed some light into this issue. However, the body of literature concerning democratic theory shows that there is a great variety in conceptions of democracy, as one would expect with central philosophical notions, that we use to think about and order society, such as equity, justice, property, privacy and freedom. These are essentially contested concepts. As John Dewey has observed long before the Internet, social media and other platforms were invented, democracy is a central concept and it implies an ongoing cooperative social experimentation process [17]. Dewey was of the opinion that we live in an ever-evolving world that requires the continuous reconstruction

of ideas and ideals to survive and thrive. The idea of democracy is no exception in this respect [121]. Therefore, it seems that the online intermediaries that fulfill a public role must take necessary measures to open to and ready to experiment with a plurality of democracy models, including ones that propagate agonistic and contestatory elements. It is possible that these two models of democracy are not quite popular and that this explains that designers are not aware of the norms and benchmarks implied by these models. It might be beneficial if the designers are exposed to a variety of conceptions and models of democracies, in order to come to realize that each models has strengths and weaknesses.

An information intermediary could include agonistic and contestatory elements in its design by 1) Ensuring that minorities and other marginalized groups receive special attention, so that they can reach a larger audience. This must be designed carefully, as research shows that minority views are usually ignored by the majority and the alternative voice only has a formal, but not a meaningful place in the debate [359]. 2) Providing mechanisms and channels of publicity and, so that the performance of the relevant parties (e.g., the government) is known. This would include highlighting information on important political issues and put it in user's newsfeed/search result, even if the algorithm would normally not do so, in order to make users aware of the oppressors 3) Designing platforms for effective contestation. If key information is present, this must ideally reach the relevant users, so that they also can contest the decision makers. 4) Allowing people to be notified or alerted when something important/relevant happens, thus not only commercially relevant, but politically as well 5) Designing the tools in a way that opposing viewpoints are actually considered and reflected upon. Otherwise simply showing contradictory views might lead to flaming [85] 6) Emphasizing to the user that algorithmic selection is always a contest, one that is choosing from contrary perspectives. This could be done by showing that the selected viewpoint is a selection out of many possible ones [71]. 7) Always offering the ability to choose between real alternative viewpoints, not just the dominant ones.

Recent studies indicate that most people are unaware of filters in social media [100, 271]. We can thus expect that the tools that we have mentioned in this paper are not widely known. Major online platforms such as Google and Facebook often argue that they are not a news platform, that they do not have an editorial role and therefore they will not design algorithms to promote diversity. For instance, Facebook's project management director for News Feed states: "there's a line that we can't cross, which is deciding that a specific piece of information—be it news, political, religious, etc.—is something we should be promoting. It's just a very, very slippery slope that I think we have to be very careful not go down." [211]. However, research shows that these platforms are increasingly used to receive diverse opinions. According to a recent study in the US, nearly half (48%) of the 10,000 panelists say they accessed news about politics and government on Facebook alone in the past week [230]. A more recent study indicates that 86% of the Millennials usually turn to social media to receive diverse opinions, more than any other media [13]. Between 2010 and 2012, the traffic to news sites from various social media platforms grew by 57 percent and leading news organizations get around a quarter of site visits from the social networking platform, some even 40 percent [194, 225, 262]. If we also consider the dominant position of these platforms in the search and social me-

dia markets worldwide [278, 304, 354, 356], we can argue that these platforms are indeed important news and opinion sources.

If we consider these platforms as important news and opinion sources, then we can argue that they should aim to increase viewpoint diversity, a value that is deemed important by almost all democracy models. They could adapt and experiment with the tools that we have listed in Section 3. Experimenting seems unavoidable as the current design attempts to break the bubbles are all experimental. Breaking bubbles requires an interdisciplinary approach, as several disciplines including human-computer interaction, multimedia information retrieval, media and communication studies or computer ethics have all something to contribute in the design of diversity-sensitive algorithms. More experiments with different contexts will need to be conducted in order to find which techniques work and which do not. Once we have more concrete results, the systems could apply different strategies for different types of users. While these different designs to fight the filter bubble are very valuable to understand how users' attitudes can be changed to remedy polarization, the actual goal must be more explicit and must be better supported with theory and public deliberation. Otherwise, user autonomy might be diminished, and in turn, the honesty and trustworthiness of the platforms could be questioned.

## 4.5. CONCLUSION

In this paper, we have pointed out, that different democracy theories emphasize different aspects of the filter bubble, whether it is the loss of autonomy, the decrease in the epistemic quality of information, losing the ability for effective contestation or losing effective channels that display the performance of the governing bodies. Most tools that aim to fight the bubbles do not define the filter bubble explicitly. They also do not reveal their goals explicitly or simply define it as "hearing the other side". Further, most of those studies are performed for US politics. As some democracy theorists and communication scholars argue, viewpoint diversity is improved not only by aiming for consensus and hearing pro/con arguments, but also allowing the minorities and marginal groups to reach a larger public or by ensuring that citizens are able to contest effectively. As we have mentioned earlier, minority reach could be a problem in social media for certain political cultures.

Our findings indicate that the majority of the tools that we have studied to combat filter bubbles are designed with norms required by liberal or deliberative models of democracy in mind. More work is needed to reveal designers' understanding of democracy and to see whether they are aware of different norms. As we have shown in this paper, all models have their weaknesses. It would thus be beneficial if the designers were exposed to other conceptions of democracy to realize that there is not just one model. As democracy itself is an ongoing cooperative social experimentation process, it would be beneficial for all to experiment with different norms of different conceptions and theories of democracy and not just the popular ones.



# 5

## ETHICS OF ONLINE SOCIAL EXPERIMENTS

*The Internet is the first thing that humanity has built that humanity doesn't understand,  
the largest experiment in anarchy that we have ever had.*

Eric Schmidt, CEO of Google, 2010

### 5.1. THE FACEBOOK EMOTIONAL CONTAGION EXPERIMENT

The article 'Experimental evidence of massive-scale emotional contagion through social networks' by Adam D.I. Kramer (Facebook), Jamie E. Guillory (University of California) and Jeffrey T. Hancock (Cornell University) was published on 17 June 2014 in Proceedings of the National Academy of Sciences of the United States of America (PNAS), a highly competitive interdisciplinary science journal [185]. The paper tested the assumption that basic emotions, positive and negative, are contagious, that is, that they spread from person to person by exposure. This had been previously tested for face-to-face communication in laboratory settings, but not online, and not using a large random sample of subjects. The authors studied roughly three million English language posts written by approximately 700,000 users in January 2012. The experimental design consisted of an adjustment of the Facebook News Feed of these users to randomly filter out specific posts with positive and negative emotion words to which they would normally have been exposed. A subsequent analysis of the emotional content of the subjects' posts in the following period was then conducted to determine whether exposure to emotional content would affect the subjects. Kramer and colleagues stressed that no content was added to the subjects' News Feed, and that the percentage of posts filtered out in this way from the News Feed was very small. The basis for the filtering decision was the Linguistic Inquiry and Word Count (LIWC) software package, developed by James Pennebaker and colleagues, which is used to correlate word usage with physical well-being [261]. LIWC's

---

This chapter has been published in Internet Policy Review [270]

origins lie in clinical environments and originally the approach was tested using diaries and other traditional written genres, rather than short Facebook status updates [142]. The study found that basic emotions are in fact contagious, though the effect that the researchers measured was quite small. The authors noted that given the large sample, the global effect was still notable, and argued that emotional contagion had not been observed before in a computer-mediated setting based purely on textual content.

The article provoked some very strong reactions both in the international news media (e.g. *The Atlantic*, *Forbes*, *Venture Beat*, *The Independent*, *The New York Times*) and among scholars (James Grimmelman, John Grohol, Tal Yarkoni, Zeynep Tufekci, Michelle N. Meyer - see [140], for a detailed collection of responses). The *New York Times*' Vinu Goel surmised that "to Facebook, we are all lab rats" and *The Atlantic*'s Robinson Meyer called the study a "secret mood manipulation experiment" [128, 224]. Responses from scholars were more mixed: a group of ethicists reacted with skepticism to the many critical media reports, arguing that they overplayed the danger of the experiment and warning that the severe attacks could have a chilling effect on research [224]. Several critics noted that the research design and the magnitude of the experiment were poorly represented by the media, while others claimed that a significant breach of research ethics had occurred, with potential legal implications [139, 324]. First author Adam D.I. Kramer responded to the criticism with a Facebook post in which he explained the team's aims and apologised for the distress that the study has caused [185].

The strong reactions provoked by the paper, especially in the media, seem related to the large scale of the study and its widespread characterisation as "a mood-altering experiment" [208]. Furthermore, the 689,003 users whose News Feeds were changed between 11 and 18 January 2012 were not aware of their participation in the experiment and had no way of knowing how exactly their News Feeds were adjusted. In their defense, Kramer and colleagues pointed out that: (1) the content omitted from the News Feed as part of the experiment was still available by going directly to the user's Wall; (2) the percentage of omitted content was very small; (3) the content of the News Feed is generally the product of algorithmic filtering rather than a verbatim reproduction of everything posted by one's contacts; and (4) no content was examined manually, that is, read by a human researcher, but that the classification was determined by LIWC automatically. Some of these aspects were misrepresented in the media reactions to the study, but more basic considerations such as how the study had been institutionally handled by Facebook, Cornell, and PNAS, and whether agreement to the terms of service constituted informed consent to participation in an experiment were also raised in the debate that followed.

## 5.2. THE UNCLEAR ETHICAL TERRAIN OF ONLINE SOCIAL EXPERIMENTS

How can the extremely divergent characterisations of the same event be explained, and what do such conflicting perspectives spell out for the ethics of large-scale online social experiments? In what follows, we will discuss these questions, drawing on multiple examples of similar studies. Researchers at Facebook have conducted other experiments, for instance studying forms of self-censorship by tracking what users type into a com-

ment box without sending it [74]; displaying products that users have claimed through Facebook offers to their friends in order to see whether a buying impulse is activated by peer behaviour [319]; showing users a picture of a friend next to an advertisement without the friend's consent [25]; hiding content from certain users to measure the influence peers exert on information sharing [26]; and offering users an 'I Voted' button at the top of their News Feeds in order to nudge family members and friends to vote and at the same time assess the influence of peer pressure on voting behaviour [42].

While the Facebook emotional contagion study caused the largest controversy, other companies actively conduct very similar experiments. OkCupid, an online dating company, undertook an experiment that consisted of displaying an incorrect matching score to a pair of users in order to assess the effect that an artificially inflated or reduced score would have on user behaviour. A couple that was shown a 90% preferential match was an actual 20% match according to the OkCupid algorithm and an actual 90% match was shown as a 20% score [279]. According to the results, the recommendation was sufficient to inspire bad matches to exchange nearly as many messages as good matches typically do [260], calling the effectiveness of the algorithm into question. Co-founder and president of OkCupid Christian Rudder responded to this criticism by claiming that: "when we tell people they are a good match, they act as if they are [...] even when they should be wrong for each other" [279]. OkCupid also removed text from users' profiles and hid photos for certain experiments in order to gauge the effect that this would have on user behaviour [34]. Similar experiments are conducted by companies such as Google, Yahoo, Amazon, Ebay and Twitter, all of which have access to large volumes of user data and increasingly employ interdisciplinary teams of research scientists that approach problems beyond the scope of traditional computer science. Such teams consist of mathematicians, psychologists, sociologists and ethnographers who analyse data from user transactions, interviews, surveys and ethnographic studies in order to optimise company services [328]. Very often (as in the Facebook case) the results of their research is presented at international conferences or published in academic journals in order to stimulate discourse with the academic community. Frequently multi-authored papers bring together company researchers and scientists at academic institutions, particularly in the United States. Therefore the questions of whether something constitutes industry research or academic research is much harder to answer than it may seem at the onset, with the lines deliberately being blurred by the quasi-academic environment cultivated at major internet companies.

### 5.3. ARGUMENTS FOR AND AGAINST ONLINE SOCIAL EXPERIMENTS

In the debate that followed the publication of the study, different stances were assumed by a range of actors including journalists, user rights advocates, government officials, company representatives, and academics from a variety of fields, a small and nonrepresentative selection of which is presented in the following (see table 1 for a summary). Our sample is based on a list compiled by legal scholar James Grimmelman [140], who collected sources and called for references from social media users in the period after the study had been widely publicised. Grimmelman does not specify exact criteria for the

items on his list, simply referring to them as “major primary sources”, but we believe that it provides a valuable overview of the types of arguments made in favour of and in opposition to the study. Many commentators reacted critically to the research, but some also expressed concerns in relation to how the study had been handled, blaming media hype and misrepresentation of the experiment for some of the negative responses. Our aim is to characterise these reactions through their implicit conceptualisations by identifying a set of recurring arguments provided in defense of the experiment. Our intent is furthermore to categorise and contrast different arguments, and to point out how they relate to the actors who benefit most from what they imply. By categorising actors along with arguments, we show that the discussion around online experiments is strongly shaped by different and at times conflicting epistemological frameworks that implicitly privilege certain viewpoints over others to attain legitimacy.

### 5.3.1. BENEFITS OF ONLINE EXPERIMENTS FOR THE INDIVIDUAL

5 A number of media reports stated that as part of the experiment, the News Feed had been “manipulated” [18, 33, 161, 200, 224], a wording that appeared problematic to some commentators, as the News Feed is generally filtered to represent a selection of status updates curated according to algorithmic criteria [45, 124]. Since the News Feed is algorithmically personalised to foster user engagement in Facebook, it is difficult to judge what kind of modifications qualify as manipulations and which constitute website optimisation. Gillespie [2014] points out that Facebook’s curation of user data in the News Feed is already part of the site’s terms of service and its data use policy. Sandvig [2014] in turn offers a list of examples outside the News Feed in which pieces of personal communication are effectively recontextualised, for example to be used as advertisements. Facebook has stated that out of an average of 1,500 updates, the News Feed algorithm selects approximately 300 items for each user with each update [21]. According to Facebook, in an unfiltered stream of information, people would be missing “something they wanted to see” [21]. Since the selection of items is achieved through constant testing of alternative site designs, content selection is the product of constant experimentation. As platforms such as Facebook are generally subject to some sort of algorithmic filtering, some commentators have argued that we are ultimately faced with “a problem with the ethics of there being an algorithm in the first place” [276].

On the other hand, research shows that most Facebook users have no precise idea about how the News Feed algorithm works, or that there is a filtering process at all [285]. Contrary to intuition, an average Facebook post reaches only 12% of a user’s followers [69]. This curation is assumed to add value, and given the amount of content that is published on Facebook, it reduces clutter. But the filtering criteria cannot be controlled by users (in contrast to, for example, privacy settings), and the precise set of criteria is not transparent. Sandvig [2014] refers to the dangers of a curation that results in a distorted sense of the social context as “corrupt personalization” which he characterises as “the process by which your attention is drawn to interests that are not your own”. He acknowledges that it is difficult to pinpoint inauthentic personal interests, but argues convincingly that a commercialisation of communication through algorithmic curation may conflict with user interests without the subject noticing that this is the case. Sandvig categorically differentiates between tailoring content to a user in her best interest

and deriving a profit from it, and prioritising commercial content over non-commercial content in a non-transparent fashion. He interprets the latter not merely as an ethical issue to be resolved, but also as a waste of the potential of algorithmic curation.

### 5.3.2. INFORMED CONSENT AND ITS MANY INTERPRETATIONS

A second point of contention is whether or not agreeing to the Facebook terms of service constitutes informed consent to an experiment in which the News Feed is manipulated in the described way. This question has narrower legal and broader ethical implications. A clause in the terms of service covers research to improve the site and make it more attractive to users, but experts disagree on whether this covers an experimental design as the one chosen by Facebook [139, 224]. The Facebook study provoked a discussion among legal scholars about the responsibility of institutional review boards (IRBs) that is still ongoing, demonstrating that massive online experiments represent uncharted territory not just from the perspective of internet companies, but also for academic regulatory bodies, who are likely to approach such experiments in markedly different ways. Grimmelmann [139] argues that “informed consent, at a minimum, includes providing a description of the research to participants, disclosing any reasonably foreseeable risks or discomforts, providing a point of contact for questions, and giving participants the ability to opt out with no penalty or loss of benefits to which the subject is otherwise entitled”, which in his view the Facebook study did not do effectively. Taking on a similar perspective, Gray [2014] points out that Facebook could have notified the participants in a follow up email, sharing the results with them and offering them a link to the happy and sad moments that they missed in their News Feed while the experiment was underway. Facebook could also have given participants the option of deleting their data after the research was concluded, which the company did not. Jeffrey Hancock, a co-author of the study, also argued for such a “notify after” approach as a response to criticism. Hancock claimed opt-in procedures to be unrealistic for online experiments due to their ubiquity. Instead, he argued in favour of retroactively informing users after an experiment has taken place, including more information about the study, and contact information for the researchers or an ombudsman [193]. Of course, user data samples based on prior consent may be less attractive to scientists than random samples [39]. But while the risk of influencing results by informing users in advance is acknowledged, legal scholars argue that this cannot be effectively weighed against informed consent, because “if it were, informed consent would never be viable” [141].

Beyond the question of what kind of provisions are covered by the terms of service in this concrete case, informed consent more generally is seen by some experts as being in need of reform. Erika C. Hayden refers to informed consent as “a broken contract” [156] and Mary DeRosa describes it as being “overdue for a wake-up call” [81]. In the context of the reactions to Facebook’s study, DeRosa discusses the difference between what may constitute legal agreement and ethical behaviour, asking: “Would anyone seriously argue that Facebook users expected this kind of manipulation of their News Feed or examination of their data for this purpose? Some consumers would knowingly consent to research like this, but it is unlikely that a single one actually did” (para 6). As DeRosa points out, a key problem is that the expectations of users are violated, rather than that consent with online experiments is necessarily per se rare.

Van de Poel [2011] argues that applying the principle of informed consent to social experiments in technology raises the question of whether it makes sense to ask people to consent to unknown hazards. As accepting to be a part of an experiment with unknown consequences seems to entail accepting all negative consequences emerging from the experiment, it is difficult to see how people could rationally agree to such an approach. However, Van de Poel argues, any social experiment involving ignorance and a lack of mutual understanding is unacceptable. Instead of directly trying to apply the principle of informed consent, it might be better to focus on the underlying moral concern on which consent is based. Instead of blindly accepting an agreement, the emphasis could rest on informing users about the experiment as such and the risks it entails, providing the option to stop participating if desired, and notifying participants once the experiment is stopped.

### 5.3.3. THE UBIQUITY OF ONLINE SOCIAL EXPERIMENTS

Some proponents of the study claim that online experiments should be accepted as a fact of life, since every social media company conducts them and they are without any feasible alternative [309]. Furthermore, some researchers argue that online experiments should not be regulated by the same ethical guidelines that are applied to offline laboratory experiments as they are unique, novel and provide a great opportunity to discover human behaviour at a large scale [39, 350]. However, experiments do not always occur in a traditional laboratory setting. Van de Poel [2009] shows that certain innovations, such as nanotechnology, cannot be developed in a laboratory setting and it is hardly possible to reliably predict risks of such technologies before they are actually employed in society. It may not be feasible to reliably predict the possible hazards to all potential users of a technology, and even when we can, we may not properly express their likelihood in numbers. Van de Poel [2009, 2001] lists conditions for the acceptability of social experiments: (1) the absence of alternatives, (2) the controllability of the experiment, (3) informed consent, (4) the proportionality of hazards and benefits, (5) the approval by democratically legitimised bodies, (6) the possibility for subjects to influence the set-up, carrying out and stop the experiment if needed, (7) the protection of potentially vulnerable subjects, and (8) careful and proportional scaling of the sample size.

Clearly many online intermediaries do not adhere to these principles, mixing different types of considerations: (1) users are rarely informed before or after an experiment is conducted, (2) experiments are approved from within the company, rather than by independent bodies, (3) the subjects cannot influence or stop the experiment, nor give feedback, (4) vulnerable subjects are not protected, (5) experiments are conducted in large scale from the start, (6) the distribution of potential hazards and benefits are not clearly shown, (7) alternatives to the experiments are not considered, and (8) experiments are not subject to the control of participants in the sense that they are able to revoke or modify their participation after the experiment has started. While the ubiquity of such experiments is a result of the pervasiveness of online platforms in which users are able to interact, this hardly makes the experiments ethically less consequential. All actors involved need to jointly discuss and devise criteria for the ethics of online experiments in accordance with existing guidelines (see for example [20]). This by no means excludes users, who also can better weigh risks and benefits when they are adequately

informed. In this vein, arguing for a better understanding of how social media platforms operate, Muench [2014] observes that it is “important for users to be aware of how these sites are designed to engage and reinforce our browsing behavior through evolutionary reward systems”.

#### 5.3.4. DIFFERENT PERCEPTIONS OF RISK IN ONLINE EXPERIMENTS

The authors of the Facebook study claimed that because Facebook did not insert emotional messages into the News Feed, but only hid certain posts for certain users, the experiment did not represent any danger to users. This argument has been opposed on the grounds that if persuasion does not happen voluntarily and if the persuader does not reveal her intentions before the persuading act takes place, this is to be considered manipulative [297, 302], making manipulation as much an issue of intent as much as an issue of effect. Others argue that involuntary persuasion is acceptable only if there is a very significant benefit for society that would outweigh possible harms [38]. In the case of the Facebook study, it is difficult to adequately judge the benefits of the research at this point, while the harm, if only in terms of public perception, has become quite obvious. Data scientist Duncan Watts optimistically argues in *The Guardian* that online social experiments will usher in “a golden age for research” [350], but this depends on each actor’s perspective. Mary L. Gray [2014] draws a comparison to early nuclear research and experiments on human subjects, and sees data science as undergoing a learning process with regards to research ethics. In reaction to Kramer’s response to the criticism, published on his personal Facebook page, individual Facebook users responded with personal accounts of emotional hardship and depression, expressing concern that Facebook would experiment on the content of the News Feed in ways that could adversely affect them. The question of risk beyond individual users seems impossible to answer without precedence, but the lack of transparency towards participants is likely to weigh more strongly in the eyes of many users than the small size of the effect reported in the study - and the details of how the filtering was conducted. Furthermore, as Kramer and colleagues point out, the impact of systematically seeking to influence users may still be strong, even if it is restricted to a small group. In a 61 million user experiment in 2010, Facebook users were shown messages at the top of their News Feeds that encouraged them to vote, pointed to nearby polling places, offered a place to click “I Voted” and displayed images of select friends who had already voted [42]. The results suggest that the Facebook social message increased turnout by close to 340,000 votes. It has consequently been argued that if Facebook can persuade users to vote, it can also persuade them to vote for a certain candidate, a kind of influence which, while hypothetical, does present obvious risks [375].

#### 5.3.5. BENEFITS OF ONLINE EXPERIMENTATION FOR THE SOCIETY

A popular argument among proponents of online social experiments resides in their potential benefits to society, and associated with these, the danger that negative responses could have a chilling effect on collaborations between industry and academics [39, 224, 350, 366]. Michelle N. Meyer [2014] makes this argument in two parts, stating first that “rigorous science helps to generate information that we need to understand our world, how it affects us and how our activities affect others”, and secondly that “permit-

ting Facebook and other companies to mine our data and study our behavior for personal profit, but penalizing it for making its data available for others to see and to learn from makes no one better off”. Similar arguments are made by Watts [2014], and also by Tarkoni [2014], who contends:

“Consider: by far the most likely outcome of the backlash Facebook is currently experiencing is that, in future, its leadership will be less likely to allow its data scientists to publish their findings in the scientific literature[...] The fact that Facebook is willing to allow its data science team to spend at least some of its time publishing basic scientific research that draws on Facebook’s unparalleled resources is something to be commended, not criticized.”

What justifies the risks, if potential, that are incurred by large-scale online social experiments? Watts draws an analogy between the rise of empiricism during the Enlightenment and the current circumstances, arguing that “the arrival of new ways to understand the world can be unsettling”. But this analogy is made at least latently problematic by the commercial interests that are at play - the opportunities of learning anything about basic human behaviour are no more pertinent than the opportunities to influence behaviour, for whatever purpose. Muench [2014] compares online social experiments to Skinnerian operant conditioning, in which strategic choices, such as exposing subjects to stimuli in randomised intervals, lead to greater engagement. To make good on the claim of societal benefit, a clearer case needs to be made for the positive impact of online social experiments, a case that is able to transcend the aim of increasing user engagement.

5

### 5.3.6. THE UNAVOIDABILITY OF ONLINE EXPERIMENTS

Advocates of online social experiments, such as OkCupid’s CEO Christian Rudder, argue that such experiments are unavoidable, because all aspects of the design of digital platforms are shaped by constant experimentation in order to make improvements:

“OkCupid doesn’t really know what it’s doing. Neither does any other website. It’s not like people have been building these things for very long, or you can go look up a blueprint or something. Most ideas are bad. Even good ideas could be better. Experiments are how you sort all this out.” [279].

He continues to argue that experiments are needed to make sure that the current algorithm works better than a random one, and that there is no alternative to such an incremental approach to optimally address user preferences. He also believes that while experiments presently cause controversies, they will be fully accepted in the future. Critics contend that the potential to innovate via experimentation must still be weighed with possible drawbacks, rather than being accepted as being without an alternative. For instance, Howell [2014] responds to Rudder arguing that he “is clearly acting wrongly”, and for (at least) two reasons: 1) He is being dishonest by providing something other than what he says he will provide. Rudder thus provides a system that performs bad matches to see how people will react, instead of their claim “our matching algorithm helps you find the right people<sup>1</sup>”. 2) he subjects his (users) to potential harm that they have actively sought to avoid. Howell [2014] further argues that the defense of the company is disingenuous: “either OkCupid believes its sales pitch or it doesn’t. If it doesn’t, we already have a moral issue. If it does, then they are doing what they believe will be harmful

<sup>1</sup>See <http://www.okcupid.com> for the claim.

to their customers". Grimmelmann [2014] shares this view when proposing that, unless risks are minimal or nonexistent, researchers cannot decide that an experiment is worth a particular risk. That decision should instead be made by users.

Table 1 summarises our observations on the arguments made by the proponents and critics of the Facebook study, and similar online experiments.

## 5.4. DISCUSSION

We have aimed to show that the ethical issues raised by social experiments can be described on multiple discursive levels, depending on the roles that the discussants assume. We have shown that the problem is complex and involves interests reflected in different arguments, such as the individual and social benefits of online experiments, their ubiquity and relevance, the fact that consent is provided and that users are not exposed to any significant risks. We have shown that some of these values themselves are dependent on specific frames of reference (e.g., the attainment of status in science) and that further debate is needed to balance their relation to one another. Perhaps our central observation is that the asymmetrical relationship between data scientists and users of social media platforms is what underpins these conflicting frames of reference. Furthermore, as long as there is no consensus regarding the ethics of online experiments that transcends a single stakeholder group, such conflicts are likely to arise again in the future, rather than abate. In this paper, we have used the Facebook experiment as a use case to discuss a range of arguments provided by different stakeholders to illustrate this conflict.

While the study has provoked strong reactions, it is worth to again point to similar research, both at Facebook and elsewhere, to clarify that this is a broader issue, rather than a singular case. In a 2012 study on information diffusion, Facebook researchers randomly blocked some status updates from the News Feeds of a pool of some 250 million users, many more than in the emotion contagion experiment. Google provides a set of tools to conduct A/B tests for website optimisation, as does Amazon. Beyond A/B testing to improve the quality of search results, issues become yet more complicated when experiments around information exposure are conducted with social improvement in mind, and without explicit consent. In research conducted at Microsoft, researchers Yom-Tov, Dumais, & Guo [2013] changed search engine results in order to promote more balanced civil discourse. In the study, the authors modified results that were displayed when users entered specific political search queries, so that subjects entering the query *obamacare* would be exposed both to liberal and conservative sources, rather than just to content biased into one ideological direction. While the researchers arguably had the best intentions, they did not notify users that their search results were being modified, neither during the experiment nor afterwards. This raises complex questions regarding the ethics of manipulation with the aim of affording social improvement. Some have claimed that when persuasion is conducted for a higher ethical goal, this can be acceptable [38], while others disagree [297, 302]. In the light of the discrepancy between the ethical standards of academic research on human subjects and the entirely different requirements of building and optimising social media platforms and search engines, it is tempting and simplistic to single out any particular company for filtering content algorithmically. New collaborative models of joint corporate and academic research are

<b>Argument theme</b>	<b>Pro experiment</b>	<b>Contra experiment</b>
Benefits of online experiments for the individual	<ol style="list-style-type: none"> <li>1. Filtering reduces clutter</li> <li>2. Users want filtered, rather than unfiltered content</li> </ol>	<ol style="list-style-type: none"> <li>1. Users are not aware of filtering</li> <li>2. Filtering cannot be controlled</li> <li>3. Filtering mechanisms are not transparent</li> </ol>
Informed consent and its many interpretations	<ol style="list-style-type: none"> <li>1. Accepting terms of service is a form of consent</li> <li>2. Opt-in is annoying to users</li> <li>3. Opt-in influences user behaviour</li> </ol>	<ol style="list-style-type: none"> <li>1. Possibility of biased user behaviour does not counter informed consent</li> <li>2. Users could be informed post-experiment</li> <li>3. Consenting to unknown hazards is problematic</li> </ol>
The ubiquity of online social experiments	<ol style="list-style-type: none"> <li>1. Experiments are essential to platform improvement</li> <li>2. Differ from offline experiments by being unique and novel</li> <li>3. Provide opportunities to study human behaviour at scale</li> </ol>	<ol style="list-style-type: none"> <li>1. Same principles that govern offline experiments can be applied</li> <li>2. Experiments should not be conducted at large scale when there is no need</li> <li>3. Alternatives should be considered</li> <li>4. Users should be able to influence or stop the experiments and provide feedback</li> </ol>
Different perceptions of risk in online experiments	<ol style="list-style-type: none"> <li>1. Withholding information does not cause danger</li> <li>2. In the long term, benefits will outweigh risks</li> </ol>	<ol style="list-style-type: none"> <li>1. If participation is not voluntary, it is manipulative</li> <li>2. Persuasion is likely to benefit the persuader at least as much as the persuaded</li> </ol>
Benefits of online experimentation for the society	<ol style="list-style-type: none"> <li>1. Online experiments create new opportunities for science and society</li> <li>2. Constant scrutiny will have a chilling effect on collaboration between industry and academia</li> </ol>	<ol style="list-style-type: none"> <li>1. Exact benefits are unclear</li> <li>2. We learn less about human interaction than about media effects</li> <li>3. It is not sufficient to equate scientific benefit with social benefit</li> </ol>
The unavoidability of online experiments	<ol style="list-style-type: none"> <li>1. Online platforms cannot be improved without experimentation</li> <li>2. Incremental improvement is the only way to succeed</li> </ol>	<ol style="list-style-type: none"> <li>1. Potential risks also need consideration</li> <li>2. Judging risks to be minimal without having considered them is premature</li> </ol>

Table 5.1: Arguments for and against online social experiments surrounding the Facebook emotional contagion study

considerably blurring the boundaries between basic and industry research, and complicating the picture of disinterested academia and result-driven commercial research.

The public outcry in reaction to the Facebook study underlines that there is a growing expectation towards more transparency regarding how content is filtered and presented, beyond assuming a 'take it or leave it' -style attitude. A company may have the interests of its users in mind, whether this goal is usability, more relevant search results, happier status updates, or a better match in dating platforms. However, users have to be able to assess these intentions for themselves, and evaluate the balance between their personal benefits and the interests of the company. There is a pronounced fear among publicly-funded academics that Facebook and other social media companies might limit the already fairly sparse access to their data, as they clearly see benefits in publishing studies based on unprecedented amounts of data - not solely for science, but also for their own careers. Competition for cutting-edge research results is neither unique to social media data nor surprising, but it spells out a potential conflict of interest between users whose sense of freedom and privacy is in potential conflict with scientists' interest in advancing a nascent field vying for scholarly acceptance through high-profile publications. To users, it remains largely unclear what exactly the benefits of such research may be. The argument made by Meyer, that "rigorous science helps to generate information that we need to better understand our world" (our emphasis), is qualified by the highly media-specific nature of such research - we learn much more about how people react to each other on Facebook than about human interaction in any broader, more universal sense.

After the controversy had erupted, the editor of the publication, Susan Fiske, noted the complexity of the situation, pointing out that the Institutional Review Board of the authors' institutions had approved the research, and arguing that Facebook could not be held to the same standards as academic institutions. Kramer and colleagues clearly saw their experiment in line with Facebook's continued efforts to optimise the News Feed, yet as we have pointed out, the arguments made in defense of this and similar experiments are strongly coloured by the interests of different parties, with users relatively far removed from the benefits in favour of which the proponents argue. Data science must show more convincingly that it balances the interests of scientists, companies and users to deliver on its many promises. Laboratories, regardless of their size, are governed by rules ensuring that the research conducted under their oversight is not just legal, but also ethical. Legalistic attempts to seek to cover behind the terms of service have failed to achieve this type of broad societal acceptance for what undoubtedly constitutes a new approach to science. While some researchers argue that online social experiments should not be subjected to the same ethical guidelines that are used for offline social experiments, we find the 'newness' of such experiments to lie in their potential scale, rather than in their ethics. The point is not to wring our hands about hypothetical potentials for abuse, but to carefully examine cases such as the Facebook study and ask why the reference points of users and data scientists are as different as they apparently are, and whether these differences can be reconciled in the future. Benefits for science should be balanced with possible hazards that may be caused by experiments, rather than precluding that such benefits outweigh the gains. Transparency towards users is paramount, as is seeking articulated consent for participation.



# 6

## CONCLUSION

*The ties which hold men together in action are numerous, though and subtle. But they are invisible and intangible. We have the physical tools as never before. The thoughts and aspirations congruous with them are not communicated, and hence are not common.*

*Without such communication the public will remain shadowy and formless, seeking spasmodically for itself, but seizing and holding its shadow rather than its substance. Till the great society is converted into a great community, the public will remain in eclipse.*

*Communication can alone create a great community.*

John Dewey, philosopher, 1927

Pariser [2011] and others have pointed us to the danger of online personalization and social filtering for society. Pariser's book has caused lengthy discussions online and offline. This thesis contributes to the public discussion and academic debate by using the Value Sensitive Design methodology in order to analyze the filter bubble phenomenon and the value of viewpoint diversity. It performed the following conceptual, empirical and technical investigations; 1) At the conceptual level, it extends the gatekeeping theory of communication studies by algorithmic gatekeeping. It also clears any conceptual muddles by introducing different dimensions of viewpoint diversity and by providing an overview of the metrics used in the literature of viewpoint diversity analysis and media ethics 2) At the empirical level, it provides a framework to analyze viewpoint diversity in Twitter for different political cultures. It shows how minority access is missing from the typical dimensions of viewpoint diversity studied by computer scientists and the impact it has on viewpoint diversity analysis. Further, it provides the results of a case study on minorities that is performed for Turkish and Dutch Twitter users and shows that minorities cannot reach a large percentage of Turkish Twitter users. It also shows that algorithms and personalization are only part of the problem, as operators of the online services and other processes may cause biases 3) At the technical level, it studies technical artifacts that are designed to combat filter bubble and shows how different understandings of democracy by the developers lead to fundamentally different designs.

In the next section, we will revisit the research questions and describe how this thesis answered them. We will discuss the limitations of the study. Finally, we will discuss future work.

## 6.1. ANSWERS TO THE RESEARCH QUESTIONS AND REVISITING THE OBJECTIVES OF THE STUDY

**RQ1:** What is filter bubble and which factors lead to it?

This thesis shows that filter bubble is the phenomenon where online platforms supply biased information due to 1) algorithms 2) human operators 3) selective exposure (users' own biases). Chapter 2 studied the algorithmic and human processes and factors within online platforms that cause filter bubbles. These included factors that came in effect after personalization technologies became prevalent, such as user interaction history, location, novelty or user preferences. Other factors such as technical limitations, information type and information age are factors that might diminish viewpoint diversity and introduce bias for non-personalized systems as well. Factors that included traditional media such as organizational factors, personal judgments, government and advertiser pressures still affect online platforms as they have operators that can prioritize/deprioritize and delete content. Chapter 3 on the other hand analyzed selective exposure and group polarization by studying user's incoming and outgoing information and found out that, depending on the studied political culture, minorities can become invisible. Chapter 4 showed that filter bubbles violate different norms required by different democracy theories. These are 1) the loss of control and autonomy over opaque filters that are prominent in modern online intermediaries and being unaware of alternative information; 2) being exposed to same viewpoint and deliberating with like-minded individuals, whether it is forced or chosen; 3) the invisibility of minorities and politically disadvantaged due to an emphasis on consensus and mainstream opinions; 4) not being aware of the oppressors, not being able to provide key information, and the lack of public channels where the performance of the decision makers can be monitored and contested.

**RQ2:** Can we conceptualize and operationalize viewpoint diversity?

Yes we can. However, there is no single metric to measure bubbles, as they violate different norms of democracy. Chapter 3 quantified viewpoint diversity by combining metrics used in different studies and by introducing new ones. Existing metrics included source diversity, seed user interaction and input-output correlation. New metrics are minority access and output diversity. The study indicated the importance of minority access by showing that minorities could not reach the majority of the users in Turkish Twitter sphere. These metrics (and possibly new ones, including topic diversity) should be repeated and tested in different (political) cultures to get a better understanding of this phenomenon.

**RQ3:** Which issues does filter bubble cause for different models of democracy and what can design do to solve them?

Chapter 4 showed that the reason why filter bubbles are an issue for democracy depends on the used model of democracy and different norms are required by different models. Liberal democracy criticizes filter bubbles, because they make the user unaware of the options, take the control from him, restrain the user and his individual liberties are curtailed. Deliberative democrats are more concerned about the diminishing of epistemic quality of information, losing respect towards one each other and the suffering of communicating, mutual understanding and sensemaking. Contestatory democrats are worried because filter bubbles diminish one's ability to contest and diminish one's awareness of the oppressors, as channels of publicity and mechanisms that show the performance of the relevant parties are hidden. Agonists are worried because the minorities and the disadvantaged could become ever more marginalized and their ability to reach a larger public becomes only a theoretical possibility.

Chapter 4 also studied different tools designed by computer scientists to combat filter bubbles. It showed that only two of the introduced democracy models (liberal and deliberative models) influenced the developed tools. The chapter concluded that the designers' exposure to online environments by itself does not provide them with a variety of democracy conceptions nor with a sensitivity to realize that 'democracy' is an essentially contested concept. We recommended that the designers should be exposed to diverse conceptions of the value democracy and discuss their goals with the general public before and after the tools become operational.

**RQ4:** What are the ethical issues associated with those solutions?

Some of the tools or methods developed by online platforms to combat filter bubbles are experiments [368]. Some authors argue that experiments performed by the platforms themselves may cause bubbles [220]. Chapter 5 discussed arguments used in discussing the ethics of online social experiments. Using the controversy around Facebook's emotional contagion study [185], it analyzes the arguments of two parties (data science advocates and data science critics) through a qualitative and non-representative discourse analysis. These included many interpretations of informed consent, benefits of online experiments for the individual and the society, the ubiquity of online social experiments, different perceptions of risk, and the unavoidability of such experiments.

**RQ:** How can we design for viewpoint diversity, so that filter bubbles are avoided?

Answers to the subquestions show us that, before designing for viewpoint diversity, we should 1) understand which factors can cause bias and diminish viewpoint diversity 2) conceptualize what filter bubble is by studying different theories, particularly different democracy models and norms 3) understand the extent of filter bubble's effect by conducting empirical studies using different metrics for different political cultures 4) implement different norms to improve viewpoint diversity by design 5) Repeat these steps and make sure that the goal of the design and studies are communicated with the public.

Almost all empirical research that analyzes viewpoint diversity (including Facebook's most recent study [24]) use deliberative model to study if two major ideologies in the US are exchanging information with each other. However, our empirical study in Chapter 3 showed that minorities are invisible in Turkish Twittersphere. This thesis thus showed that diversity can be measured differently depending on the chosen democracy theory

and can produce different results for different political cultures. While online intermediaries have an important role in measuring and understanding viewpoint diversity, finding out whether bubbles exist is no easy task and it cannot be determined by one single study. There is a broader need for scientists, activists, individuals and social critics to study the online intermediaries on a long term, as the algorithms used by the platforms and the behavior of the users may change. Informed by the findings, designers can produce different solutions.

This thesis also showed that designing for viewpoint diversity is no straightforward matter. This is due to different norms required by different models of democracy. While almost all of them value viewpoint diversity, how it conceptualized can be quite different. Most tools that combat filter bubbles either implement the deliberative or the liberal models. However, these models have their weaknesses and are being criticized by other political theorists. For this reason, before and after the tools are developed, the designers must discuss their goals and the used norms with the public explicitly.

## 6.2. LIMITATIONS OF THE STUDY

In this thesis, we have applied Value Sensitive Design (VSD) methodology and focused on viewpoint diversity, freedom from bias and democracy. Implications for other values we have identified in Chapter 1 (privacy and identity) are not studied. A known challenge of VSD is that values often conflict with each other. Value conflicts might arise between moral values (privacy vs. autonomy, or autonomy vs. viewpoint diversity) or between moral values and economical values such as efficiency, productivity and profitability [215]. Resolving these value conflicts or value tradeoffs is no trivial matter [341]. One can argue that the pressures of the marketplace will make it very costly for a company to implement viewpoint diversity in its design. Efforts in implementing VSD in the industry in some instances were not successful. For instance Manders-Huits and Zimmer [2012] argue that they could not convince engineers and managers to include privacy in the design of a software system, because different parties had different conceptions of privacy and some did not see privacy as an issue. Further, concerns over efficiency and costs outweighed the benefits and translating certain ethical values into design requirements was not successful [215]. Advertising, profit or other factors that play a role in commercial environments could affect the implementation of these values in design. We only focused on moral values and did not focus on other values, such as economic, cultural or aesthetic ones. Therefore this remains a limitation.

Chapter 3 suffers from limitations that any big data analysis does. First, big data analytics observes the traces left in the media and can rarely observe the driving factors that cause these behaviors [147]. For instance users will make different uses of Twitter. Some might use it as their primary news source, therefore following mainstream items, while others will use it to be informed of the opposing political view or to find items missing in the traditional media. Therefore, we do not know why some users only follow sources from a specific political category. More qualitative studies are needed. Second, there is also a difference between endorsement retweets (created by pushing the retweet button) and informal tweets (where users include most of the same text often prefixed by 'RT' or similar but also add their own comments before or after the tweet). These two actions measure a different interaction. Informal retweets and replies could also express dis-

agreement and show us deliberation. In order to make a distinction between these two types of tweets, we need semantic analysis. We are not aware of the availability of such tools for the Turkish language, therefore we were unable to perform such an analysis. While information diversity is important not only for deliberative models of democracy, it would be very useful to study deliberation on Twitter by using such tools in the future. Third, information intended to deceive can spread through social media in the same way as valid information. This raises questions on how to detect different types of deception (e.g., manipulating information, changing context, or outright fabrication) in different social channels and formats [192]. In our case, users could retweet or reply with bad intentions, such as trolling. For retweets, we only measured users' retweets to original tweets created by seed users. We assume that, those powerful political actors would not take part in trolling. Users can retweet a seed user's tweet randomly or for trolling purposes. The same issue can manifest itself in replies. Since we did not perform a semantic analysis, this remains a limitation. Fourth, while the results give us an idea on the political landscape of the studied countries, Twitter does not represent 'all people'. As boyd and Crawford [2011] have stated, "many journalists and researchers refer to 'people' and 'Twitter users' as synonymous (...) Some users have multiple accounts. Some accounts are used by multiple people. Some people never establish an account, and simply access Twitter via the web". Therefore we cannot conclude that our sample represent the real population of the studied countries.

### 6.3. FUTURE RESEARCH

This thesis showed that the design of filtering algorithms is an interdisciplinary problem. It has several technical challenges and it requires quantitative analysis. However, as we have shown in this thesis, such quantitative analysis must be informed by conceptual investigations. Otherwise it has the risk of missing important human behavior and ethical values that people might consider important. While this thesis has contributed to the literature most importantly at the conceptual and empirical level, much work remain to be done. This section points out the future steps for different disciplines.

#### 6.3.1. VIEWPOINT DIVERSITY RESEARCH

The most important continuation of this thesis would be to extend viewpoint diversity analysis. We need to know whether viewpoint diversity differs per culture and whether political culture makes a difference in changing attitudes towards a more balanced information diet. Data analysis, accompanied by surveys and experiments could show interesting results. Such experiments can be performed by the online platforms themselves or their benchmarks could be open to scrutiny for the researchers. The framework we have presented in Chapter 3 and benchmarks we have presented in Chapter 4 can be used as a starting point to conduct an ongoing diversity analysis. However, as Chapter 5 and the recent Facebook experiment controversy show, it is vital to communicate any experiment, its goals and its results with the users. A/B testing and registering the user clicks and other actions seems to be the industry standard for capturing user behavior. In order to avoid controversies, design features could be discussed with the users before it is implemented and released. Value Sensitive Design [116] could be very useful here.

Empirically, we only measured group polarization in Twitter. Repeating these studies in Facebook or other platforms where algorithmic filtering is taking place could add to our insights. Facebook itself performed two filter bubble studies to understand whether online polarization occurs. The first one only measured novelty and diversity of URLs, not viewpoints [26]. The second one measured viewpoint diversity by focusing on 10 million Facebook users who have labeled themselves politically [24]. The company studied the links shared by those individuals by assigning each shared article a score based on the political stance of the people who shared it. Facebook concluded that while the filters decrease viewpoint diversity, the role of users' networks and their own bias is larger. The study has been criticized due to weaknesses in its methodology [154, 173, 257], selection bias [154], its reflection of only a limited period [199, 257], problems with ideological tagging [257], the effect of algorithm in user choices and false claim of neutrality, [173, 325], irreproducibility [199, 257]<sup>1</sup> and its incorrect conclusions [257, 283, 325]. These criticisms show that there is a broader need for scientists, knowledgeable individuals and social critics to study systems like Facebook.

Another issue with Facebook research and others that study filter bubble is the lack of diversity in democracy conceptions. As Chapter 4 showed, democracy is a contested concept and different models imply different norms which might lead to different metrics and designs. Most studies focus on norms required by liberal and deliberative notions of democracy, while there are many other models and other norms. For instance agonists criticize the consensus goal of deliberative democracy and argue that radical voices, including those of the minorities, must be heard, not just formally, but also effectively. Future research should thus study different norms of different democracy models, not just the popular ones. Open questions include whether the information intermediary prioritizes some voices over others, and whether certain types of subjects are highlighted by the algorithm in a way that systematically undermines certain topics, groups or individuals.

Chapter 3 showed that minorities cannot reach the majority of the studied users in Turkish Twittersphere. Many political theorists argue that the minority improves public judgment by activating a validation process in the otherwise conformist majority and respecting the minority discourages deluded consensus [212, 337]. One could argue that the minority posts could get prioritized in online platforms, so that they can effectively reach the population, rather than the *possibility* to reach. Research shows that even though an alternative or radical voice reaches a group online, it is not successful in opening up a dialogue [359]. In Witschge's study, participants were unanimous in trying to find ways to exclude the minority voices. They did this not by addressing the content of the message, but by "eliminating" it. The participants or the content of the post is not acknowledged. Witschge [2008] concludes that while the discussion platform initially allowed for the voice to be included, the participants were not receptive. Recent incidents also show that when minority tweets reach a large audience through retweets, often harassment follows [103]. Future research should also look for ways to make radical voices heard effectively and their arguments to be discussed. Mackie [2006] provides empirical evidence where minority view takes longer to affect the deliberating individ-

<sup>1</sup>Also see <https://developers.facebook.com/docs/apps/upgrading> for restrictions of data collection using Facebook API

uals. Informed by such theories and evidence on attitude change, online platforms can experiment with different methods on providing minority viewpoints to the larger public.

### 6.3.2. TRANSPARENCY

Google amounts to 64.4% of search traffic in US and 92% in Europe [277, 305, 355]. Over 1.4 billion users are active on Facebook [357]. These platforms “have taken on the role of quasi-governmental bodies today, regulating what their users can and cannot do” [50]. With its “Ads Preferences Manager” Google displays the attributes it has assigned to a user for ad prediction and allows the user to decline any incorrect or undesired attribute [133]. However, it provides no such option for search results. Facebook allows the user to change the News Feed based on “Top Stories” and “Most Recent”, but the user cannot know why a particular item appears in one feed, and which items were selected for which reason. Recent research shows that users often do not know that their results are being personalized, that they do not receive all the updates from their family and friends, and, as such, not all users may understand why their web experience is different from that of their friends [100, 271]. Even if users are aware of personalization, they have very limited control to make corrections. Recently, U.S. Federal Trade Commission (FTC)’s chief technologist argued that consumers interact with algorithms of online intermediaries on a daily basis, yet we have very little insights on how they operate and the incentives behind them [246]. FTC’s Bureau of Consumer Protection established an office dedicated to better understand how these algorithms work and which assumptions are taken in their design.

Transparency however, is a quite a challenging value to implement. Writing on search engine bias, Introna and Nissenbaum [2000] argued that the marketplace alone will not ensure transparency and that we should “demand full and truthful disclosure of the underlying algorithms that govern indexing, searching, and prioritizing, stated in a way meaningful to most Web users”. However, this might not provide the transparency we want, as it will be very difficult for an average user to comprehend the algorithm and determining the output and consequences from source code will be quite difficult, even for an expert. As we have argued in Chapter 2, algorithms are not the only factors that determine the output of an online service, whether it is search ranking of a search engine or the news feed of a social networking platform. All major services employ human operators that might remove, deprioritize content, URL or the entire domain name. The decisions are often political and the criteria in which they make their decisions are often unknown. Chapter 2 showed us that it is not only the workings of the algorithm that must be explained, but also the decisions of those human operators. The system might aim to increase the diversity of a user’s newsfeed or search results, or try to narrow it down for a more exact result. This can be the result of an algorithm or manual human operators. In both cases, the decisions must be explained.

Security researcher and philosopher Pieters argues that explanations, should (1) aim for the right goal (why or how) and (2) carry the right amount of information, in order to provide informed consent to the user, and thereby keep (human) responsibilities clear [266]. Pieters [2010] suggests that, in case the users confide in the system (has no choice), the system should go for justification and explain *why* certain decisions are

made in the algorithm and why things work the way they are. If the users trust in the system (while having a choice, they chose to trust this system by making themselves vulnerable), the system should go for transparency. Pieters argues that as long as the users have confidence in the decisions of the system, they may not be interested in how it works. Therefore, the explanations of expert systems are mainly explanations-for-confidence. Only when the users suspect something, transparency will be required by means of explanations-for-trust.

Applying these principles to the area of diversity, we can argue that the information intermediaries must first provide explanations-for-confidence. This would include listing their benchmarks, that is explaining the goal, purpose, intent and editorial goals of the filtering algorithm [84] and the decisions made by human operators. For instance, the system can show what it is optimizing for, along with user preferences. It can show why certain search results are personalized for certain keywords (“you have searched for this information before”) or why the news feed is personalized (“we are showing you 150 out of 1050 possible items to decrease information overload”), or why the news feed contains certain news items from the other end of political spectrum (“to increase diversity”). In a social networking environment, it can also show how many people one’s post has reached. This way, the user can have a rough idea about the items he is *not* seeing, and items that he has posted but are not seen by others. It would allow the user to “to compare and contrast” [84]. Further, the platform can also provide a list of modifications performed by human operators to make the process more accountable.

Explanations for algorithms are being used by some companies in the industry and are being developed in the academia. For instance, El-Arini et al. [2012] have developed a transparent user-interpretable recommendation system where users are put in clusters and were informed that they were in this cluster by displaying badges such as “vegetarian” or “liberal”. Users have the option to correct this categorization [95]. Amazon’s recommendation system aims to achieve this by displaying the recommendation reason (“because you purchased”). If the recommendation is questionable, a user is allowed to correct the system by selecting a “because you purchased” item and indicating “don’t use for recommendations”. Loepp et al [2015] have implemented a faceted filtering method, which they call *blended recommending*, in which users can interact with a set of filter facets representing criteria that can be used as an input for collaborative and content-based recommenders. Users can not only select filter criteria, but they can also provide different weights to influence the recommendations.

As Chapter 2 of this thesis shows, there is no such thing as an objective filter. Therefore, the benchmarks of online intermediaries must be open to scrutiny, if there are suspicions about the published benchmarks. This would entail some sort of algorithmic auditing [284]. This can for instance be achieved by tools to do some black-box testing, to see if the announced benchmarks are actually being used, and if there are some unannounced benchmarks. For instance the tool Booble [365] repeats a user supplied query for generated and aged profiles and then shows the user which results he received and the other profiles did not, and vice versa. Such specialized tools are being developed in the academia and in the industry. However, those tools are also being banned by the online services, often providing reasons such as “unauthorized access” and “attack” [186, 372]. Such tools should thus be added to a whitelist by the online intermediaries

and they should be able to test the algorithms. Future research should devise ways to audit these algorithms by testing the benchmarks in a repeatable and verifiable fashion.



# ACKNOWLEDGEMENTS

Pursuing a PhD, at least in my experience, while being a valuable process, is ultimately a lonely path with unknown horizons. Yet, without the help and support of some people, this path would never be completed. This section is created to mention some of them, while noting that the list is in reality much longer.

First, I would like to thank my promoters Ibo van de Poel and Jeroen van den Hoven. Ibo: thank you for reading the text I have sent often last minute and yet coming up with useful feedback in such a short time. Without your constructive criticism and to the point feedback this thesis would never have the current quality. Jeroen: thank you for giving me the opportunity to carry out this PhD project in Values, Technology and Innovation department and providing important literature suggestions. Thanks to Martijn Warnier for correcting spelling errors for some of the chapters. I also would like to thank the rest of my thesis committee, Christian Sandvig, Mutlu Binark, Richard Rogers, Alan Hanjalic and Marijn Janssen for reading my thesis and accepting to join the committee.

I would like to thank my current employer SIG and especially Joost Visser for offering the flexibility during the final few months of writing this thesis and encouraging me to get my PhD as soon as possible. I also would like to mention all my colleagues who showed special interest in my PhD project: Xander, Barbara, Haiyun, Werner, and all the interns: thank you.

I further would like to thank researchers with whom I had the chance to work during my thesis: Job Timmermans, Qi Gao, Geert-Jan Houben and Cornelius Puschmann. I also would like to thank the following scholars for exchanging ideas during my research: Pascal Jürgens, Seda Gürses, Joris van Hoboken, Frederik Borgesius, Wolter Pieters, Batya Friedman, Ammar Halabi, Kari Karppinen, Johnny Hartz Søraker and Aisling O’Kane. I also would like to thank Helen Nissenbaum, Mutlu Binark and Miriam Rasch for inviting me to present my work in their home institutions.

I would like to express my gratitude to Humboldt Institute for Society & Technology, who hosted me for two months in Berlin as a research fellow. I especially would like to thank Simon and Jana for organizing everything so perfectly and Uta, Lies, Antonio, Kaja and Kirsten for their occasional chats.

I would like to thank Özgür Gültekin for designing the cover of this thesis. Special thanks to Victoria Griesdoorn for translating the summary into Dutch.

My friends from the faculty TPM Sertaç Oruç, Yakup Koç and Zülküf Genç deserve special mention for all their support and friendship that they have given me during my PhD. The topic of our chats included (among others) the life and unknown future of a PhD student, (sad) news from our “lonely and beautiful country”, and Turkish football to which we unfortunately are all addicted. Sertaç: thank you for being my paranymph, all the road-trips in US with Çağla and Sercan, game nights and indoor and outdoor football adventures. Yakup: thank you for the road-trips in the US with Aylin, attending film festivals, endless chats about house renovation and all the other fun stuff. Zülküf: thank

you for coorganizing the weekly football tournament and all the interesting experiences including Marktplaats operations, shirt shopping saga, and the coffee ordering etiquette in Starbucks. Next to the fun moments, thank you Sertaç for reviewing my empirical work and thank you Yakup for helping me with the formulas.

Thanks to my officemates Behnam, Filippo and Neelke. Behnam Reis: thanks for the fun times at the conferences and road trips we have survived together and for your hospitality in Boston and in Delft. Also many thanks for translating the propositions of this thesis into Dutch. Kheili mamnoon! Filippo: Thank you for the chats about football (Torino and Galatasaray in particular) and our discussions about films. Grazie! Neelke: Sorry for my guests who always messed up with your chair. Wasn't me! I also would like to thank fellow PhD's of the philosophy section for creating a friendly working environment. Hendrik and Thomas: thank you for your work in summer school and occasional discussions. Thanks to Henneke for her immense help in organizing the summer school. Thanks to fellow researchers Chris Detweiler, Alina Huldgtren and Maaïke Harbers of EWI. Chris & Alina: I had a great time in Paris and Lisbon with you guys and Monsieur Fredrik. Chris: it was awesome planning a summer school and a course from scratch last minute with you. Merci beaucoup.

Many others in the Netherlands and abroad have brought color to rather black & white days. First, I have to specially mention Sinem and Orkun, for their hospitality and for being there during difficult times. I also would like to thank Huseyin Şen in particular: thanks to his suggestion I applied for the research grant (literally 1 minute before the deadline) which eventually led to the realization of this thesis. Others, including Aylin, Çağla (and Şiraz), Sophia & Rouzbeh, Koray & Deniz, Caner & Nalan, Reyhan, Cafer, Özge, Giray & Pervin, Ibrahim & Sevinç, Vedran, Burak, Kenan, Sezi, Luca&Barbara,: Thank you for the chats, food, holidays, drinks, concerts and many other fun memories that we had together. Thanks to my football buddies Can, Emre and Mustafa for the weekly games. My friends that live abroad: Arno, Özgür & Emanuelle, Irem & Konuralp, Hakan & Latife, Sadi, Yermiz, Seda, Murat & Nildan & Ekin. It was great to see you and talk to you during the writing of this thesis. Thank you for your hospitality and kindness.

Thanks to my parents who were there supporting me as always. Special thanks to my dad who encouraged me to pursue a PhD. I love you both. Further, I would like to thank my other family members in Europe and in Turkey who supported me during the writing process: Nihal & Özgün & Can & Hüseyin, Mahmut & Pınar & Umut, Hakan & Nilgün & Ela, Levent & Aysin & Poyraz & Rüzgar, Halime & Mehmet Ali & Deniz & Hazal. İyi ki varsınız!

Finally, last but not least, thank you Hande. You have supported me with your love, endless support, intellect and delicious food during the stressful writing process. While I lived in the Netherlands over 10 years now, I have enjoyed the years I have spent with you the most. I have found you when I was not looking. Eksik bir şey yok artık.

As a friend of mine once has written, the acknowledgements is the most fun, emotional and challenging part to write in the thesis. When one focuses on a topic for so long, he might forget some names. There are many others that deserve to be mentioned here: my gratitude goes out to all of you. Burada yazmayı unuttuysam da sizleri unutmadım. Hepinize teşekkürler!

## SUMMARY

The filter bubble is the theory that argues that personalization algorithms used in search engines, social networking platforms and other online services have too much focus on relevance. According to this theory, search engines may customize search results for users and provide different results for the same query for different users. For instance, one user who searches for the keyword “climate change” may see websites that argue that climate change is not real, while another user sees the opposite. Next to the algorithms, users themselves suffer from biases in information seeking and they may either search for biased information or form their online social networks only from like-minded users. In the end, users are exposed to opaque filters that they have no control over, receive only biased information and viewpoints that they agree with. As a result users rarely get challenged by opposing perspectives and ideas. This leads to the diminishment of viewpoint diversity, which is an undesired consequence for almost all theories of democracy.

The main research question of this thesis is the following: “How can we design for viewpoint diversity, so that filter bubbles are avoided?”. In order to answer this question, this thesis uses Value Sensitive Design (VSD) methodology. VSD consists of three steps: 1) a conceptual analysis where different values implicated by a certain technology are discovered and investigated using theories from other disciplines; 2) an empirical analysis that includes qualitative and quantitative methods such as observations, interviews and surveys; 3) a technical analysis to see how existing technological properties can support or hinder human values and how identified values can be translated, operationalized and implemented in system design. This thesis applies VSD to this research question by following the next three steps 1) conceptualize the phenomenon filter bubble by defining what it is, by identifying the factors leading to it and the consequences this brings for different models of democracy; 2) study the relevant value viewpoint diversity in a social networking platform empirically. This can be done by extending existing viewpoint diversity analyses in computer science by using existing theory in other disciplines such as media and communication and media ethics; 3) analyze existing design attempts and software tools to identify the democracy norms they promote and identify the missing norms.

While many authors and scholars have discussed the dangers of the filter bubble for democracy, they tend to focus on particular democracy models such as deliberative and liberal democracy. Further, existing work does not focus on the internal processes of these platforms and different factors for each process that can introduce bias and diminish diversity of information. Chapter 2 of this thesis uses and extends the traditional gatekeeping theory from media and communications field. It demonstrates the factors that affect personalization algorithms (e.g., user interaction history, location, novelty or user preferences), other filtering algorithms (e.g., information type and information age), and human operators of online services (e.g., pressure from advertisers or governments, personal judgments, organization policies). It focuses on processes such as col-

lection, selection, prioritization, deletion, withholding, disregarding, localization, customization or channeling. This chapter shows that the gatekeeping function of the media is not entirely replaced by users; platforms still hold control on which information gets out of the filter and which does not. Further, this chapter shows that bias is not only introduced by the algorithms. Human operators of online platforms may introduce bias after the algorithms become operational.

Chapter 3 of this thesis is one of the first studies that aims to measure the filter bubble empirically. It is one of the first studies that tests the filter bubble theory for different political cultures (Dutch and Turkish). It introduces different definitions of information diversity using the theory from communication studies and political philosophy. Later, it provides various metrics used in viewpoint diversity analysis and adds a new one, namely, minority access. Finally it presents the result of an empirical study performed by using these metrics. It demonstrates the importance of the newly introduced metric minority access by showing that minorities could not reach the majority of the studied Turkish Twittersphere, while this was not the case for the Dutch users.

Chapter 4 of this thesis analyzes software tools and design attempts to combat filter bubbles and how their goals compare to the norms required by different democracy models. To do this, this chapter first studies different democracy models and identifies reasons why the filter bubble is a problem for each model. Reasons include users having no control on opaque algorithms, diminishing epistemic quality of information, loss of respect towards different perspectives, the invisibility of minorities and politically disadvantaged due to an emphasis on consensus and mainstream opinions, not being aware of the oppressors, not being able to provide key information for political change, and the lack of public channels where the performance of the decision makers can be monitored and contested. It shows that the norms of only two of the democracy models can be found in the developed tools (liberal and deliberative democracy), while norms that belong to other models (such as agonism or contestatory democracy) are missing. This chapter argues that it would be beneficial if the designers were exposed to other conceptions of democracy to further the realisation that there is not just one concept of democracy. The chapter argues that as democracy itself is an ongoing cooperative social experimentation process, it would be beneficial for all to experiment with different norms of different theories and not just the popular ones.

Chapter 5 of this thesis discusses ethics of online social experiments using the controversy surrounding the so-called Facebook emotional contagion study [185] as the prime example. Online social experiments are relevant for the filter bubble because some experiments conducted by online platforms may cause filter bubbles and most tools developed to break the filter bubble themselves are experimental. This chapter describes arguments used by those who support such experiments in its current form and those who criticize them. Addressed topics include the benefits of these experiments for the individual and the society, many interpretations of the concept “informed consent”, the ubiquity of online social experiments, different perception of risk in online experiments, and the unavoidability of such experiments. This chapter argues that the asymmetrical relationship between data scientists and the users of social media platforms is what underpins the conflicting frames of reference. Furthermore, this chapter argues that, as long as there is no consensus regarding the ethics of online experiments

for all stakeholders, such conflicts are likely to arise again in the future. The chapter concludes that the public outcry in reaction to the Facebook study underlines that there is a growing expectation towards more transparency regarding how content is filtered and presented, beyond a zero-sum approach with unnecessary tradeoffs. While a company may have the interests of its users in mind, the chapter argues that users have to be able to assess these intentions for themselves, and evaluate the balance between their personal benefits and the interests of the company.

This thesis shows that viewpoint diversity can be measured differently depending on the chosen democracy theory and can produce different results for different political cultures. While online intermediaries have an important role in measuring and understanding viewpoint diversity, finding out whether bubbles exist is no easy task and it cannot be determined by one single study. There is a broader need for scientists, activists, individuals and social critics to study the online intermediaries on a long term, as the algorithms used by the platforms and the behavior of the users may change. Informed by the findings, designers can produce different solutions. This thesis also shows that designing for viewpoint diversity is no straightforward matter. This is due to different norms required by different models of democracy. While almost all of them value viewpoint diversity, how it conceptualized can be quite different. For this reason, this thesis argues that the designers must discuss their goals and the used norms with the public explicitly.



# SAMENVATTING

De filterbubble is de theorie dat beargumenteert dat personalisatie algoritmen die in zoekmachines, sociaal netwerkplatforms en andere online diensten gebruikt worden te veel de focus leggen op relevantie. Volgens deze theorie kunnen zoekmachines resultaten toespitsen op gebruikers en verschillende resultaten tonen aan verschillende gebruikers voor dezelfde zoekopdrachten. Bijvoorbeeld een gebruiker die zoekt op het trefwoord “klimaatverandering” kan websites zien die beargumenteren dat klimaatverandering niet waar is terwijl een andere gebruiker het tegenovergestelde ziet. Bovendien zijn gebruikers zelf aan vooroordelen in zoekstrategieën onderhevig en zoeken mogelijk naar bevooroordeelde informatie of onderhouden een sociaal netwerk met gelijkgezinde individuen. Uiteindelijk worden gebruikers blootgesteld aan ondoorzichtige filters waar zij geen controle over hebben, en ontvangen zij alleen bevooroordeelde informatie en optieken waar zij het mee eens zijn. Als resultaat worden gebruikers zelden blootgesteld aan contrasterende perspectieven en ideeën. Dit leidt tot het afnemen van opiniediversiteit welke een ongewenste consequentie is in vrijwel alle democratietheorieën.

De hoofdvraag van dit onderzoeksproefschrift is de volgende: “Hoe kan er ontworpen worden voor opiniediversiteit zodat de filterbubble vermeden wordt?”. Om deze vraag te beantwoorden gebruikt dit proefschrift de Value Sensitive Design (VSD; Waardegevoelig ontwerp) methodiek. VSD bestaat uit drie stappen: 1) een conceptuele analyse in welke verschillende waarden geïmpliceerd door een zekere technologie ontdekt en onderzocht worden met behulp van theorieën uit andere disciplines; 2) een empirische analyse die kwalitatieve en kwantitatieve methoden bevat zoals observaties, interviews en enquêtes; 3) een technische analyse om te zien hoe bestaande technologische eigenschappen menselijke waarden kunnen dragen of hinderen en hoe geïdentificeerde waarden vertaald, geoperationaliseerd en geïmplementeerd kunnen worden in systeemontwerp. Dit proefschrift past VSD op deze onderzoeksvraag toe in de volgende drie stappen: 1) het conceptualiseren van het filterbubble fenomeen door te definiëren wat het is, door het identificeren van de factoren die ertoe leiden en door de consequenties vast te stellen die het fenomeen heeft voor de verschillende democratiemodellen; 2) het empirisch bestuderen van de relevante waarde van opiniediversiteit in een sociaal netwerkplatform. Dit kan bereikt worden door het uitbreiden van opiniediversiteit analyses in de informatica met het gebruik van theorieën uit andere disciplines zoals media, communicatie en mediaethiek; 3) het analyseren van software implementaties en bestaande pogingen tot ontwerp om de democratiënormen vast te stellen die deze elementen bevorderen en zo ook de ontbrekende normen daarin.

Terwijl vele auteurs de gevaren van de filterbubble voor de democratie bediscussieerd hebben neigen zij te concentreren op bepaalde democratiemodellen zoals de deliberatieve en liberale democratie. Bestaande werken concentreren bovendien niet op de interne processen van deze platforms en verschillende factoren van elk proces die beooroordeling kunnen introduceren en diversiteit van informatie kunnen verminderen.

ren. Hoofdstuk 2 van dit proefschrift gebruikt en breid de conventionele gatekeeping theorie uit vanuit het veld van de media en communicatie. Het demonstreert de factoren die personalisatie algoritmen beïnvloeden (e.g. gebruikersinteractie geschiedenis, locatie, nieuwigheid, gebruikersinstellingen), andere filteringalgoritmen (e.g. informatietype, informatieleeftijd), menselijke bestuurders van online diensten (e.g. druk van adverteerders of overheden, persoonlijk oordeel, organisatiebeleid). Het concentreert op processen zoals verzamelen, selectie, prioriteren, verwijderen, onderdrukken, negeren, lokaliseren, toespitsen, sturen. Dit hoofdstuk toont dat de gatekeeperfunctie van de media niet geheel vervangen wordt door gebruikers; platforms behouden controle over welke informatie door het filter heen komt en welke niet. Verder laat dit hoofdstuk zien dat bevoordeling niet alleen geïntroduceerd wordt door algoritmen. Menselijke bestuurders van online platforms kunnen bevoordeling introduceren nadat algoritmen operationeel worden.

Hoofdstuk 3 van dit proefschrift is één van de eerste studies die zich richt op het empirisch vaststellen van de filterbubble. Het is één van de eerste studies die de filterbubble theorie test in verschillende politieke culturen (de Nederlandse en de Turkse). Het introduceert verschillende definities van informatiediversiteit en gebruikt daarvoor theorie uit de communicatiewetenschap en politieke filosofie. Verder verschaft het uiteenlopende metrieken die gebruikt worden in opiniediversiteit analyse en draagt een nieuwe bij, namelijk minderheidsbereik. Tenslotte toont het de resultaten van een empirisch onderzoek uitgevoerd met deze metrieken. Het demonstreert het belang van de hier geïntroduceerde minderheidsbereik metriek door het vaststellen dat minderheden niet de meerderheid van de Turkse Twittersfeer kan bereiken, terwijl dit wel het geval was voor de Nederlandse gebruikers.

Hoofdstuk 4 van dit proefschrift analyseert software implementaties en ontwerppogingen die filterbubbles bestrijden en hoe vergelijkbaar hun doelen zijn met de normen vereist door de verschillende democratiemodellen. Om dit te doen bestudeert dit hoofdstuk eerst verschillende democratiemodellen en identificeert de redenen waarom de filterbubble een probleem is voor elk model. Redenen omvatten gebruikers die geen controle hebben over ondoorzichtige algoritmen, de verminderende epistemische kwaliteit van informatie, het verlies van respect voor afwijkende optieken, de onzichtbaarheid van minderheden en de politieke minderbevoorechten door een nadruk op consensus en gemiddelde meningen, het niet bewust zijn van onderdrukkers, het niet in staat zijn om cruciale informatie nodig voor politieke verandering te verstrekken, het gebrek aan publieke kanalen waardoor toezicht gehouden kan worden op de prestatie van beleidsvormers en deze kan worden betwist. Het toont dat de normen van slechts twee democratiemodellen (de deliberatieve en liberale democratie) terug gevonden kunnen worden in de ontwikkelde implementaties, terwijl de normen die behoren tot andere modellen (zoals het agonisme en de betwistingsdemocratie) ontbreken. Dit hoofdstuk beargumenteert dat het nuttig zou zijn als de ontwerpers blootgesteld zouden zijn aan andere democratische concepten en zich zouden realiseren dat er niet slechts één concept van democratie bestaat. Het hoofdstuk beargumenteert dat omdat democratie zelf een voortdurend coöperatief sociaal experimenteel proces is, het voor allen bevorderlijk zou zijn om te experimenteren met de verschillende normen van verschillende theorieën, en niet slecht de populaire.

Hoofdstuk 5 van dit proefschrift handelt over de ethiek van online sociale experimenten met als schoolvoorbeeld de controverse rondom de zogenoemde “Facebook emotional contagion study” (Facebook emotionele besmettelijkheidsstudie [185]). Online sociale experimenten zijn relevant voor de filterbubble omdat sommige experimenten die gehouden worden door online platforms filterbubbles kunnen veroorzaken en de meeste implementaties om filterbubbles te doorbreken zijn experimenteel. Dit hoofdstuk beschrijft argumenten gebruikt door degenen die zulke experimenten in de huidige vorm steunen en door hen die ze bekritisieren. Geadresseerde ontwerpen zijn de voordelen van deze experimenten voor het individu en de samenleving, de vele interpretaties van het concept “informed consent” (weloverwogen instemming), de alomtegenwoordigheid van online sociale experimenten, verschillende percepties van risico in online experimenten, de onontkoombaarheid van zulke experimenten. Dit hoofdstuk beargumenteert dat de asymmetrische relatie tussen datawetenschappers en de gebruikers van sociale media platforms hetgeen is dat de contrasterende referentiekaders onderbouwt. Verder beargumenteert dit hoofdstuk dat zolang er geen consensus is over de ethiek van online experimenten voor alle belanghebbenden zulke conflicten zich waarschijnlijk wederom zullen voordoen in de toekomst. Dit hoofdstuk concludeert dat het publieke protest in reactie tegen de Facebook studie onderschrijft dat er een groeiende verwachting is van meer transparantie omtrent hoe materiaal gefilterd en gepresenteerd wordt, in verzet tegen een stagnering met als gevolg onnodige inlevering. Terwijl een bedrijf wellicht het belang van de gebruiker als oogpunt heeft, beargumenteert dit hoofdstuk dat de gebruikers in staat moeten zijn deze intenties zelf te beoordelen en het evenwicht tussen hun persoonlijke voordelen en de belangen van het bedrijf te bepalen.

Dit proefschrift toont dat opiniediversiteit verschillend gemeten kan worden afhankelijk van de gekozen democratietheorie en verschillende resultaten kan produceren voor verschillende politieke culturen. Terwijl online intermediairen een belangrijke rol hebben in het meten en begrijpen van opiniediversiteit, is het geen gemakkelijke taak om uit te zoeken of een bubble bestaat en kan het niet met een enkele studie vastgesteld worden. Er is een brede noodzaak voor het bestuderen door wetenschappers, activisten, individuen en sociaal critici van de online intermediairen op de lange termijn, aangezien algoritmen gebruikt door platforms en het gedrag van gebruikers kunnen veranderen. Geïnformeerd door resultaten kunnen ontwerpers verschillende oplossingen produceren. Dit proefschrift toont ook aan dat ontwerpen met opiniediversiteit als oogpunt geen eenvoudige zaak is. Dit door de verschillende normen vereist door verschillende democratiemodellen. Terwijl bijna alle modellen opiniediversiteit waarderen, kan hoe het geconceptualiseerd wordt nogal verschillend zijn. Om deze reden beargumenteert dit proefschrift dat de ontwerpers hun doelen en de gebruikte normen met het publiek expliciet moeten bespreken.



# CURRICULUM VITÆ

## Engin BOZDAĞ

Engin Bozdag was born in Malatya, Turkey on 1 January 1982. He obtained both his BSc and MSc degrees in computer science from Delft University of Technology in 2006 and 2008, respectively. Between September 2008 and November 2010, he worked in the industry, including Google and ASR Insurances. From November 2010 to November 2014, he was a PhD candidate in the Values and Technology Department at the Delft University of Technology and a fellow of Alexander von Humboldt Institute for Internet and Society. In 2010, he was awarded a prestigious Mozaiek grant, issued by Netherlands Organisation for Scientific Research (NWO).

## 6.4. SCIENTIFIC PUBLICATIONS

### 6.4.1. JOURNAL ARTICLES

1. **E. Bozdag**, Q. Gao, Houben, G. J., and M. Warnier, *Does Offline Political Segregation Affect the Filter Bubble? An Empirical Analysis of Information Diversity for Dutch and Turkish Twitter Users*, Computers in Human Behavior, Elsevier, 41, 405-415, (2014).
2. C. Puschmann, and **E. Bozdag**, *Staking out the unclear ethical terrain of online social experiments*, Internet Policy Review, 3, 4, (2014).
3. **E. Bozdag**, *Bias in algorithmic filtering and personalization*, Ethics and Information Technology, Springer, 15,3, 209-227, (2013)
4. **E. Bozdag**, A. Mesbah, and A. van Deursen, *Performance testing of data delivery techniques for AJAX applications*, Journal of Web Engineering 8, 287 (2009).

### 6.4.2. CONFERENCE PROCEEDINGS

1. **E. Bozdag**, Q. Gao, M. Warnier, and G. Houben, *Analyzing Viewpoint Diversity in Twitter*, Proceedings of “Designing Social Media for Change” Workshop, The ACM SIGCHI Conference on Human Factors in Computing Systems, (2013).
2. **E.Bozdag**, I. van de Poel, *Designing for diversity in online personalization systems*, Technology Management for Emerging Technologies Conference (PICMET), Proceedings of PICMET '13, (2013).

3. **E.Bozdag** and Q. Gao, *An Empirical Research Design for Analyzing Information Diversity in Twitter for Dutch and Turkish Users*, Proceedings of ETHICOMP 2013, University of Southern Denmark, Kolding, Denmark, 12-14 June, (2013).
4. **E. Bozdag** and J.Timmermans, *Values in the Filter Bubble: Ethics of Personalization Algorithms in Cloud Computing*, Proceedings of the 13th IFIP TC13 Conference on Human-Computer Interaction (Interact), Workshop on Values in Design - Building Bridges between HCI, RE and Ethics, Lisbon, Portugal, (2011).
5. J. Timmermans, B. C. Stahl, V. Ikonen, and **E. Bozdag**, *The Ethics of Cloud Computing, a Conceptual Review*, in 2nd IEEE International Conference on Cloud Computing Technology and Science, Indianapolis, USA, pp. 614-620,(2010).
6. **E. Bozdag** and A. van Deursen, *An adaptive push/pull algorithm for AJAX applications*, Third International Workshop on Adaptation and Evolution in Web Systems Engineering, (AEWSE 08), (2008).
7. A. Mesbah, **E. Bozdag**, and A. van Deursen, *Crawling AJAX by inferring user interface state changes*, In 8th International Conference on Web Engineering (ICWE08) (2007).
8. **E. Bozdag**, A. Mesbah, and A. van Deursen, *A comparison of push and pull techniques for AJAX*, Proceedings of the 9th IEEE International Symposium on Web Site Evolution (WSE), pages 15- 22. IEEE Computer Society (2007).

## 6.5. PROFESSIONAL/POPULAR PUBLICATIONS

1. **E. Bozdag** and M.Warnier, *Social Media and the Protests in Istanbul, democratizing or polarizing?* (Dutch), Volkskrant, (2013).
2. **E. Bozdag** and M.Warnier, *Polarisatie door sociale media Turkije*, (Dutch), NRC Handelsblad, (2013).
3. **E. Bozdag**, *Facebook'un online deneyleri* (Turkish), Radikal Online (2014).

## REFERENCES

- [1] K Acar. Globalization And Language: English In Turkey. *Sosyal Bilimler*, 2(1), 2004.
- [2] Accuracast. Facebook Advertising Policies Homophobic, may 2010. URL <http://searchdailynews.blogspot.com/2010/05/facebook-advertising-policies.html>.
- [3] La Adamic and N Glance. The political blogosphere and the 2004 US election: divided they blog. *Proceedings of the 3rd international workshop on ...*, pages 36–43, 2005. ISSN 1595932151. doi: 10.1145/1134271.1134277. URL <http://dl.acm.org/citation.cfm?id=1134271.1134277>.
- [4] Gediminas Adomavicius, Ramesh Sankaranarayanan, Shahana Sen, and Alexander Tuzhilin. Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Trans. Inf. Syst.*, 23(1):103–145, jan 2005. ISSN 1046-8188. doi: 10.1145/1055709.1055714. URL <http://doi.acm.org/10.1145/1055709.1055714>.
- [5] John Agada. Inner-City Gatekeepers: An Exploratory Survey of Their Information Use Environment. *Journal of the American Society for Information Science*, 50(1):74–85, 1999. ISSN ISSN-0002-8231. URL <http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ582286>.
- [6] Bekir Agirdir. Polarization in Politics and the Society. Technical report, Konda Research, Istanbul, 2010.
- [7] Kristoffer Ahlström. Why Deliberative Democracy is (Still) Untenable. *Public Affairs Quarterly*, 26(3), 2012.
- [8] Chloe Albanesius. Schmidt, Yelp Clash Over Google's Search Tactics, 2011. URL <http://www.pcmag.com/article2/0,2817,2393369,00.asp>.
- [9] Anders Albrechtslund. Ethics and technology design. *Ethics and Information Technology*, 9(1):63–72, 2007. ISSN 13881957. doi: 10.1007/s10676-006-9129-8.
- [10] Tamara Alsheikh and JA Rode. (Whose) value-sensitive design: a study of long-distance relationships in an Arabic cultural context. *of the ACM 2011 conference on*, 2011. URL <http://dl.acm.org/citation.cfm?id=1958836>.
- [11] S.L. Althaus and D. Tewksbury. Agenda setting and the “new” news. *Communication Research*, 29(2):180, 2002.

- [12] Hakan Altintas. The Impact of Political Parties and Urbanisation on the Process of Polarisation in the Political System. *Akdeniz İktisadi ve İdari Bilimler Fakültesi Dergisi*, 3(5):1–31, 2003.
- [13] American Press Institute. The Personal News Cycle: How Americans choose to get their news. Technical report, 2014. URL <http://www.americanpressinstitute.org/publications/reports/survey-research/personal-news-cycle/>.
- [14] J An, Meeyoung Cha, and KP Gummadi. Visualizing media bias through Twitter. *AAAI Technical Report WS-12-01 The Potential of Social Media Tools and Data for Journalists in the News Media Industry*, 2012.
- [15] Jisun An, Meeyoung Cha, Krishna Gummadi, and Jon Crowcroft. Media Landscape in Twitter: A World of New Conventions and Political Diversity. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pages 18–25, may 2011.
- [16] Jisun An, Daniele Quercia, and Jon Crowcroft. Partisan Sharing : Facebook Evidence and Societal Consequences. In *COSN' 14*, number Figure 1, 2014.
- [17] Elizabeth Anderson. The Epistemology of Democracy. *Episteme*, 3(1-2):8–22, jan 2006. ISSN 1742-3600. doi: 10.3366/epi.2006.3.1-2.8.
- [18] C Arthur. Facebook emotion study breached ethical guidelines, researchers say, 2014. URL <http://www.theguardian.com/technology/2014/jun/30/facebook-emotion-study-breached-ethical-guidelines-researchers-say>.
- [19] M Aslama, E de Bens, J van Cuilenburg, K Nordenstreng, W Schulz, and R van der Wurff. Measuring and assessing empirical media diversity: Some European cases. In Els de Bens, editor, *Media between Culture and Commerce*, pages 55–98. Intellect, Bristol, 2007.
- [20] Association of Internet Researchers. Ethical decision-making and internet research: Recommendations from the AoIR Ethics Working Committee (Version 2.0), 2012. URL <http://aoir.org/reports/ethics2.pdf>.
- [21] L Backstrom. News Feed FYI: a window into News Feed., 2012. URL <https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed>.
- [22] David Badash. Has Facebook Censorship Gone Too Far?, 2011. URL <http://thenewcivilrightsmovement.com/has-facebook-censorship-gone-too-far/politics/2011/11/07/29714>.
- [23] Ben H. Bagdikian. *The New Media Monopoly: A Completely Revised and Updated Edition With Seven New Chapters*. Beacon Press, may 2004. ISBN 0807061875.

- [24] E. Bakshy, S. Messing, and L. Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, pages science.aaa1160–, may 2015. ISSN 0036-8075. doi: 10.1126/science.aaa1160. URL <http://www.sciencemag.org/content/early/2015/05/08/science.aaa1160.abstract>.
- [25] Eytan Bakshy, Dean Eckles, Rong Yan, and Itamar Rosenn. Social influence in social advertising. In *Proceedings of the 13th ACM Conference on Electronic Commerce - EC '12*, number 212, page 146, 2012. ISBN 9781450314152. doi: 10.1145/2229012.2229027. URL <http://arxiv.org/abs/1206.4327%5C%22delimit%20E3B2%24%22http://dl.acm.org/citation.cfm?doid=2229012.2229027>.
- [26] Eytan Bakshy, I Rosenn, Cameron Marlow, and L Adamic. The role of social networks in information diffusion. *Proceedings of the 21st international conference on World Wide Web*, pages 519–528, 2012. URL <http://dl.acm.org/citation.cfm?id=2187907>.
- [27] Judit Bar-Ilan, Kevin Keenoy, Mark Levene, and Eti Yaari. Presentation bias is significant in determining user preference for search results-A user study. *Journal of the American Society for Information Science and Technology*, 60(1):135–149, jan 2009. ISSN 15322882, 15322890. doi: 10.1002/asi.20941. URL <http://doi.wiley.com/10.1002/asi.20941>.
- [28] Pablo Barbera. How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S. 2014. URL <http://j.mp/BarberaPolarization>.
- [29] Pablo Barbera and Gonzalo Rivero. Understanding the political representativeness of Twitter users. *Social Science Computer Review*, Published, 2014.
- [30] Ruken Baris. The Turkish Media Landscape. In Georgios Terzis, editor, *European Media Governance: National and Regional Dimensions*, pages 289–301. Intellect Books, 2007. ISBN 9781841501925.
- [31] K. Barzilai-Nahon. Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American society for information science and technology*, 59(9):1493–1512, 2008.
- [32] K. Barzilai-Nahon. Gatekeeping: A critical review. *Annual Review of Information Science and Technology*, 43(1):1–79, 2009.
- [33] BBC. Official complaint filed over Facebook emotion study, 2014. URL <http://www.bbc.com/news/technology-28157889>.
- [34] BBC. OKCupid experiments with 'bad' dating matches, 2014. URL <http://www.bbc.com/news/technology-28542642>.
- [35] Michael a. Beam and Gerald M. Kosicki. Personalized News Portals: Filtering Systems and Increased News Exposure. *Journalism & Mass Communication Quarterly*, 91(1):59–77, 2014. ISSN 1077-6990. doi: 10.1177/1077699013514411. URL <http://jmq.sagepub.com/lookup/doi/10.1177/1077699013514411>.

- [36] Yochai Benkler and Helen Nissenbaum. Commons-based Peer Production and Virtue. *Journal of Political Philosophy*, 14(4):394–419, dec 2006. ISSN 0963-8016. doi: 10.1111/j.1467-9760.2006.00235.x. URL <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1467-9760.2006.00235.x>.
- [37] Jeremy Bentham. An Introduction to the Principles of Morals and Legislation, 1780. URL <http://www.econlib.org/library/Bentham/bnthPML.html>.
- [38] Daniel Berdichevsky and Erik Neuenschwander. Toward an ethics of persuasive technology. *Communications of the ACM*, 42(5):51–58, may 1999. ISSN 00010782. doi: 10.1145/301353.301410. URL <http://portal.acm.org/citation.cfm?doid=301353.301410>.
- [39] M. Bernstein. The destructive silence of social computing researchers., 2014. URL <https://medium.com/@msbernst/the-destructive-silence-of-social-computing-researchers-9155cdf659>.
- [40] Susan Bickford. *The Dissonance of Democracy, Listening, Conflict, and Citizenship*. Cornell University Press, 1996. URL <http://www.cornellpress.cornell.edu/book/?GCOI=80140100358380>.
- [41] James Bohman. Deliberative Democracy and the Epistemic Benefits of Diversity. *Episteme*, 3(03):175–191, 2006. doi: 10.3366/epi.2006.3.3.175.
- [42] Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. A 61-million-person experiment in social influence and political mobilization, 2012. ISSN 0028-0836.
- [43] Alan Borning and Michael Muller. Next Steps for Value Sensitive Design. In *CHI '12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1125–1134, 2012. ISBN 9781450310154.
- [44] Danah Boyd and Kate Crawford. Six Provocations for Big Data. In *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, Oxford, sep 2011. doi: 10.2139/ssrn.1926431.
- [45] Engin Bozdag. Bias in algorithmic filtering and personalization. *Ethics and Information Technology, Springer*, 15(3):209–227, 2013.
- [46] Engin Bozdag and Job Timmermans. Values in the Filter Bubble: Ethics of Personalization Algorithms in Cloud Computing. In *Workshop on Values in Design - Building Bridges between RE, HCI & Ethics held in conjunction with INTERACT 2011 13th IFIP TC13 Conference on Human-Computer Interaction*, 2011.
- [47] Kees Brants. Guest Editor's Introduction: The Internet and the Public Sphere. *Political Communication*, 22(2):143–146, apr 2005. ISSN 1058-4609. doi: 10.1080/10584600590933133.
- [48] A. Bruns. Gatewatching, Gatecrashing: Futures for Tactical News Media. In *Digital media and democracy: tactics in hard times*, pages 247–271. The MIT Press, 2008.

- [49] Axel Bruns. Gatekeeping, Gatewatching, Real-time Feedback. *Brazilian Journalism Research*, 7:117–136, 2011.
- [50] T. Busch and T. Shepherd. Doing well by doing good? Normative tensions underlying Twitter’s corporate social responsibility ethos. *Convergence: The International Journal of Research into New Media Technologies*, 20(3):293–315, 2014. ISSN 1354-8565. doi: 10.1177/1354856514531533. URL <http://con.sagepub.com/content/20/3/293.abstract.html?etoc>.
- [51] Bryan Caplan. *The Myth of the Rational Voter: Why Democracies Choose Bad Policies*. Princeton University Press, Princeton, N.J.; Woodstock, new editio edition, aug 2008. ISBN 9780691138732.
- [52] Ali Carkoglu and Ersin Kalaycioglu. *Turkish Democracy Today: Elections, Protest and Stability in an Islamic Society*. I.B.Tauris, New York, 2007. ISBN 1845111850.
- [53] Manuel Castells. *Communication Power*. Oxford University Press, USA, 2011. ISBN 0199595690.
- [54] Rory Cellan-Jones. Facebook tightens its grip on news, 2015. URL <http://www.bbc.com/news/technology-32707014>.
- [55] Elfreda A. Chatman. Opinion Leadership, Poverty, and Information Sharing. *RQ*, 26(3):341–53, 1987. URL <http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ354348>.
- [56] Ching-chih Chen and Peter Hernon. *Information seeking: assessing and anticipating user needs*. Neal-Schuman Publishers, 1982. ISBN 9780918212504.
- [57] Jilin Chen, Rowan Nairn, Les Nelson, Michael Bernstein, and Ed Chi. Short and tweet: experiments on recommending content from information streams. In *CHI 2010, CHI ’10*, pages 1185–1194, Atlanta, 2010. ACM. ISBN 978-1-60558-929-9. doi: 10.1145/1753326.1753503.
- [58] Steve Chibnall. The Crime Reporter: A Study in the Production of Commercial Knowledge. *Sociology*, 9(1):49 –66, jan 1975. doi: 10.1177/003803857500900103. URL <http://soc.sagepub.com/content/9/1/49.abstract><http://soc.sagepub.com/content/9/1/49.short>.
- [59] ChillingEffects. Scientology Complains that Advanced Technology appears in Google Groups, 2005. URL <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=2355>.
- [60] Tom Christiano. Democracy, jul 2006. URL <http://plato.stanford.edu/entries/democracy/>.
- [61] John Christman. Liberalism and Individual Positive Freedom \*. *Ethics*, 101(2): 343–359, 2011.

- [62] Cisco. Cisco Visual Networking Index: Forecast and Methodology, whitepaper., 2011.
- [63] Joshua Cohen. An epistemic conception of democracy. *Ethics*, 97(1):26–38, 1986. URL <http://www.jstor.org/stable/2381404>.
- [64] Joshua Cohen. Procedure and Substance in Deliberative Democracy. In James Bohman and William Rehg, editors, *Deliberative Democracy: Essays on Reason and Politics*, page 447. MIT Press, 1997. ISBN 0262522411.
- [65] Joshua Cohen. Reflections on Deliberative Democracy. In Thomas Christiano and John Christman, editors, *Contemporary Debates in Political Philosophy*, chapter 14. Blackwell, West-Sussex, UK, 2009.
- [66] Stephen Coleman and John Götze. Bowling Together: Online Public Engagement in Policy Deliberation. *Information Polity The International Journal of Government Democracy in the Information Age*, page 48, 2002. ISSN 1570-1255. URL <http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=10388900&site=ehost-live>.
- [67] Elanor Colleoni, Alessandro Rozza, and Adam Arvidsson. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in twitter using big data, 2014. ISSN 00219916.
- [68] Michael Conover, Jacob Ratkiewicz, and Matthew Francisco. Political polarization on twitter. *ICWSM*, 133(26):89–96, 2011. ISSN 15205126. doi: 10.1021/ja202932e.
- [69] J Constine. Your average Facebook post only reaches 12% of your friends. *TechCrunch*, 2012. URL <http://techcrunch.com/2012/02/29/facebook-post-reach-16-friends/>.
- [70] Dick Costolo. The trends are algorithmic, not chosen by us but we edit out any w/ obscenities, jul 2011. URL <https://twitter.com/{#}!/dickc/status/97686216681594880>.
- [71] Kate Crawford. Can an algorithm be agonistic? Ten scenes about living in calculated publics. In *Governing Algorithms 2013*, 2013.
- [72] Lincoln Dahlberg. Computer-Mediated Communication and The Public Sphere: A Critical Analysis. *Journal of Computer-Mediated Communication*, 7(1), jun 2001. ISSN 10836101. doi: 10.1111/j.1083-6101.2001.tb00137.x. URL <http://doi.wiley.com/10.1111/j.1083-6101.2001.tb00137.x>.
- [73] Christopher A Le Dantec, Erika Shehan Poole, and Susan P Wyche. Values as Lived Experience : Evolving Value Sensitive Design in Support of Value Discovery. *Design*, pages 1141–1150, 2009.
- [74] Sauvik Das and Adam Kramer. Self-Censorship on Facebook. *Seventh International AAAI Conference on Weblogs and Social Media*, pages 120–127, 2013. doi: 10.1007/b104039.

- [75] Janet Davis. Household indicators: design to inform and engage citizens. In *Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems*, volume 2, pages 688–693, 2006. ISBN 1595932984. doi: 10.1145/1125451.1125591. URL <http://doi.acm.org/10.1145/1125451.1125591>.
- [76] Janet L. Davis and Lisa P. Nathan. Value Sensitive Design: Applications, Adaptations and Critiques. In Jeroen Van den hoven, Ibo Van de poel, and Pieter Vermaas, editors, *Handbook of Ethics, Values, and Technological Design*, pages 1–26. Springer Netherlands, 2014.
- [77] I Van de Poel. Values in engineering design. *Handbook of the philosophy of science*, 9, 2009.
- [78] Vincent Dekker. Google: een zwijgzame rechter en politieagent, 2006.
- [79] S DellaVigna and E Kaplan. The Fox News Effect: Media Bias and Voting. *The Quarterly Journal of Economics*, 2007.
- [80] T. Demir and E. Ben-Zadok. Politically Driven Regulations and Enforcement: Monitoring Kurd and Fundamentalist Broadcasts in Turkey. *Administration & Society*, 39(2):262–293, apr 2007. ISSN 0095-3997.
- [81] M DeRosa. How informed consent has failed. *TechCrunch*, 2014. URL <http://techcrunch.com/2014/07/26/how-informed-consent-has-failed/>.
- [82] John Dewey. *The Public and Its Problems*. Holt, New York, 1927. ISBN 0271055693.
- [83] Nicholas Diakopoulos. Algorithmic accountability reporting: On the investigation of black boxes. *Tow Center for Digital Journalism Brief, Columbia ...*, 2014.
- [84] Nicholas Diakopoulos. Towards a Standard for Algorithmic Transparency in the Media. Technical report, Tow Center for Digital Journalism, 2015. URL <http://towcenter.org/towards-a-standard-for-algorithmic-transparency-in-the-media/>.
- [85] Nicholas Diakopoulos and Mor Naaman. Towards quality discourse in online news comments. *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 133–142, 2011. doi: 10.1145/1958824.1958844. URL <http://dx.doi.org/10.1145/1958824.1958844>.
- [86] A Diaz. Through the Google Goggles : Sociopolitical Bias in Search Engine Design. *Knowledge Management*, 2008.
- [87] Carl DiSalvo. *Adversarial Design*. MIT Press, 2012.
- [88] Abraham Doris-Down, Husayn Versee, and Eric Gilbert. Political Blend: An Application Designed to Bring People Together Based on Political Differences. In *Proc. Intl. Conf. on Communities and Technologies (C&T)*, pages 120–130, 2013. ISBN 9781450321044. doi: 10.1145/2482991.2483002.

- [89] Steve Dorsey. Turkey's Social Media And Smartphones Key To 'Occupy Gezi' Protests, 2013.
- [90] John S. Dryzek. *Discursive Democracy: Politics, Policy, and Political Science*. Cambridge University Press, 1994. ISBN 0521478278.
- [91] John Dunn. *Western Political Theory in the Face of the Future*, volume 3. Cambridge University Press, 1979. ISBN 0521295785.
- [92] Joan C. Durrance. *Armed for Action Library Response to Citizen Information Needs*. Neal Schuman Pub, jan 1984. ISBN 0918212715.
- [93] B. Edelman. Bias in Search Results?: Diagnosis and Response. *Indian Journal of Law and Technology*, 7(0), 2011.
- [94] Amir Efrati. Rivals Say Google Plays Favorites, dec 2010. ISSN 0099-9660. URL <http://online.wsj.com/article/SB10001424052748704058704576015630188568972.html>.
- [95] Khalid El-Arini, Ulrich Paquet, Ralf Herbrich, Jurgen Van Gael, and Blaise Agüera y Arcas. Transparent user models for personalization. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '12*, pages 678–686, New York, New York, USA, aug 2012. ACM Press. ISBN 9781450314626. doi: 10.1145/2339530.2339639. URL <http://dl.acm.org/citation.cfm?id=2339530.2339639>.
- [96] Mike Elgan. How to pop your Internet 'filter bubble'. *Computerworld*, 2011.
- [97] Jon Elster. The Market and the Forum: Three Varieties of Political Theory. In James Bohman and William Rehg, editors, *Deliberative Democracy: Essays on Reason and Politics*, chapter 1, pages 3–34. The MIT Press, 1997. ISBN 0521596963. doi: 10.1177/019145370102700505. URL <http://www.amazon.com/dp/0262522411>.
- [98] Engin Bozdag, Qi Gao, Geert Jan Houben, and Martijn Warnier. Does Offline Political Segregation Affect the Filter Bubble? An Empirical Analysis of Information Diversity for Dutch and Turkish Twitter Users. *Computers in Human Behavior*, 41: 405 – 415, 2014.
- [99] Martin J. Eppler and Jeanne Mengis. The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *The Information Society*, 20(5):325–344, 2004. ISSN 0197-2243. doi: 10.1080/01972240490507974. URL <http://www.tandfonline.com/doi/abs/10.1080/01972240490507974>.
- [100] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. “I always assumed that I wasn't really that close to [her]”: Reasoning about invisible algorithms in the news feed. In *Human Factors in Computing Systems Conference (CHI), 2015.*, Seoul, Korea, 2015.

- [101] Deborah Fallows. Search Engine Users. Technical report, Pew Research, 2005. URL <http://www.pewinternet.org/Reports/2005/Search-Engine-Users/8-Conclusions/Conclusions.aspx>.
- [102] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. Opinion space: a scalable tool for browsing online comments. *Proceedings of the 28th international conference on Human factors in computing systems*, pages 1175–1184, 2010. doi: 10.1145/1753326.1753502. URL <http://dl.acm.org/citation.cfm?id=1753502>.
- [103] Feminist Frequency. One Week of Harassment on Twitter, 2015. URL <http://femfreq.tumblr.com/post/109319269825/one-week-of-harassment-on-twitter>.
- [104] Leon Festinger. *A theory of cognitive dissonance*, volume 207. Stanford University Press, 1957. ISBN 0804709114. doi: 10.1037/10318-001.
- [105] P Fischer, D Frey, C Peus, and A Kastenmuller. The theory of cognitive dissonance: state of the science and directions for future research. *Clashes of Knowledge*, pages 189–198, 2008.
- [106] James S. Fishkin. *Democracy and Deliberation: New Directions for Democratic Reform*. Yale University Press, 1993. ISBN 0300051638.
- [107] Mark Fishman. *Manufacturing the News*. Univ of Texas Pr, sep 1988. ISBN 0292751044.
- [108] Mary Flanagan, Daniel C. Howe, and Helen Nissenbaum. Embodying Values in Technology Theory and Practice. In Jeroen van den Hoven and John Weckert, editors, *Information Technology and Moral Philosophy*. Cambridge University Press, 2008. ISBN 1139469711.
- [109] Jennifer Fong. Facebook’s Bias Against 3rd Party Apps, 2011. URL <http://www.jenfongspeaks.com/facebooks-bias-against-3rd-party-apps/>.
- [110] D. Freelon, M. Lynch, and S. Aday. Online Fragmentation in Wartime: A Longitudinal Analysis of Tweets about Syria, 2011-2013. *The ANNALS of the American Academy of Political and Social Science*, 659(1):166–179, apr 2015. ISSN 0002-7162. doi: 10.1177/0002716214563921. URL <http://ann.sagepub.com/content/659/1/166.abstract>.
- [111] Deen G. Freelon, Travis Kriplean, Jonathan Morgan, W. Lance Bennett, and Alan Borning. Facilitating Diverse Political Engagement with the Living Voters Guide. *Journal of Information Technology & Politics*, 9(3):279–297, jul 2012. ISSN 1933-1681. doi: 10.1080/19331681.2012.665755. URL <http://dx.doi.org/10.1080/19331681.2012.665755>.
- [112] B. Friedman, N. Freier, and Jr. Kahn, P.H. Office window of the future? Two case studies of an augmented window. *Extended Abstracts of CHI 2004 Conference on Human Factors in Computing Systems (CHI EA '04)*, page 1559, 2004. ISSN 1581137036. doi: 10.1145/985921.986135.

- [113] Batya Friedman and Helen Nissenbaum. Bias in Computer Systems. *ACM Transactions on Information Systems*, 14(3):330—347, 1996.
- [114] Batya Friedman and Helen Nissenbaum. Software agents and user autonomy. *Proceedings of the first international conference on Autonomous agents - AGENTS '97*, pages 466–469, 1997. doi: 10.1145/267658.267772. URL <http://portal.acm.org/citation.cfm?doid=267658.267772>.
- [115] Batya Friedman, Peyina Lin, and JK Miller. Informed consent by design. In *Security and Usability*, number 2001, pages 503–530. O'Reilly Media, 2005. ISBN 0-596-00827-9.
- [116] Batya Friedman, Peter H Kahn Jr., and Alan Borning. Value Sensitive Design and Information Systems: Three Case Studies. In *Human-Computer Interaction and Management Information Systems: Foundations*. M.E. Sharp Inc, New York, 2006. ISBN 9400778430.
- [117] Batya Friedman, Alan Borning, Janet L Davis, Brian T Gill, Peter H Kahn, Travis Kriplean, and Peyina Lin. Laying the Foundations for Public Participation and Value Advocacy : Interaction Design for a Large Scale Urban Simulation. *Design*, 1(2):305–314, 2008. URL <http://portal.acm.org/citation.cfm?id=1367883>.
- [118] Batya Friedman, Kristina Hook, Brian Gill, Lina Eidmar, Catherine Sallmander Prien, and Rachel L. Severson. Personlig integritet: A comparative study of perceptions of privacy in public places in Sweden and the United States. In *Proceedings of the 5th Nordic Conference on Human-Computer Interaction Building Bridges (NordiCHI '08)*, pages 142–151, 2008. ISBN 9781595937049. doi: 10.1145/1463160.1463176.
- [119] Herbert J. Gans. *Deciding What's News: A Study of CBS Evening News, NBC Nightly News, Newsweek, and Time*. Northwestern University Press, 2 edition, feb 2005. ISBN 0810122375.
- [120] Hector Garcia-Molina, Georgia Koutrika, and Aditya Parameswaran. Information seeking. *Communications of the ACM*, 54(11):121, nov 2011. ISSN 00010782. doi: 10.1145/2018396.2018423.
- [121] Jim Garrison. Reconstructing Democracy and Recontextualizing Deweyan Pragmatism. In Jim Garrison, editor, *Reconstructing Democracy, Recontextualizing Dewey: Pragmatism and Interactive Constructivism in the Twenty-First Century*, chapter 1, pages 1–17. State University of New York Press, Albany, 2008.
- [122] S. Gauch, M. Speretta, A. Chandramouli, and A. Micarelli. User profiles for personalized information access. *The Adaptive Web*, pages 54–89, 2007.
- [123] J. Gerhards and M. S. Schafer. Is the internet a better public sphere? Comparing old and new media in the USA and Germany. *New Media & Society*, 12(1):143–160, jan 2010. ISSN 1461-4448. doi: 10.1177/1461444809341444.

- [124] T Gillespie. Facebook's algorithm - why our assumptions are wrong, and our concerns are right. *Culture Digitally*, 2014.
- [125] Tarleton Gillespie. Can an Algorithm be Wrong? *Limn*, (2), 2012. URL <http://limn.it/can-an-algorithm-be-wrong/>.
- [126] Dan Gillmor. *We the Media: Grassroots Journalism By the People, For the People*, volume 6. O'Reilly Media, 2006. ISBN 0596553919.
- [127] Carroll J. Glynn, Susan Herbst, Garrett J. O'Keefe, Robert Y. Shapiro, and Mark Lindeman. *Public Opinion*. Westview Press, 2004. ISBN 0813341728.
- [128] V Goel. Facebook tinkers with users' emotions in News Feed experiment, stirring outcry, 2014.
- [129] E. Goldman. Search engine bias and the demise of search engine utopianism. *Yale JL & Tech.*, 8:188, 2005.
- [130] Eric Goldman. Revisiting search engine bias 38.1 (2011): 12-11. *William Mitchell Law Review*, 38(1.12-11), 2011.
- [131] Google. We knew the web was big..., 2008. URL <http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>.
- [132] Google. Transparency Report, 2011. URL <http://www.google.com/transparencyreport/governmentrequests/>.
- [133] Google. Increasing transparency and choice with ads on search and Gmail, 2011. URL <http://googleblog.blogspot.nl/2011/10/>.
- [134] Google. Search, plus your world. Your photos, your friends, your stuff., 2012. URL <http://www.google.com/insidesearch/features/plus/index.html>.
- [135] Laura a. Granka. The Politics of Search: A Decade Retrospective. *The Information Society*, 26(5):364–374, sep 2010. ISSN 0197-2243. doi: 10.1080/01972243.2010.511560. URL <http://tandfprod.literatumonline.com/doi/abs/10.1080/01972243.2010.511560>.
- [136] Mark S. Granovetter. *The strength of weak ties : a network theory revisited*. State University of New York, Department of Sociology, 1981.
- [137] M Gray. When science, customer service, and human subjects research collide. Now what? *Ethnography Matters*, 2014.
- [138] Catherine Grevet, Loren G Terveen, and Eric Gilbert. Managing Political Differences in Social Media. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1400–1408, 2014. doi: 10.1145/2531602.2531676.
- [139] J Grimmelmann. The Facebook emotional manipulation study: sources, 2014. URL <http://goo.gl/zFcAIF>.

- [140] J Grimmelmann. Letter to Inder M. Verma, editor-in-chief, PNAS, 2014.
- [141] J Grimmelmann. Illegal, Immoral, and Mood-Altering. How Facebook and OkCupid Broke the Law When They Experimented on Users. *Medium*, 2014. URL <http://goo.gl/TGwUXp>.
- [142] JM Grohol. Emotional contagion on Facebook? More like bad research methods. *PsychCentral*, 2014.
- [143] M. Gromping. 'Echo Chambers': Partisan Facebook Groups during the 2014 Thai Election. *Asia Pacific Media Educator*, 24(1):39–59, 2014. ISSN 1326-365X. doi: 10.1177/1326365X14539185. URL <http://ame.sagepub.com/lookup/doi/10.1177/1326365X14539185>.
- [144] J Groot. Trouw wekenlang niet te vinden op Google. *Webwereld*, 2004.
- [145] Anatolii Gruzd and Jeffrey Roy. Investigating Political Polarization on Twitter: A Canadian Perspective. *Policy & Internet*, 6(1):28–45, 2014. ISSN 19442866. doi: 10.1002/1944-2866.POI354. URL <http://doi.wiley.com/10.1002/1944-2866.POI354>.
- [146] Saikat Guha, Bin Cheng, and Paul Francis. Challenges in Measuring Online Advertising Systems. *Public Policy*, 2010.
- [147] Pritam Gundecha and Huan Liu. Mining Social Media: A Brief Introduction. *Tutorials in Operations Research*, (Dmml):1–17, 2012. doi: <http://dx.doi.org/10.1287/educ.1120.0105>.
- [148] Amy Gutmann and Dennis Thompson. *Why Deliberative Democracy?* Princeton University Press, Princeton, N.J, aug 2004. ISBN 9780691120195.
- [149] Juergen Habermas. *Legitimation Crisis*. Beacon Press, 1975. ISBN 0807015210.
- [150] Jurgen Habermas. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. MIT Press, Cambridge, Mass., reprint ed edition, mar 1998. ISBN 9780262581622.
- [151] K. S. Hahn, S. Ryu, and S. Park. Fragmentation in the Twitter Following of News Outlets: The Representation of South Korean Users' Ideological and Generational Cleavage. *Journalism & Mass Communication Quarterly*, 92(1):56–76, 2015. ISSN 1077-6990. doi: 10.1177/1077699014559499. URL <http://jmq.sagepub.com/lookup/doi/10.1177/1077699014559499>.
- [152] Timur Hammond and Elizabeth Angell. Is Everywhere Taksim?: Public Space and Possible Publics. *Jadaliyya*, 2013. URL <http://goo.gl/gR7Y9v>.
- [153] Russell Hardin. Deliberative Democracy. In Thomas Christiano and John Christman, editors, *Contemporary Debates in Political Philosophy*, chapter 13. Blackwell, West-Sussex, UK, 2009.

- [154] Eszter Hargittai. Why doesn't Science publish important methods info prominently?, 2015. URL <http://crookedtimber.org/2015/05/07/why-doesnt-science-publish-important-methods-info-prominently/>.
- [155] Eszter Hargittai, Jason Gallo, and Matthew Kane. Cross-ideological discussions among conservative and liberal bloggers. *Public Choice*, 134(1-2):67–86, 2008. ISSN 00485829. doi: 10.1007/s11127-007-9201-x.
- [156] Erika Check Hayden. Informed consent: A broken contract, 2012. ISSN 0028-0836.
- [157] N. Helberger. Diversity by Design. *Journal of Information Policy*, 1, 2011. URL <http://jip.vmhost.psu.edu/ojs/index.php/jip/article/viewArticle/59>.
- [158] David Held. *Models of Democracy, 3rd Edition*. Stanford University Press, 2006. ISBN 0804754721.
- [159] Alfred Hermida. Tweets and Truth: Journalism as a discipline of collaborative verification. *Journalism Practice*, pages 1–14, 2012. ISSN 1751-2786. doi: 10.1080/17512786.2012.667269. URL <http://www.caerdydd.ac.uk/jomec/resources/foj2011/foj2011-Hermida.pdf>.
- [160] Martin Hilbert. Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin*, 138(2):211–237, 2012. ISSN 1939-1455, 0033-2909. doi: 10.1037/a0025940. URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0025940>.
- [161] K Hill. Facebook manipulated 689,003 users' emotions for science. *Forbes*, 2014. URL <http://www.forbes.com/sites/kashmirhill/2014/06/28/facebook-manipulated-689003-users-emotions-for-science/>.
- [162] Matthew Hindman. *The Myth of Digital Democracy*. Princeton University Press, 2008. ISBN 0691138680.
- [163] Hitwise. Social networks now more popular than search engines in the UK, 2010. URL <http://www.hitwise.com/index.php/uk/press-centre/press-releases/2010/social-media-alert-june-2010/>.
- [164] Lindsay H. Hoffman and Hui Fang. Quantifying Political Behavior on Mobile Devices over Time: A User Evaluation Study. *Journal of Information Technology & Politics*, (June):140602091913000, 2014. ISSN 1933-1681. doi: 10.1080/19331681.2014.929996. URL <http://dx.doi.org/10.1080/19331681.2014.929996>.
- [165] R Howell. OK, Stupid-OK Cupid's Ethical Confusion, 2014. URL <http://rjhjr.com/dailysabbatical/?p=639>.
- [166] IBM. Bringing smarter computing to big data, 2011. URL <http://goo.gl/H566Jz>.

- [167] Mathew Ingram. The downside of Facebook as a public space: Censorship, jun 2011. URL <http://gigaom.com/2011/06/21/the-downside-of-facebook-as-a-public-space-censorship/>.
- [168] Lucas Introna. Defining the Web : The Politics of Search Engines. *Computer*, pages 54–62, 2000.
- [169] Glenn Jacobs. TechRadar: How to optimise your site for Google Caffeine, apr 2010. URL <http://www.techradar.com/news/internet/how-to-optimise-your-site-for-google-caffeine-685436>.
- [170] T. Joachims and F. Radlinski. Search engines that learn from implicit feedback. *Computer*, 40(8):34–40, 2007.
- [171] P. Jürgens, A. Jungherr, and H. Schoen. Small worlds with a difference: New gatekeepers and the filtering of political information on twitter. In *WebSci '11*, 2011.
- [172] Pascal Jürgens. Measuring Personalization: An Experimental Framework for Testing Technological Black Boxes. In *Society of the Query*, 2013. URL <http://networkcultures.org/wpmu/query/tag/pascal-jurgens/>.
- [173] Nathan Jurgenson. Facebook: Fair and Balanced, 2015. URL <http://thesocietypages.org/cyborgology/2015/05/07/facebook-fair-and-balanced/>.
- [174] K. Karppinen. Rethinking media pluralism and communicative abundance. *Observatorio (OBS\*)*, 3(4), 2009.
- [175] Kari Karppinen. *Rethinking Media Pluralism*. Fordham University Press, onald mcga edition, 2013. ISBN 0823245128.
- [176] Elihu Katz. And Deliver Us from Segmentation. *Annals of the American Academy of Political and Social Science*, 546:22–33, jul 1996. ISSN 0002-7162. URL <http://www.jstor.org/stable/1048167><http://www.jstor.org/stable/pdfplus/1048167.pdf?acceptTC=true>.
- [177] Elihu Katz and Paul Lazarsfeld. *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Transaction Publishers, oct 2005. ISBN 1412805074.
- [178] Minjeong Kim and Han Woo Park. Measuring Twitter-based political participation and deliberation in the South Korean context by using social network and Triple Helix indicators. *Scientometrics*, 90(1):121–140, jan 2012. URL <http://link.springer.com/10.1007/s11192-011-0508-5>.
- [179] Jason Kincaid. TechCrunch | Today's Lesson: Make Facebook Angry, And They'll Censor You Into Oblivion, 2010. URL <http://techcrunch.com/2010/11/22/facebook-censorship/>.

- [180] Hakan M. Kiris. Parti Sisteminde Kutuplasma ve Türk Parti Örneği. *Amme İdaresi Dergisi*, 44(4):33–67, 2011.
- [181] Jacob Klein. A Web Marketer's Guide to Reddit, dec 2011. URL <http://www.distilled.net/blog/social-media/a-web-marketers-guide-to-reddit/>.
- [182] Will Knight. Google Hopes to Make Friends with a More Social Search - Technology Review, 2012. URL <http://www.technologyreview.com/computing/39444/> <http://www.technologyreview.com/computing/39444/?p1=MstRcnt>.
- [183] Andy Kohut. Pew surveys of audience habits suggest perilous future for news, 2013. URL <http://www.poynter.org/news/mediawire/225139/pew-surveys-of-audience-habits-suggest-perilous-future-for-news/>.
- [184] Aleksandra Korolova. Privacy Violations Using Microtargeted Ads : A Case Study. In *IEEE international conference on data mining workshops (ICDMW '10)*, 2010.
- [185] Adam D I Kramer, Jamie E Guillory, and Jeffrey T Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 111(24): 8788–90, 2014. ISSN 1091-6490. doi: 10.1073/pnas.1320040111. URL <http://www.ncbi.nlm.nih.gov/pubmed/24889601>.
- [186] David Kravets. A bot exposes Twitter's financials—was the scraping an illegal hack? *Arstechnica*, 2015. URL <http://arstechnica.com/tech-policy/2015/04/a-bot-exposes-twitters-financials-was-the-scraping-an-illegal-hack/>.
- [187] Travis Kriplean, Michael Toomim, JT Morgan, Alan Borning, and AJ Ko. REFLECT: Supporting active listening and grounding on the Web through restatement. In *Computer Supported Cooperative Work (CSCW)*, 2011. ISBN 9781450305563.
- [188] Travis Kriplean, Jonathan Morgan, Deen Freelon, Alan Borning, and Lance Bennett. Supporting reflective public thought with considerit. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12*, page 265, New York, New York, USA, feb 2012. ACM Press. ISBN 9781450310864. doi: 10.1145/2145204.2145249. URL <http://dl.acm.org/citation.cfm?id=2145204.2145249>.
- [189] Andre Krouwel. Zo links zijn de kranten niet. Technical report, Dag, 2008.
- [190] Andre Krouwel. Regeerakkoord analyse: Kieskompas, 2012.
- [191] JP Kruijt. De begrippen verzuiling en ontzuiling. *Socialisme en Democratie*, pages 791–797, 1962.
- [192] Shamanth Kumar, Fred Morstatter, and Huan Liu. *Twitter Data Analytics*. Springer Briefs in Computer Science, 2013.

- [193] A. Lafrance. How Much Should You Know About the Way Facebook Works? *the Atlantic*, 2014. URL <http://www.theatlantic.com/technology/archive/2014/08/how-much-should-you-know-about-how-facebook-works/378812/>.
- [194] Adrienne Lafrance. Facebook is Eating the Internet, 2015. URL <http://www.theatlantic.com/technology/archive/2015/04/facebook-is-eating-the-internet/391766/>.
- [195] H. Landemore. Democratic Reason: the Mechanisms of Collective Intelligence in Politics. In *Collective Wisdom: Principle and Mechanisms*, number April 2011. Cambridge University Press, Cambridge, 2012.
- [196] DL Lasorsa, SC Lewis, and A Holton. Normalizing Twitter-Journalism practice in an emerging communication space. *Journalism Studies*, 13(1):19–36, 2012.
- [197] Dominic L. Lasorsa, Seth C. Lewis, and Avery E. Holton. Normalizing Twitter. *Journalism Studies*, 13(1):19–36, feb 2012. ISSN 1461-670X. doi: 10.1080/1461670X.2011.571825.
- [198] T. Lavie, M. Sela, I. Oppenheim, O. Inbar, and J. Meyer. User attitudes towards news content personalization. *International Journal of Human-Computer Studies*, 68(8):483–495, 2009.
- [199] David Lazer. The rise of the social algorithm. *Science*, 2015. doi: 10.1126/science.aab1422. URL <http://www.sciencemag.org/content/early/2015/05/06/science.aab1422.full{#}target-1>.
- [200] N Lennard. The troubling link between Facebook's emotion study and Pentagon research. *Vice*, 2014. URL <https://goo.gl/t3Z20P>.
- [201] Paul Levinson. *Digital McLuhan: A Guide to the Information Millennium*. Routledge, 1 edition, apr 1999. ISBN 041519251X. URL <http://www.amazon.com/dp/041519251X>.
- [202] QV Liao and WT Fu. Beyond the filter bubble: interactive effects of perceived threat and topic involvement on selective exposure to information. *Proceedings of the SIGCHI Conference on Human ...*, 2013. doi: 10.1145/2470654.2481326. URL <http://dl.acm.org/citation.cfm?id=2481326>.
- [203] QV Liao and WT Fu. Expert voices in echo chambers: effects of source expertise indicators on exposure to diverse opinions. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 2745–2754, 2014.
- [204] A Lijphart. *The politics of Accommodation: Pluralism and Democracy in the Netherlands*. University of California Press, 1968.
- [205] Zhe Liu and Ingmar Weber. Is Twitter a Public Sphere for Online Conflicts? A Cross-Ideological and Cross-Hierarchical Look. *Social Informatics Lecture Notes in Computer Science Volume*, 8851:336–347, 2014.

- [206] Zhe Liu and Ingmar Weber. Predicting ideological friends and foes in Twitter conflicts. In *Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion*, pages 575–576, Seoul, Korea, apr 2014. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-2745-9. doi: 10.1145/2567948.2576964. URL <http://dl.acm.org/citation.cfm?id=2567948.2576964>.
- [207] Benedikt Loepp and Jürgen Ziegler. Blended Recommending : Integrating Interactive Information Filtering and Algorithmic Recommender Techniques. In *CHI 2015*, pages 975–984, 2015.
- [208] T Lorenz. Plugin allows you to recreate Facebook’s controversial mood-altering experiment on YOUR News Feed, 2014.
- [209] Gilad Lotan. Data Reveals That “Occupying” Twitter Trending Topics is Harder Than it Looks!, 2011. URL <http://goo.gl/Ok9UXb>.
- [210] Y. Lu. The human in human information acquisition: Understanding gatekeeping and proposing new directions in scholarship. *Library & information science research*, 29(1):103–123, 2007.
- [211] Victor Luckerson. Here’s How Facebook’s News Feed Actually Works. *Time*, 2015. URL <http://time.com/3950525/facebook-news-feed-algorithm/>.
- [212] G. Mackie. Does democratic deliberation change minds? *Politics, Philosophy & Economics*, 5(3):279–303, oct 2006. ISSN 1470-594X. doi: 10.1177/1470594X06068301. URL <http://ppe.sagepub.com/cgi/doi/10.1177/1470594X06068301>.
- [213] James Madison. Federalist 10. *The Federalist Papers*, (10):1–7, 1787.
- [214] Noëmi Manders-Huits. What Values in Design? The Challenge of Incorporating Moral Values into Design. *Science and Engineering Ethics*, 17(2):271–287, 2011. ISSN 13533452. doi: 10.1007/s11948-010-9198-2.
- [215] Noëmi Manders-Huits and Michael Zimmer. Values and Pragmatic Action: The Challenges of Introducing Ethical Intelligence in Technical Design Communities. In Felton Emma, Zelenko Oksana, and Vaughan Suzi, editors, *Design and Ethics: Reflections on Practice*, pages 57–73. Routledge, 2012.
- [216] Bernard Manin. *The Principles of Representative Government*. Cambridge University Press, 1997. URL <http://www.cambridge.org/us/academic/subjects/politics-international-relations/political-theory/principles-representative-government>.
- [217] James Manyika, Michael Chui, Brad Brown, Jacques Buighin, Richard Dobbs, and Roxburgh. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute report. Technical report, McKinsey Global Institute, 2011. URL <http://goo.gl/fgGfbW>.

- [218] G Marsman. *De katholieke dagbladders in sociologisch perspectief*. Van Gorcum, Nimegue, 1967.
- [219] D. McQuail and J. J. van Cuilenburg. Diversity as a Media Policy Goal: a Strategy for Evaluative Research and a Netherlands Case Study. *International Communication Gazette*, 31(3):145–162, jan 1983. ISSN 1748-0485.
- [220] Derek Mead. Facebook Is Making Us All Live Inside Emotional 'Filter Bubbles'. *Motherboard (Vice)*, 2014. URL <http://motherboard.vice.com/read/facebook-is-making-us-all-live-inside-emotional-filter-bubbles>.
- [221] Mediamonitor. The concept of pluralism: media diversity, 2013. URL <http://www.mediamonitor.nl/english/the-concept-of-pluralism-media-diversity/>.
- [222] Cade Metz. Google contradicts own counsel in face of antitrust probe, admits existence of search algorithm whitelists, 2011. URL <http://goo.gl/1pL254>.
- [223] Cade Metz. Google opens curtain on 'manual' search penalties, 2011. URL <http://goo.gl/fsLuiE>.
- [224] M Meyer. Misjudgements will drive social trials underground. *Nature*, 511(7509): 265, 2014.
- [225] Robinson Meyer. Facebook as a Press Baron. *The Atlantic*, 2015. URL <http://www.theatlantic.com/technology/archive/2015/03/facebook-as-a-press-baron/388559/>.
- [226] JS Mill. *On liberty*. Batoche, Kitchener, jun 1859. ISBN 9780486421308.
- [227] J.K. Miller, B. Friedman, G. Jancke, and Brian Gill. Value tensions in design: The value sensitive design, development, and appropriation of a corporation's groupware system. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work (GROUP '07)*, pages 281–290, 2007. ISBN 9781595938459. doi: 10.1145/1316624.1316668.
- [228] Patrick R. Miller and Pamela Johnston Conover. Why partisan warriors don't listen: the gendered dynamics of intergroup anxiety and partisan conflict. *Politics, Groups, and Identities*, 3(1):21–39, 2014. ISSN 2156-5503. doi: 10.1080/21565503.2014.992795. URL <http://www.tandfonline.com/doi/abs/10.1080/21565503.2014.992795>.
- [229] LI Millett, Batya Friedman, and Edward Felten. Cookies and Web browser design: toward realizing informed consent online. In *Proceeding CHI '01 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, volume 3, pages 46–52, 2001. ISBN 1581133278. doi: 10.1145/365024.365034.
- [230] Amy Mitchell, Jeffrey Gottfried, Jocelyn Kiley, and Katerina Matsa. Political Polarization & Media Habits. Technical report, Pew Research Center, 2014. URL <http://www.journalism.org/files/2014/10/>

[Political-Polarization-and-Media-Habits-FINAL-REPORT-11-10-14-2.pdf](#).

- [231] Evgeny Morozov. Your Own Facts. Book review of "The Filter Bubble, What the Internet Is Hiding From You", 2011. URL [http://www.nytimes.com/2011/06/12/books/review/book-review-the-filter-bubble-by-eli-pariser.html?\\_r=1&pagewanted=all](http://www.nytimes.com/2011/06/12/books/review/book-review-the-filter-bubble-by-eli-pariser.html?_r=1&pagewanted=all).
- [232] Chantal Mouffe. Deliberative Democracy or Agonistic Pluralism? *Social Research*, 66(3):745–758, 1999. ISSN 0037783X. doi: Article.
- [233] Chantal Mouffe. *The Democratic Paradox*. Verso, 2009.
- [234] A. Mowshowitz and A. Kawaguchi. Bias on the Web. *Communications of the ACM*, 45(9):56–60, 2002.
- [235] F Muench. The new Skinner Box: web and mobile analytics. *Psychology Today*, 2014. URL <https://www.psychologytoday.com/blog/more-tech-support/201403/the-new-skinner-box-web-and-mobile-analytics>.
- [236] Meltem Müftüler Baç. Turkey's Political Reforms and the Impact of the European Union. *South European Society and Politics*, 10(1):17–31, apr 2005. ISSN 1360-8746.
- [237] Meltem Muftuler-Bas and Fuat E. Keyman. Turkey Under the AKP. *Journal of Democracy*, 23(1):85–99, 2012.
- [238] SA Munson and P Resnick. Presenting diverse political opinions: how and how much. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2010*, pages 1457–1466, New York, New York, USA, 2010.
- [239] Sean A. Munson, Stephanie Y. Lee, and Paul Resnick. Encouraging Reading of Diverse Political Viewpoints with a Browser Widget. In *International Conference on Weblogs and Social Media (ICWSM)*, Boston, 2013.
- [240] DC Mutz. *Hearing the Other Side: Deliberative versus Participatory Democracy*. Cambridge University Press, 2006.
- [241] Gautham Nagesh. Privacy advocates want Facebook probed on recent changes. *The Hill*, 2011.
- [242] Sayooran Nagulendra and Julita Vassileva. Understanding and Controlling the Filter Bubble through Interactive Visualization : A User Study. In *HT 14*, pages 107–115, 2014. ISBN 9781450329545.
- [243] Pm Napoli. Deconstructing the diversity principle. *Journal of Communication*, 49(4):7–34, 1999. ISSN 1460-2466. doi: 10.1111/j.1460-2466.1999.tb02815.x.
- [244] Dimitar Nikolov, Diego F M Oliveira, Alessandro Flammini, Filippo Menczer, and S I Feb. Measuring Online Social Bubbles. 2015.

- [245] Helen Nissenbaum, Lucas D. Inrona. Shaping the Web: Why the Politics of Search Engines Matters. *The Information Society*, 16(3):169–185, jul 2000. ISSN 0197-2243. doi: 10.1080/01972240050133634.
- [246] Katherine Noyes. The FTC is worried about algorithmic transparency, and you should be too. *PCWorld*, 2015.
- [247] Jolie O’Dell. Facebook’s Ad Revenue Hit \$1.86B for 2010, 2011. URL <http://mashable.com/2011/01/17/facebooks-ad-revenue-hit-1-86b-for-2010/>.
- [248] Claus Offe and Ulrich Klaus Preuss. Democratic Institutions and Moral Resources. In *Political Theory Today*, pages 143–171. Stanford University Press, 1991.
- [249] Kieron O’Hara. In worship of an echo. *IEEE Internet Computing*, 18(4):79–83, 2014. ISSN 10897801. doi: 10.1109/MIC.2014.71.
- [250] Kieron O’Hara and David Stevens. Echo Chambers and Online Radicalism : Assessing the Internet ’ s Complicity in Violent Extremism. *Policy and Internet*, Online Fir, 2015.
- [251] Kerem Oktem. Why Turkey’s mainstream media chose to show penguins rather than protests. *Guardian*, 2013. URL <http://www.theguardian.com/commentisfree/2013/jun/09/turkey-mainstream-media-penguins-protests>.
- [252] Kurt Opsahl. Google Begins Behavioral Targeting Ad Program, 2009. URL <https://www.eff.org/deeplinks/2009/03/google-begins-behavioral-targeting-ad-program>.
- [253] Muhsin Ozturk. Agirdir: Bugunku kutuplasma 1970’lerdekinden daha vahim. *Aksiyon*, 2013. URL <http://t24.com.tr/haber/agirdir-bugunku-kutuplasma-1970lerdekinden-daha-vahim/241939>.
- [254] Matthew Panzarino. Twitter’s Timeline Could Get (More) Algorithmic. *TechCrunch*, 2014.
- [255] Z. Papacharissi. The virtual sphere: The internet as a public sphere. *New Media & Society*, 4(1):9–27, feb 2002. ISSN 1461-4448. doi: 10.1177/14614440222226244. URL <http://nms.sagepub.com/content/4/1/9.short>.
- [256] Eli Pariser. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Press, New York, may 2011. ISBN 1594203008.
- [257] Eli Pariser. Did Facebook’s Big New Study Kill My Filter Bubble Thesis? *Medium*, 2015. URL <https://goo.gl/AI8jn2>.
- [258] Souneil Park, Seungwoo Kang, Sangyoung Chung, and Junehwa Song. NewsCube: Delivering Multiple Aspects of News to Mitigate Media Bias. In *Proceedings of CHI ’09, the SIGCHI Conference on Human Factors in Computing Systems*, pages 443–453, 2009. ISBN 9781605582467. doi: 10.1145/1518701.1518772.

- [259] Souneil Park, KyungSoon Lee, and Junehwa Song. Contrasting Opposing Views of News Articles on Contentious Issues. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, pages 340–349, 2011. ISBN 9781932432879. URL <https://www.aclweb.org/anthology-new/P/P11/P11-1035.pdf>.
- [260] N Paumgarten. Make me a match. *New Yorker, The*, 2014. URL <http://www.newyorker.com/magazine/2014/08/25/2710913>.
- [261] JW Pennebaker, RJ Booth, and ME Francis. Linguistic Inquiry and Word Count: LIWC2007, 2007.
- [262] Iryna Pentina and Monideepa Tarafdar. From "information" to "knowing": Exploring the role of social media in contemporary news consumption. *Computers in Human Behavior*, 35:211–223, 2014. ISSN 07475632. doi: 10.1016/j.chb.2014.02.045.
- [263] M. Peter van der Hoek. Does the dutch model really exist? *International Advances in Economic Research*, 6(3):387–403, aug 2000. ISSN 1083-0898. doi: 10.1007/BF02294959.
- [264] Philip Pettit. *Republicanism: A Theory of Freedom and Government*. Oxford University Press, 1999.
- [265] Pew Research. Bias Perceived in News Coverage, 2012. URL <http://www.pewresearch.org/daily-number/bias-perceived-in-news-coverage/>.
- [266] Wolter Pieters. Explanation and trust: what to tell the user in security and AI? *Ethics and Information Technology*, 13(1):53–64, nov 2010. ISSN 1388-1957. doi: 10.1007/s10676-010-9253-3. URL <http://www.springerlink.com/index/10.1007/s10676-010-9253-3>.
- [267] Alina Pommeranz, Joost Broekens, Pascal Wiggers, Willem-Paul Brinkman, and Catholijn M. Jonker. Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process. *User Modeling and User-Adapted Interaction*, mar 2012. ISSN 0924-1868. doi: 10.1007/s11257-011-9116-6. URL <http://www.springerlink.com/index/10.1007/s11257-011-9116-6>.
- [268] H Post. *Pillarization: an analysis of Dutch and Belgian society*. Gower Publishing Company, Vermont, 1989.
- [269] M Priestley. Honest News in the Slashdot Decade. *First Monday*, 4(8), 1999. URL <http://www.firstmonday.org/issues/issue4{ }8/priestley/index.html>.
- [270] Cornelius Puschmann and Engin Bozdag. Staking out the unclear ethical terrain of online social experiments. *Internet Policy Review*, 3(4), 2014.

- [271] Emilee Rader and Rebecca Gray. Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed. In *CHI '15 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 173–182, New York, USA, 2015. ACM. ISBN 9781450331456. URL <http://dl.acm.org/citation.cfm?id=2702174>.
- [272] John Rawls. *A Theory of Justice*. Harvard University Press, 1971.
- [273] John Rawls. The Idea of Public Reason. In James Bohman and William Rehg, editors, *Deliberative Democracy: Essays on Reason and Politics*, page 447. MIT Press, 1997. ISBN 0262522411.
- [274] Joanna Redden and Tamara Witschge. A new news order? Online news content examined. In Natalie Fenton, editor, *New media, old news: Journalism and democracy in the digital age*, pages 171–186. SAGE Publications Ltd, 2009.
- [275] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. GroupLens: an open architecture for collaborative filtering of netnews. *Information Retrieval*, pages 175–186, 1994. doi: 10.1145/192844.192905. URL <http://doi.acm.org/10.1145/192844.192905><http://dl.acm.org/ft{ }gateway.cfm?id=192905{&}type=pdf>.
- [276] M Robbins. Does OKCupid need our consent ?, 2014. URL <http://www.theguardian.com/science/the-lay-scientist/2014/jul/30/does-okcupid-need-our-consent>.
- [277] Matt Rosoff. Here's How Dominant Google Is In Europe. *Business Insider UK*, 2014.
- [278] Matt Rosoff. Here's How Dominant Google Is In Europe. *Business Insider*, 2015.
- [279] C Rudder. We Experiment On Human Beings!, 2014. URL <http://blog.okcupid.com/index.php/we-experiment-on-human-beings/>.
- [280] D Saez-Trumper, Carlos Castillo, and Mounia Lalmas. Social media news communities: gatekeeping, coverage, and statement bias. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1679–1684, New York, NY, USA, 2013. ACM. ISBN 9781450322638.
- [281] Amir Salihefendic. How Reddit ranking algorithms work, 2010. URL <http://amix.dk/blog/post/19588>.
- [282] Christian Sandvig. Corrupt personalization, 2014. URL <http://socialmediacollective.org/2014/06/26/corrupt-personalization>.
- [283] Christian Sandvig. What Facebook's 'It's Not Our Fault' Study Really Means. *Wired*, 2015. URL <http://www.wired.com/2015/05/facebook-not-fault-study/>.
- [284] Christian Sandvig, Kevin Hamilton, Karrie G Karahalios, and Cedric Landbort. Auditing Algorithms: Research Methods for Detecting Discrimination on Internet

- Platforms. In *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*, 64th Annual Meeting of the International Communication Association, Seattle, 2014.
- [285] Christian Sandvig, Karrie G Karahalios, and Cedric Langbort. Look inside the Facebook News Feed, 2014. URL <http://goo.gl/yBc5B4>.
- [286] Steven Michael Schneider. *Expanding the Public Sphere through Computer-Mediated Communication: Political Discussion about Abortion*. PhD thesis, Massachusetts Institute of Technology, 1997.
- [287] Stan Schroeder. Twitter Ad Revenue May Reach \$150 Million This Year, 2011. URL <http://mashable.com/2011/01/24/twitter-revenue-150-million/>.
- [288] Stefan Schulz-Hardt, Dieter Frey, Carsten Lüthgens, and Serge Moscovici. Biased information search in group decision making. *Journal of Personality and Social Psychology*, 78(4):655–669, 2000. ISSN 1939-1315(Electronic);0022-3514(Print). doi: 10.1037/0022-3514.78.4.655.
- [289] Searchenginewatch. Twitter: Google Search Plus Your World Bad for Web Users, 2012. URL <http://searchenginewatch.com/article/2136873/Twitter-Google-Search-Plus-Your-World-Bad-for-Web-Users>.
- [290] David Segal. Search Optimization and Its Dirty Little Secrets, feb 2011. ISSN 0362-4331. URL <http://www.nytimes.com/2011/02/13/business/13search.html>[http://www.nytimes.com/2011/02/13/business/13search.html?\\_r=3&pagewanted=all](http://www.nytimes.com/2011/02/13/business/13search.html?_r=3&pagewanted=all).
- [291] CE Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379 – 423, 1948.
- [292] U. Shardanand and P. Maes. Social information filtering: algorithms for automating “word of mouth”. In *CHI '95 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 210–217, 1995.
- [293] Katie Shilton, Jes Koepfler, and Kenneth Fleischmann. How to see values in social computing: methods for studying values dimensions. *Computer-Supported Cooperative Work and Social Computing*, 2014. URL <http://terpconnect.umd.edu/~kshilton/ShiltonCSCW2014preprint.pdf>.
- [294] Pamela J. Shoemaker and Timothy Vos. *Gatekeeping Theory*. Routledge, 1 edition, apr 2009. ISBN 0415981395. URL <http://www.amazon.com/dp/0415981395>.
- [295] P.J. Shoemaker, T.P. Vos, and S.D. Reese. Journalists as gatekeepers. *Handbook of Journalism Studies*, pages 73–87, 2008.
- [296] Philip Elliot Slater. Role Differentiation in Small Groups. *American Sociological Review*, 20(3):300–310, 1955.

- [297] Jilles Smids. The voluntariness of persuasive technology. *Persuasive Technology Design for Health and Safety*, pages 123–132, 2012. URL [http://link.springer.com/chapter/10.1007/978-3-642-31037-9\\_11](http://link.springer.com/chapter/10.1007/978-3-642-31037-9_11).
- [298] Graham Smith. Taking deliberation seriously: institutional design and green politics. *Environmental Politics*, 10(3):72–93, 2001. URL <http://www.tandfonline.com/doi/abs/10.1080/714000562>.
- [299] Marc Smith, Lee Rainie, Ben Shneiderman, and Itai Himelboim. Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters. Technical report, Pew Research Center and Social Media Research Foundation, 2014.
- [300] Barry Smyth. A Community-Based Approach to Personalizing Web Search. *Computer*, 40(8):42–50, aug 2007. ISSN 0018-9162. doi: 10.1109/MC.2007.259. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4292011>.
- [301] Lawrence C. Soley. *Censorship, Inc: the corporate threat to free speech in the United States*. Monthly Review Press, jun 2002. ISBN 9781583670668. URL <http://books.google.nl/books?id=VREoAQAAMAAJ>.
- [302] Andreas Spahn. And lead us (not) into persuasion...? Persuasive technology and the ethics of communication. *Science and engineering ethics*, 18(4):633–50, dec 2012. ISSN 1471-5546. doi: 10.1007/s11948-011-9278-y. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3513602&tool=pmcentrez&rendertype=abstract>.
- [303] R Steininger. Pillarization and Political Parties. *Sociologische Gids*, pages 242–257, 1977.
- [304] Greg Sterling. Bing Reaches 20 Percent Search Market Share Milestone In US. *Search Engine Land*, 2015.
- [305] Greg Sterling. Bing Reaches 20 Percent Search Market Share Milestone In US. *Search Engine Land*, 2015. URL <http://searchengineland.com/bing-reaches-20-percent-search-milestone-in-us-market-218574>.
- [306] Natalie Jomini Stroud. *Niche News : The Politics of News Choice Abstract and Keywords*. Number March 2014. Oxford University Press, 2011. ISBN 9780199755509. doi: 10.1093/acprof.
- [307] Paul Sturges. Gatekeepers and other intermediaries. *Aslib Proceedings*, 53(2):62–67, jan 2001. ISSN 0001-253X. doi: 10.1108/EUM0000000007039.
- [308] Danny Sullivan. Google's Results Get More Personal With "Search Plus Your World", 2012. URL <http://searchengineland.com/googles-results-get-more-personal-with-search-plus-your-world-107285>.
- [309] M Sullivan. Facebook investor Andreessen sarcastically defends Facebook's mood experiment, 2014. URL <http://goo.gl/41lpZs>.

- [310] Tao Sun, M Zhang, and Qiaozhu Mei. Unexpected Relevance: An Empirical Study of Serendipity in Retweets. In *Seventh International AAAI Conference on Weblogs and Social Media (ICWSM'13)*, 2013.
- [311] Cass R. Sunstein. *Republic.com*. Princeton University Press, aug 2002. ISBN 9780691143286.
- [312] Cass R Sunstein. The Law of Group Polarization. *Journal of Political Philosophy*, 10(2):175–195, dec 2002. ISSN 0963-8016. doi: 10.1111/1467-9760.00148.
- [313] Cass R. Sunstein. Preferences, Paternalism, and Liberty. *Royal Institute of Philosophy Supplements*, 59:233–264, 2006. doi: 10.1017/S135824610605911X.
- [314] Cass R. Sunstein. *Republic.com 2.0*. Princeton University Press, first edit edition, aug 2007. ISBN 0691133565.
- [315] Cass R. Sunstein. *Infotopia: How Many Minds Produce Knowledge*. Oxford University Press, USA, jul 2008. ISBN 0195340671. URL <http://www.amazon.com/dp/0195340671>.
- [316] Juha Suoranta and Tere Vadén. *Wikiworld*. Pluto Press, 2009. ISBN 0745328911.
- [317] R. B. Talisse. Deliberativist responses to activist challenges: A continuation of Young's dialectic, 2005. ISSN 0191-4537.
- [318] Dan Taylor. Everything you need to know about Facebook's EdgeRank, 2011. URL <http://thenextweb.com/socialmedia/2011/05/09/everything-you-need-to-know-about-facebooks-edgerank/>.
- [319] SJ Taylor, E Bakshy, and S Aral. Selection Effects in Online Sharing: Consequences for Peer Adoption. In *Proceedings of the fourteenth ACM conference on Electronic commerce.*, pages 821–836, 2013. ISBN 9781450319621. doi: 10.1145/2482540.2482604. URL [http://papers.ssrn.com/sol3/papers.cfm?abstract={\\_}id=2258864\\$delimiter"026E3B2\\$nhhttp://dl.acm.org/citation.cfm?id=2482604](http://papers.ssrn.com/sol3/papers.cfm?abstract={_}id=2258864$delimiter).
- [320] Techcrunch. EdgeRank: The Secret Sauce That Makes Facebook's News Feed Tick, 2011. URL <http://techcrunch.com/2010/04/22/facebook-edgerank/>.
- [321] M Tessler and E Altinoglu. Political culture in Turkey: Connections among attitudes toward democracy, the military and Islam. *Democratization*, 11(1):22–51, 2004.
- [322] By David Tewksbury. What Do Americans Really Want to Know ? Tracking the Behavior of News Readers on. *World*, (December):694–710, 2003.
- [323] Trendlight. AD is de krant voor Henk en Ingrid | TrendLight, 2012. URL <http://trendlight.eu/ad-is-de-krant-voor-henk-en-ingrid/>.
- [324] Zeynep Tufekci. Facebook and engineering the public., 2014.

- [325] Zeynep Tufekci. How Facebook's Algorithm Suppresses Content Diversity (Modestly) and How the Newsfeed Rules Your Clicks. *Medium*, 2015.
- [326] Twitter. To Trend or Not to Trend, 2010. URL <http://blog.twitter.com/2010/12/to-trend-or-not-to-trend.html>.
- [327] Dan Tynan. How Companies Buy Facebook Friends, Likes, and Buzz. *PCWorld*, 2012.
- [328] N Ungerleider. eBay Is Running Its Own Sociology Experiments, 2014. URL <http://www.fastcolabs.com/3033885/ebay-is-running-its-own-sociology-experiments>.
- [329] Akin Unver. Clash of Communities: Turkey's Dormant Domestic Cold War. *Political Reflection Quarterly Journal* 1, 2(3), 2011.
- [330] Bruce Upbin. Facebook Ushers In Era Of New Social Gestures - Forbes, 2011. URL <http://www.forbes.com/sites/bruceupbin/2011/09/22/facebook-ushers-in-era-of-new-social-gestures/>.
- [331] U.S. Securities and Exchange Commission. Google Inc., Consolidated Balance Sheets, 2009. URL <http://www.sec.gov/Archives/edgar/data/1288776/000119312509150129/dex992.htm>.
- [332] Adil Usturali. *Deliberative and Agonistic Theories of Democracy and their Moral Theories: A New Path towards a Synthesis between the Two?* Master's Thesis in Political Science. PhD thesis, Uppsala University, 2012.
- [333] T. Vaden and J. Suoranta. A definition and criticism of cybercommunism. *Capital & Class*, 33(1):159–177, jan 2009. ISSN 0309-8168. doi: 10.1177/030981680909700109. URL <http://cnc.sagepub.com/content/33/1/159.short>.
- [334] Elizabeth Van Couvering. Is Relevance Relevant? Market, Science, and War: Discourses of Search Engine Quality. *Journal of Computer-Mediated Communication*, 12(3):866–887, apr 2007. ISSN 10836101. doi: 10.1111/j.1083-6101.2007.00354.x. URL <http://doi.wiley.com/10.1111/j.1083-6101.2007.00354.x>.
- [335] J. van Cuilenburg. On Competition, Access and Diversity in Media, Old and New Some Remarks for Communications Policy in the Information Age. *New media & society*, 1(2):183–207, 1999.
- [336] Jan van Cuilenburg. The media diversity concept and European Perspectives. In *Media Economics, Content and Diversity Seminar Finnish Academy of Sciences*, Helsinki, 2002.
- [337] Jan van Cuilenburg. The media diversity concept and European perspectives, 2002.

- [338] Jan van Cuilenburg and Denis McQuail. Media Policy Paradigm Shifts: Towards a New Communications Policy Paradigm. *European Journal of Communication*, 18 (2):181–207, jun 2003. ISSN 00000000.
- [339] Jan van Cuilenburg and R van der Wurff. Toward Easy-to-Measure Media Diversity Indicators. In Els De Bens, editor, *Media Between Culture and Commerce*, page 258. Intellect Books, jul 2007. ISBN 9781841501659.
- [340] I van de Poel. Investigating ethical issues in engineering design. *Science and engineering ethics*, 7(3):429–46, jul 2001. ISSN 1353-3452. URL <http://www.ncbi.nlm.nih.gov/pubmed/11506428>.
- [341] Ibo van de Poel. Values in Engineering Design. *Philosophy of Technology and Engineering Sciences*, 9:973–1006, 2009. doi: 10.1016/B978-0-444-51667-1.50040-9.
- [342] Ibo van de Poel. Nuclear Energy as a social experiment. *Ethics, Policy & Environment*, 14(3):285–290, 2011.
- [343] Jeroen van den Hoven. E-democracy, E-contestation and the monitorial citizen. *Ethics and Information Technology*, 7(2):51–59, 2005. ISSN 13881957. doi: 10.1007/s10676-005-4581-4.
- [344] Jeroen van den Hoven and Emma Rooksby. Distributive Justice and the Value of Information: A (Broadly) Rawlsian Approach. In M.J. van den Hoven and J. Weckert, editors, *Information Technology and Moral Philosophy*, Cambridge Studies in Philosophy and Public Policy. Cambridge University Press, 2008. ISBN 9780511498725.
- [345] Cees van der Eijk. The Netherlands: Media and Politics between Segmented Pluralism and Market Forces. In Richard Gunther and Anthony Mughan, editors, *Democracy and the Media: A Comparative Perspective*, page 510. Cambridge University Press, aug 2000.
- [346] S. Van der Hof and C. Prins. Personalisation and its influence on identities, behaviour and social values. *Profiling the European citizen: cross-disciplinary perspectives*. Springer, pages 111–127, 2008.
- [347] J.J.A van Doorn. Verzuiling: een eigentijds systeem van social controle. *Sociologische Gids*, 3:41–49, 1956.
- [348] L. Vaughan and M. Thelwall. Search engine coverage bias: evidence and possible causes. *Information Processing & Management*, 40(4):693–707, 2004.
- [349] MP Vink. Dutch “Multiculturalism” Beyond the Pillarisation Myth. *Political Studies Review*, 5:337–350, 2007.
- [350] Duncan J Watts. Stop complaining about the Facebook study. It’s a golden age for research, 2014. URL <http://www.theguardian.com/commentisfree/2014/jul/07/facebook-study-science-experiment-research>.

- [351] Zhongyu Wei, Yulan He, Wei Gao, Binyang Li, Lanjun Zhou, and K Wong. Mainstream media behavior analysis on Twitter: a case study on UK general election. *Labour*, (May):174–178, 2013.
- [352] David Weinberger. Echo Chambers= Democracy. In A Fine, M Sifry, A Rasiej, and J Levy, editors, *Rebooting America*, pages 32–37. Personal Democracy Press, New York, 2008.
- [353] Jacob Weisberg. Bubble Trouble Is Web personalization turning us into solipsistic twits?, 2011. URL <http://goo.gl/ET1p06>.
- [354] Aoife White. Google Accused of Abusing Power on Search as Android Probed. *Bloomberg*, 2015. URL <http://www.bloomberg.com/news/articles/2015-04-15/eu-accuses-google-of-antitrust-violations-starts-android-probe>.
- [355] Aoife White. Google Accused of Abusing Power on Search as Android Probed. *Bloomberg*, 2015. URL <http://www.bloomberg.com/news/articles/2015-04-15/eu-accuses-google-of-antitrust-violations-starts-android-probe>.
- [356] Zack Whittaker. Facebook Q1: Mixed earnings; 1.44 billion monthly active users. *ZDNet*, 2015. URL <http://www.zdnet.com/article/facebook-q1-2015-earnings>.
- [357] Zack Whittaker. Facebook Q1: Mixed earnings; 1.44 billion monthly active users. *ZDNet*, 2015.
- [358] Bruce A. Williams and Michael X. Delli Carpini. Monica and Bill All the Time and Everywhere: The Collapse of Gatekeeping and Agenda Setting in the New Media Environment. *American Behavioral Scientist*, 47(9):1208–1230, may 2004. ISSN 00027642. doi: 10.1177/0002764203262344. URL <http://abs.sagepub.com/content/47/9/1208>.
- [359] T Witschge. Examining online public discourse in context: A mixed method approach. *Javnost-the Public*, 15(2):75–91, 2008.
- [360] Ian H Witten, Marco Gori, and Teresa Numerico. *Web Dragons: Inside the Myths of Search Engine Technology*. Morgan Kaufmann, San Francisco, 2006.
- [361] Chad Wittman. Comments 4x More Valuable Than Likes, 2011. URL <https://www.socialbakers.com/edgerankchecker/blog/2011/11/comments-4x-more-valuable-than-likes/>.
- [362] Jill Palzkill Woelfer and David G. Hendry. Homeless young people's experiences with information systems: Life and work in a community technology center. In *Proceedings of the 28th International Conference on Human factors in Computing Systems (CHI '10)*, CHI '10, pages 1291–1300, New York, NY, USA, 2010. ACM. ISBN 9781605589299. doi: 10.1145/1753326.1753520. URL <http://doi.acm.org/10.1145/1753326.1753520>.

- [363] Richard Wray. Search for answers to Google's power leaves UK internet firm baffled. *The Guardian*, 2009. URL <http://www.theguardian.com/business/2009/aug/17/google-search-baffles-internet-firm>.
- [364] Joshua Wright. Defining and Measuring Search Bias: Some Preliminary Evidence. Technical report, International Center for Law & Economics, 2011.
- [365] Xinyu Xing, Wei Meng, and Dan Doozan. Exposing Inconsistent Web Search Results with Bobble. In *PAM 2014*, 2014.
- [366] Tal Yarkoni. In defense of Facebook, 2014. URL <http://www.talyarkoni.org/blog/2014/06/28/in-defense-of-facebook/>.
- [367] Cengin Yilmaz, Oguzhan Aygoren, and Ozlem Ozdemir. Türkiye'de siyasi kutuplaşmayı oluşturan unsurlar: Secmen tercihlerinde ekonomik oy verme davranışından toplumsal travma kuramına kadar bir dizi etkenin görece etkileri. *Iktisat Isletme ve Finans*, 27(311):09–39, 2012.
- [368] E. Yom-Tov, S. Dumais, and Q. Guo. Promoting Civil Discourse Through Search Engine Diversity. *Social Science Computer Review*, pages 0894439313506838–, nov 2013. ISSN 0894-4393. doi: 10.1177/0894439313506838. URL <http://ssc.sagepub.com/content/early/2013/11/24/0894439313506838.abstract>.
- [369] Iris Marion Young. *Inclusion and Democracy (Oxford Political Theory)*. Oxford University Press, USA, 2002. ISBN 0198297556.
- [370] C. Yu, L. Lakshmanan, and S. Amer-Yahia. Recommendation diversification using explanations. In *Data Engineering, 2009. ICDE '09. IEEE 25th International Conference on*, pages 1299–1302, Shanghai, 2009.
- [371] Y. Yue, R. Patel, and H. Roehrig. Beyond position bias: examining result attractiveness as a source of presentation bias in clickthrough data. pages 1011–1018, 2010.
- [372] Kim Zetter. Hacker Lexicon: What Is the Computer Fraud and Abuse Act? *Wired*, 2014. URL <http://www.wired.com/2014/11/hacker-lexicon-computer-fraud-abuse-act/>.
- [373] Mi Zhang and Neil Hurley. Avoiding monotony: improving the diversity of recommendation lists. *RecSys*, pages 123–130, 2008. doi: 10.1145/1454008.1454030. URL <http://portal.acm.org/citation.cfm?doid=1454008.1454030>.
- [374] Michael Zimmer. Facebook's Censorship Problem, jun 2011. URL <http://goo.gl/a89rXX>.
- [375] Jonathan Zittrain. Engineering an election. *Harvard Law Review Forum*, 127:335–341, 2014.